DOCTOR OF PHILOSOPHY

Picture name agreement as a measure of lexical co-activation in word production: Behavioral and electrophysiological insights

Balatsou, Evangelia

Award date:
2020

Awarding institution:
Bangor University

Picture name agreement as a measure of lexical co-activation in word production: Behavioral and electrophysiological insights

Evangelia Balatsou

PRIFYSGOL
BANGOR
UNIVERSITY

This thesis is submitted in partial fulfilment of the requirement for the degree of Doctor in Philosophy, completed in the school of Psychology, Bangor University.

Yr wyf drwy hyn yn datgan mai canlyniad fy ymchwil fy hun yw'r thesis hwn, ac eithrio lle nodir yn wahanol. Caiff ffynonellau eraill eu cydnabod gan droednodiadau yn rhoi cyfeiriadau eglur. Nid yw sylwedd y gwaith hwn wedi cael ei dderbyn o'r blaen ar gyfer unrhyw radd, ac nid yw'n cael ei gyflwyno ar yr un pryd mewn ymgeisiaeth am unrhyw radd oni bai ei fod, fel y cytunwyd gan y Brifysgol, am gymwysterau deuol cymeradwy.

I hereby declare that this thesis is the results of my own investigations, except where otherwise stated. All other sources are acknowledged by bibliographic references. This work has not previously been accepted in substance for any degree and is not being concurrently submitted in candidature for any degree unless, as agreed by the University, for approved dual awards.

Acknowledgments

First of all, I would like to thank my principal supervisor Dr Gary Oppenheim for giving me the opportunity to undertake this PhD. This work would not have been possible without his support, encouragement and transfer of knowledge. I value his mentorship and I am lucky to become a scientist next to him. I also owe a lot to my second supervisor, Professor Guillaume Thierry, for his constant scientific guidance, for teaching me ERPs, but also for the huge support and understanding-I will always be grateful. I also thank Dr Audrey Bürki for the short doctoral visit to her lab at the University of Potsdam and her introduction to state-of the art analytical methods of electrophysiological signal.

Thanks to Dr Kami Koldewyn for her support and encouragement as the chair of my committee. I would also like to thank the department of Psychology for funding this PhD, the staff and the amazing BLPL and BULET labs, which became my second homes. A big thank you to all the students I taught in SWAC and in the Psychology of Language and to the students I mentored in research, who also assisted in data collection: Annie, James, Sabrina, Sam, Alexina, David and Kieran. Thanks to my colleagues, labmates and friends in Bangor for the stimulating discussions, the fun and the siblinghood: Elena, Pauliina, Myrto, Jen, Pierre, Inga, Rafal, Martina, Michela and Candice.

Thanks to every single author in the reference list of this thesis for their contribution to my better understanding of how language, the brain and stats work. “We die. That may be the meaning of life. But we do language. That may be the measure of our lives.”, as Toni Morrison said. I will always cherish everything I have learned all these years as a language researcher.

I am forever grateful to my family and friends in Greece for the love and support throughout the years. My mom, Chrysa, may not understand a lot about cognitive neuroscience, but she always had my back and encouraged my intellectual curiosity. Thanks to my late dad, Thanasis, my stepfather, Sotiris, my brother, Vaios, my grandfathers Stamatis and Vaios and my grandmothers, the Evangelias.

Finally, I want to thank Agis. Through his love and support he has helped to make the end of this journey much easier and this author much happier.

Abstract

What happens in our mind before we produce a word? It is known that naming a picture of a dog is faster than naming a picture of a couch. One possibility why this happens is because a picture of a dog will almost always be named as “dog”, whereas a picture of a couch may elicit the name “sofa” or even “settee”, indicating lower picture-name agreement. *Picture name agreement* is a measure of the proportion of speakers who independently produce a picture’s modal name when asked to name it. This measure of lexical availability for pictures is associated with robust effects in word production, which have been assumed to index the competition between lexical representations for selection. On this account, “couch” and “sofa” actively race for selection, delaying production speed, until the best option is ultimately chosen. But is such competition warranted? The research reported in the current thesis examined picture name agreement as a measure of lexical co-activation in word production in a bid to clarify whether selecting words for speech requires such an active competition between representations. By measuring speakers’ word choices, naming latencies and electrophysiological activity as they named pictures with high and low name agreement in a variety of simple tasks, I was able to show that these effects point to the co-activation of linguistic representations in their minds, but also index speakers’ unique idiosyncratic preferences for specific words. Overall, variations in picture name agreement fail to provide strong support for a competitive account of lexical selection, but instead favor a view in which co-activation of words in the mental lexicon appears to be effortful, but eventually leads to the successful production of the best candidate for each individual speaker.

Table of Contents

CHAPTER 1-General Introduction

1.1 On language and word production

Language is, undeniably, the most complex communication system in the known universe and the act of speaking alone is an acquired mastery developed through years of neurobiological evolution from the earliest ancestors (Fisher, 2019). The average baby will articulate their first words at around 10 months of age (Schneider, Yurovsky, & Frank, 2015), while the average monolingual adult says about 16 thousand words per day in their native language (Mehl, Vazire, Ramírez-Esparza, Slatcher & Pennebaker, 2007) speaking at a rate of 1 to 2 words per second (Yuan, Liberman & Cieri, 2006). Through decades of intensive research in cognitive science fields, speaking has extensively been investigated as both a unique and complex mastery of everyday communication and as a window into further understanding the human brain.

Thirty-one years after Levelt's (1989) seminal work "Speaking" was first published, the field of language production is still growing, while, at the same time, new findings provide invaluable insight into the core questions related to speaking. In order to broadly understand how communicative language production works, it is necessary to have a well-established model of how the brain processes the smallest, independent, semantically meaningful unit: words. The literature abounds with invaluable empirical insights about the cognitive processes that precede single word retrieval. For instance, it is generally agreed that even before selecting a word of their choice, speakers are processing the conceptual, semantic, lexical and phonological properties of the intended word and other similar words in their mental lexicons (Dell, 1986; Levelt, Roelofs & Meyer, 1999). However, several narrower theoretical questions remain either partially unanswered or hotly debated to this day: How does the human brain select one appropriate word, out of the tens of thousands that one knows, for articulation? Do speakers continuously monitor their speech in typical everyday production? What is the precise nature of the mechanism that is involved in word selection? The current work aims to contribute to a better understanding of the answers to these smaller, yet very important, questions.

A central area of interest in the current work is how the brain selects a single word for production. In order to produce the intended word, speakers must initially select it amongst numerous equally suitable alternatives from their mental lexicon, a process known as lexical selection, and then compute the articulatory movements of production based on its phonological properties (Levelt, 1991). Lexical selection is perhaps the most integral aspect of language production and, despite being a remarkably fast and automatic process, it can also become unexpectedly challenging, for instance for individuals with language disorders (e.g., Dell, Schwartz, Martin, Saffran & Gagnon, 1997). The nature and mechanisms that underlie lexical selection in typically developed adults have been extensively investigated and concurrently hotly debated in the field, with one prominent view suggesting that selection speed simply depends on the weight of the target word only, independent of other words in the mental lexicon (Dell, 1986; Mahon, Costa, Peterson, Vargas & Caramazza, 2007), and the alternative hypothesis arguing that selection is an actively competitive process that depends on the weight of other co-activated lexical representations (Levelt et al., 1999). The nature of the selection mechanism is usually investigated by examining retrieval speed in variations of picture naming tasks, like simple picture naming (Bose & Schafer, 2017), picture naming norms (Bates et al., 2003), picture word interference (Glase & Düngelhoff, 1984), blocked cyclic naming (Crowther & Martin, 2014) and continuous naming (Aristei, Zwitserlood & Abdel Rahman, 2012).

In simple picture naming, researchers try to address the nature of lexical selection by evaluating the behavioral effects of word production in relation to the proportion of participants in a norming study who volunteer a picture's modal name (i.e., the most frequent non-omission response), which is referred to as *picture name agreement*. Pictures with high name agreement (e.g., dog) clearly elicit a single dominant name in participants' responses, while low name agreement pictures (e.g., couch) are named with a wider variety of, often, similarly appropriate words (e.g., sofa, settee or even armchair). Low name agreement is associated with robust behavioral and electrophysiological effects in naming and most studies interpret those as directly indexing within-speaker ambiguity in the form of

competition between possible names (e.g., Shao, Roelofs, Acheson, & Meyer, 2014; Bose & Schafer, 2017). For instance, naming a picture of a dog presumably requires fewer cognitive resources than naming a picture of a couch, which can also be called a sofa or a settee, and significantly fewer compared to a picture with increased name uncertainty, like that of an electric can opener, for which individuals produce sixteen different names (Székely et al., 2003).Under the assumption that weaker dominant name agreement or higher numbers of alternatives imply stronger competition, the behavioral and neural effects of this increased variability are often claimed to reflect the ongoing, systematic competition between "couch", "sofa" and "armchair" to be selected for articulation (e.g., Bates et al., 2003; Szekely et al., 2004; Alario et al., 2004; Kan & Thompson-Schill, 2004; Novick, Kan, Trueswell & Thompson-Schill, 2009; Rodríguez-Ferreiro, Menéndez, Ribacoba, & Cuetos, 2009; Bose & Schafer, 2017; Cheng, Schafer, & Akyürek, 2010; Shao et al., 2014). However, while the behavioral and neural costs of low picture name agreement are very robust, and the competition explanation of them seems prima facie cognitively plausible, the existing evidence does not specifically establish that the effects stem from active lexical competition, instead of merely reflecting the cost that comes with weak target activation and increased co-activation of alternative words.

1.2 Overview of the thesis

The current thesis aims to elucidate the psychological and neural bases of word production, by questioning and empirically assessing several key assumptions that underlie the common interpretation of name agreement effects as indexing a competitive lexical selection process. In Chapter 1, I review the current literature on word production and picture name agreement and discuss some of the important conceptual issues that drive the current experimental work. Chapter 2 presents a detailed overview of the behavioral, electrophysiological, and analytical techniques used in the following experimental chapters. A key issue towards understanding the impact of picture name agreement on word production processes is to initially assess whether name agreement is a reliable measure of lexical

co-activation within the individual speaker. Chapter 3 reports an investigation of whether such variations in name agreement directly index the options that each individual considers in each word retrieval or instead reflect heterogeneity among individuals' stable word preferences, i.e., their *idiolects*. Chapter 4 provides another important step towards understanding how lexical selection typically works, by evaluating whether the mechanisms that are associated with low name agreement (choosing 'sofa' over 'couch' in simple picture naming) resemble those that speakers engage to accomplish directed name changes (overriding an existing preference for 'sofa' in order to produce 'couch' instead). Finally, Chapter 5 provides an exploratory baseline assessment of electrophysiological modulations associated with low name agreement, specifically focusing on cases of limited within-speaker lexical co-activation. By understanding how picture name agreement variations affect word production processes, the current thesis aims to further clarify the nature of the mechanism involved in lexical selection and at the same time bridge the gap between the cognitively plausible models of word production and the experimentally supported evidence for either one of them.

1.3 Review of the literature on word production

So how does the intention to communicate an idea become the act of articulating a single word? The most prominent models of language production agree on the existence of three necessary processing levels before producing a single word: conceptualisation, formulation and articulation (Caramazza, 1997; Dell, 1986; Dell et al., 1997; Levelt et al., 1999). During conceptualisation, the abstract concept of speech is determined, which can also be triggered by external perceptual visual or auditory stimuli (Levelt et al., 1999). Although it is yet inconclusive whether the conceptualizer involves any linguistic knowledge (Levinson, 1997) or not (Bierwisch & Schreuder, 1992), it is generally assumed that during conceptualization, the content (i.e., the *what*) of speaking is determined. Following, the abstract representation is further processed for lexical retrieval in formulation, where its word form is accessed (word or "lemma" selection), along with its phonological properties

(phonological encoding). Finally, during articulation, the lexical and phonological features of the intended word are executed into motor programmes, resulting in the actual act of speaking (Griffin & Ferreira, 2006). An additional optional process of self-monitoring is thought to be involved in word production, whereby speakers monitor their responses according to certain metrics or heuristics, either via the comprehension system (Levelt et al., 1999) or by a production-based process (Nozari, Dell, & Schwartz, 2011).

1.3.1 Major differences between models of word production

Although the literature generally agrees on the existence of these three essential processes involved in speech production, there are certain qualitative differences in both the interaction between these processing levels and in how knowledge is individually represented at each level. Discrete/serial models of word production (Levelt, 1989; Levelt et al., 1991; Levelt et al., 1999; Schriefers, Meyer, & Levelt, 1990), view conceptualization, formulation and articulation as modular processes (Fodor, 1983) and argue that there is no feedback activation from lower levels to the higher levels of processing (feed-forward only processing). According to this view, only after the target lexical representation has been selected, it can pass on activation to its corresponding phonological nodes. For instance, only after the selection of “dog” has been completed, / /d/, /ɒ/, and ɡ/ can receive activation in phonological encoding. Cascaded and interactive models of lexical access (Caramazza, 1997; Costa, Caramazza, & Sebastian-Galles, 2000; Cutting & Ferreira, 1999; Dell, 1986; Dell et al., 1997; Dell & O'Seaghdha, 1991; Morsella & Miozzo, 2002; Peterson & Savoy, 1998) on the other hand, are rooted on the idea of spreading activation within the system, wherein the processing levels are not discrete but interconnected and the activation can flow from the lowest points of sublexical levels to higher levels of processing (feedback processing). In these models, activation can proportionally flow from lexical representations to phonological nodes even before lexical selection has been completed, with the addition that it can also flow backwards: from phonological nodes to their corresponding lexical

representations. Following the example above, /d/, /ɒ/, and ɡ/ could have received activation even before the selection of "dog" has been completed, while they can also proportionally pass on their activation back to the level of lexical selection. In noisy production circumstances, false nodes may receive higher activation than the right nodes, resulting in a speech error, like erroneously producing the word "dag" instead. The nature of such mixed errors in normal and aphasic speech strengthens the possibility that activation at the phonological level may reflect back on activation at earlier processes (Dell 1986; Martin, 1996; Rapp & Goldrick, 2000), however serial models instead ascribe such error patterns to a post-lexical self-monitoring mechanism that is not part of the production system (Levelt et al., 1991;1999; Roelofs, 2004).

Another major difference between these two classes of models lies in their view of how linguistic information is retrieved from the mental lexicon during the process of formulation. To this date, there is an inconclusive debate on the nature of the mechanism that determines lexical selection. A family of production models suggest that lexical selection is a competitive process (e.g., Levelt et al., 1999), while others view lexical selection as non-competitive (e.g., Mahon et al., 2007). But what exactly is *lexical competition*? It is generally agreed that the activated semantic nodes (e.g., four-legged, pet, furry) pass on their proportional activation to their corresponding words or "lemmas" (e.g.,"dog", "cat") (Dell, 1986; Levelt et al., 1999). The competition hypothesis assumes that, in cases where more than one lexical node receives activation (e.g., dog and cat are both activated in the "lemma stratum"), the process of lexical selection is delayed (Levelt et al., 1999; Roelofs 1992; 1993; 1997; 2018), meaning that the time required to select the target word is negatively affected by the *relative* activation of all other nontarget words in the mental lexicon. One possibility of how this happens in the production system is via lateral inhibition: concurrently activated nodes can pass on inhibitory activation -which is relative to their own activation levels- to other nodes and this process delays retrieval speed (Cutting & Ferreira, 1999; Howard, Nickels, Coltheart, & Cole-Virtue, 2006). An alternative way to model the principle of competition, in which co-activated nodes do not modulate

each other in real time, was proposed by Roelofs (1997) and extended by Levelt et al. (1999) and that is via a probabilistic selection process determined by the Luce choice ratio (Luce, 1959), which suggests that selection probability of a single word equals the ratio of its activation to that of all the other activated words in the system. This probability of a particular word to be selected at any given time is the most critical principle of competition in Levelt et al.'s (1999) model, because it suggests that competitive lexical selection should have profound chronometric effects (i.e., effects in response times). In this framework, the hypothesis of lexical competition was formulated with a focus on studying reaction time (RT) effects in "normal" word production and not by examining speech errors, which Levelt et al. (1999) considered as "infrequent derailments of the process" (p.2).

Other accounts either view lexical selection as strictly non-competitive, or are agnostic in relation to the chronometric effects of lexical selection. In non-competitive theories, the activation of a certain lexical representation relative to that of any alternatives still affects its *probability* of selection via a winner-take-all function (e.g., the Luce Choice ratio), but its *latency* of selection depends on its activation alone, irrespective of the number of co-activated alternative words or the total activation of other candidates (e.g., Dell et al. 1997; Oppenheim, Dell & Schwartz, 2010). Thus, non-competitive models do not assume that formulation of the target lexical representation is slowed down by the temporary activation of other candidates in the lexicon and therefore suggest lexical selection is fundamentally a non-competitive process (Mahon et al. 2007). Several other word production theories do not directly address questions of timing, but they do not incorporate a competitive selection mechanism in the narrow sense of Levelt et al.'s (1999) theory either (e.g., Dell, 1986; Stemberger, 1985; Caramazza, 1997; Rapp & Goldrick, 2000). However, even though non-competitive models do not implement competition in the lexical selection process they do not necessarily eliminate competition in the broader sense: some theories include competition-like features in other processing levels or forms, such as competitive learning (Oppenheim et al., 2010) or later-stage response interference (Mahon et al, 2007).

1.3.2 Lexical selection in word production: evidence for and against competition

Since the main prediction of the lexical competition theory is that the naming latencies of the target word will increase as the levels of co-activation of similar lexical entries increase, the majority of the empirical evidence supporting competition derive from paradigms which attempt to manipulate the activation of nontarget words and in particular, the semantic interference effect observed in the picture-word interference tasks. In the picture-word interference paradigm, participants are asked to name a picture, which is usually a line drawing, as accurately and quickly as possible while ignoring a distractor word (either in written or auditory format) presented at the same time with the picture (see Bürki, Elbuy, Madec & Vasishth, 2020, for a recent review of the empirical findings in this paradigm). Because of the simultaneous processing this paradigm requires, it has been intensely debated whether it is a lexical-semantic Stroop task, with different theories attributing the locus of interference either at the perceptual encoding stage or during the response selection stage (see Dell'Acqua, Job, Peressotti & Pascali, 2007 vs van Maanen, van Rijn & Borst, 2009). The major finding of this task is that picture naming is generally faster without any present distractor words (de Zubicaray & McMahon, 2009), non-lexical distractors (such as rows of symbols) (Hirschfeld, Jansma, Bölte, & Zwitserlood, 2008), or pictorial distractors (Bloem, & La Heij, 2003). The core finding in picture-word interference studies is that the naming latency of the target picture (e.g., horse) increases in the presence of categorically semantically related distractor words (e.g., dog) compared to unrelated words (e.g., pencil) (e.g., Schriefers et al., 1990). In contrast, distractor words with phonological similarity (e.g., cap) are associated with faster target naming latencies (e.g., cat), as are semantically non-categorical relationships (e.g., part-whole in piano and keys) (Costa, Alario, & Caramazza, 2005) and associative relationships (e.g., bone-dog) (Alario, Segui & Ferrand, 2000; Abel et al., 2009; de Zubicaray, Hansen & McMahon, 2013).

This interference caused by semantically related distractors has to this day been the most prominent finding in the literature of word production in favour of the lexical selection by competition

hypothesis. The main assumption is that semantically related visual distractors co-activate alternative lexical representations (i.e., their corresponding lexical nodes which are similar to the target word) causing retrieval delay for the target word or "lemma" in production (La Heij, 1988), because the selection mechanism is competitive (Levelt et al., 1999): response delay for the target word in picture-word interference tasks is assumed to reflect the conflict at the level of formulation and therefore this behavioral cost should be directly indicative of a selection mechanism that tries to decide between two equally appropriate lexical representations. This interpretation has been widely adopted in the literature (Schriefers, et al.,1990; Levelt et al., 1999; Roelofs, 1992; 1993; 2001; 2003), while the interference effect has been replicated in a number of different paradigms and tasks (Humphreys, Lloyd-Jones & Fias, 1995; Starreveld & La Heij, 1995; Vitkovitch, & Tyrrell, 1999; Damian, & Bowers, 2003; Bloem, van den Boogaard, & La Heij, 2004).

As an alternative, non-competitive explanation of this effect, Mahon et al. (2007) argued that the interference previously reported does not arise at the lexical level and presented robust empirical evidence which indicates that increased co-activation does not reflect the increased competition during lexical selection that Levelt et al., (1999) have described. Mahon et al. (2007) showed that as distractor words (e.g., zebra) become more semantically similar to the target word (e.g., horse), naming latencies decrease significantly, compared to within-category semantically more distant distractors (e.g., whale). The finding that co-activated lexical representations that are semantically closer to the target words facilitate the production of the target, imposes an important limitation on the nature of the chronometric effects of the competition hypothesis and more specifically on plausibility of using the Luce choice rule to explain naming latencies. In line with cascaded models of word production, Mahon et al. (2007), claimed that the levels of activation of the non-target nodes do not affect the time required to select the target node (as in Dell, 1986) and additionally suggested an alternative explanation of the latency effects in picture-word interference tasks: the interference arises at a post-lexical, output level, possibly reflecting a decision mechanism that is sensitive to the response-relevant criteria of the paradigm,

rather than systematic competition due to increased endogenous co-activation (also known as the Response Exclusion Hypothesis). The Response Exclusion Hypothesis, though, is primarily supported by empirical findings of semantic facilitation and not semantic interference (see Abdel Rahman & Melinger, 2009 for a discussion), which-to date-remains the strongest ally of lexical competition.

While the majority of converging evidence in favour of competition mainly originates from picture-word interference findings, the alternative theory, that selection is a non-competitive process, gains ground from both the speech error literature and empirical findings in other picture naming paradigms. In traditionally non-competitive theories, which suggest that the time required to select the target word is not dependent on the levels of co-activation of non-target words in the system (e.g., Caramazza, 1997; Dell, 1986; Rapp & Goldrick, 2000; Oppenheim, et al.2010), instead of a Luce-choice-inspired selection principle or lateral inhibition that determines selection latency, selection and naming latencies are successfully modeled with non-competitive criteria. A target word is selected for production when it reaches an *absolute* threshold (e.g., Dell et al., 1987; Oppenheim et al., 2010; Mahon et al., 2007), irrespective of the *relative* threshold of activation of other words in the system (Roelofs 1992; 1993; 1997; 2018; Levelt et al., 1999).

Although non-competitive theories were not primarily modeled to account for the chronometric effects during lexical retrieval (e.g., Dell et al., 1987) in the way competition did, computational principles, like that of cascading activation, feedback processing, and interactivity have historically been more robust in explaining the majority of speech error patterns (e.g., Dell, 1986; Dell et al., 1997). Other non-competitive theories which do not implement interactivity, but rather suggest that the locus of response delay is outside the lexical level (e.g., Mahon et al. 2007) or that latency effects result from interference in the form of error-driven lexical learning, have also explained a range of behavioral observations by typically developed and clinical populations in word naming tasks without lexical-level competition (e.g., Oppenheim et al., 2010; Oppenheim, Tainturier & Barr, 2016). Specifically, and although it seems rather impossible to directly estimate endogenous lexical-level co-activation

within the same trial, the hypothesis of a non-competitive lexical selection has also been tested by manipulating co-activation levels across naming instances, as in the blocked cyclic and continuous naming paradigms.

In picture naming, it is generally accepted that retrieving a particular word from the mental lexicon once, makes it easier, less error-prone and faster to retrieve the same word in subsequent trials during the same task (Mitchell & Brown, 1988). This is known as repetition priming, a mechanism associated with various other cognitive processes, such as lexical access (e.g., Forster & Davis, 1984) word identification (e.g., Bodner & Masson, 1997) face recognition (e.g., Ellis, Young, Flude & Hay, 1987) and visual attention (e.g., Kristjánsson & Campana, 2010). In particular, even a single naming episode in picture naming tasks can have very long-lasting effects, facilitating retrieval speed up to 48 weeks after initial exposure (e.g., Cave, 1997). However, this benefit of repetition in naming for the target word also comes with a cost for the ease of accessibility of the other non-selected words in the mental lexicon: retrieving a target word once (e.g., cat) delays the subsequent retrieval of other words from the same semantic category, which were co-activated but not eventually selected (e.g., dog or tiger) (Howard et al., 2006; Vigliocco, Vinson, Damian & Levelt, 2002; Wheeldon & Monsell, 1994; Oppenheim et al., 2010; Abdel Rahman & Melinger, 2007; Belke, 2008). This is another form of semantically-driven interference in continuous naming, in which naming latencies for within-semantic category items (i.e., animals) increase linearly over each ordinal position within that specific category (e.g., Howard et al.,2006; Belke, 2013; Costa, Strijkers, Martin, & Thierry, 2009). A typical extension of this paradigm is blocked cyclic naming, in which participants name items in separate cycles, either in homogeneous blocks consisting of items from the same semantic category (i.e., a block consisting only of pictures of animals) or in heterogeneous blocks, which consist of items from different categories (i.e., a block consisting of pictures of tools, animals, clothing etc.). Naming latencies in the semantically homogenous blocks are almost always longer than in the heterogenous blocks over

increased naming cycles (e. g., Belke et al., 2005; Harvey & Schnur, 2016; Schnur, Schwartz, Brecher, & Hodgson, 2006).

In the literature, this effect has also been explained in terms of lexical competition, by assuming that semantically related words are slowing down the retrieval of the target, either by modulating the activation of the competitors during speech planning (Howard et al., 2006) or by assuming a conceptual bias which reverberates lexical activation before a competitive selection emerges (Roelofs, 2018). However, as with picture word interference, it has also been shown that a competitive lexical selection is not required to account for the interference in continuous and blocked cyclic naming. Instead, the interference can be explained by an incremental learning mechanism which, after each retrieval episode, readjusts the connections from semantic features into words, by strengthening those of the target word (e.g., cat) and weakening those of the unselected competitors (e.g., dog, tiger and horse) (Oppenheim et al,. 2010). In paradigms where interference accumulates, lexical selection is delayed in the same fashion: after each retrieval the competitor words become less and less accessible, so as their ordinal position increases (i.e., as more items from the same semantic category are named), they receive more inhibitory activation. This competitive learning and unlearning follow the delta rule, which is a core training algorithm in most connectionist models. It is, though, possible that the notion of continuous priming as error-based lexical learning based on the delta rule that Oppenheim et al. (2010) illustrated as a driving force of the cumulative semantic interference effect, may extended into other production tasks as well (see Oppenheim & Balatsou, 2019).

In any case, while the chronometric effects observed in simpler paradigms, like in continuous naming and blocked cyclic naming, can be explained with and without a competitive lexical selection process, the majority of the empirical evidence in favour of lexical competition still derive from picture-word interference findings (Roelofs, 2018). There remains, thus, a need to address the nature of lexical selection either by developing new experimental approaches which can be provide strong support for either one of these hypotheses or by revisiting the debate and viewing lexical selection as

a not necessarily dichotomous process (i.e., either strictly competitive or strictly non-competitive) (Nozari & Hepner, 2019; cf. Oppenheim & Balatsou, 2019). In an attempt to bridge the gap between strictly competitive and non-competitive theories and the empirical findings supporting them, Nozari and Hepner (2019) recently put forward the idea of a flexible criterion determining lexical selection. In this framework, the nature of lexical selection is determined by task goals (e.g., speed versus accuracy trade-off) and the general capacity of the production system to achieve them (i.e., the inherent features of lexical activation), highlighting the difference between the necessary process of co-activation and a flexible lexical selection, which can emerge as situationally competitive according to task demands. When task demands introduce external conflict in the system, which can either be by a response trade-off in the task or the demand to name a picture while ignoring a competitor word, the criterion for selection is flexibly adjusted to incorporate this competition, while in cases where there is not a specific response demand, lexical selection is determined according to the inherent properties of the words. A crucial component of the flexible criterion theory is that the selection mechanism itself is naturally adjustable and capable of incorporating such constraints in the same level of processing and not at a separate, later post-hoc level, as predicted by other theories (e.g., Mahon et al., 2007).

With these in mind, we (Oppenheim & Balatsou, 2019) have recently advocated for the need to distinguish between "studying language production as it is and modifying it to fit particular laboratory constraints that we think might highlight particular aspects of the process" (p. 3) and proposed that, a necessary step to resolve the debate is to evaluate the chronometric effects of selection in paradigms less prone to misinterpretations and more closely resembling communicative speech, such as simple picture naming. Simple picture naming, unlike picture-word interference and blocked-cyclic studies, allows the observation of the behavioral effects associated solely with the ease of lexical selection by omitting experiment-specific and circumstantial confounds, like the interference and facilitation caused by repetition or by the simultaneous processing of pictures and words. In simple picture naming, researchers usually evaluate the behavioral patterns of each word retrieval and their

association to common variables of production difficulty, thus, allowing them to make more direct associations between the empirical findings in the lab and the nature of the encoding processes.

1.4 Simple picture naming and traditional uses of name agreement

Simple picture naming is one of the most widely used tasks in cognitive psychology and an elementary process of production closely related to the everyday use of language (Glazer, 1992). Picture naming is sometimes considered a semantic task, because it requires an obligatory access to the semantic system (e.g., Bookheimer, Zeffiro, Blaxton, Gaillard, & Theodore, 1995; Caramazza, Hillis, Rapp, & Romani, 1990). Naming a picture also requires object identification, name activation and response generation, which corresponds well to the necessary processing stages of word production: conceptualization, formulation and articulation (Johnson, Paivio & Clark, 1996). Although it is also not yet established whether these processes are serial (e.g., Alario et al., 2004) or cascaded (e.g., Humphreys, Riddoch & Quinlan, 1988) in picture naming, along with the general debate on the nature of these processes (Dell et al., 1986 vs Levelt et al., 1999), this task has been extensively used in word production literature to test hypotheses about the cognitive mechanisms associated with single word retrieval.

Naming a picture was assumed to be an indicative measure to test mental processes from early on (e.g. Cattell, 1886; Fraisse, 1968; 1969). In psychological research, timed picture naming has been used to investigate language production in typically developed adults (Lachman, 1973; Lachman, Shaffer, & Hennrikus, 1974; Sanfeliu & Fernandez, 1996; Snodgrass & Vanderwart, 1980; Snodgrass & Yuditsky, 1996; 1999; Torrance et al., 2018), patients with brain injury (Kohn & Goodglass, 1985; Chen & Bates, 1998; Druks, 2002; Druks & Shallice, 2000; Goodglass, 1993; Murtha, Chertkow, Beauregard, & Evans, 1999; Nilipour, Bakhtiar, Momenian, & Weekes, 2017) as well as both typically developed children and children with language impairments (Cycowicz, Friedman, Rothstein, & Snodgrass, 1997; D'Amico, Devescovi, & Bates, 2001; Davidoff & Masterson, 1996; Dockrell,

Messer, & George, 2001; Nation, Marshall, & Snowling, 2001; Nilipour, Pourshahbaz, & Momenian, 2017). In brain research, simple picture naming has been extensively used as a task in fMRI studies (e.g., Spitzer, Kwong, Kennedy, Rosen, & Belliveau, 1995; Damasio et al., 2001; Hernandez, Dapretto, Mazziotta, & Bookheimer, 2001; de Zubicaray, McMahon, Eastburn & Pringle; 2006; Saccuman et al, 2006; de Zubicaray, & McMahon, 2009; Liljeström, Hultén, Parkkonen & Salmelin, 2009; Bose & Schafer, 2017) as well as in Event-Related Potentials (ERPs) studies (e.g., Barrett & Rugg, 1990; Schmitt, Münte, & Kutas, 2000; van Turennout, Hagoort, & Brown, 1997;1998;1999; Wicha, Bates, Moreno, & Kutas, 2000; Verhoef, Roelofs & Chwilla, 2009; Cheng et al., 2010; Laganaro, Python & Toepel, 2013; Shao et al., 2014).

Timed picture naming is also methodologically useful in cognitive psychology in order to collect *norms*, a set of standardized pictures that can be later used by other researchers who wish to include valid stimuli when designing experiments (Snodgrass & Vanderwart, 1980; Snodgrass & Yuditsky, 1996; Bates et al., 2000; 2003; Szekely et al., 2003). Researchers typically select the better recognized pictures amongst the pool of norms, in order to use them as unbiased material to test a variety of cognitive processes in typical adults (e.g., Laganaro, Valente & Cyril, 2012; Valente & Laganaro, 2015) and clinical populations (Bormann, Kulke, Wallesch, & Blanken, 2008; Fieder, Nickels, Biedermann, & Best, 2014). In norms and in other simple timed picture naming studies, the variables that are associated with each pictorial stimulus and their corresponding lexical responses are usually assessed and evaluated in relation to naming latencies (e.g., Bates et al., 2003; Szekely et al., 2003; Alario et al., 2004). Such variables can be classified as visual (e.g., visual complexity, image agreement), visual-to-semantic (e.g., concept familiarity), semantic (e.g., imageability), semantic-to-lexical (e.g., name agreement), lexical (e.g., word frequency and age of acquisition) or phonological (e.g., number of phonemes and syllables of a word) (Alario et al., 2004). From early on in picture naming, age of acquisition (an estimate of the age in which a word has been learned) and lexical frequency (the degree of use of a word in a language) appeared to have strong effects on retrieval

speed, but name agreement soon appeared to have unique effects on naming latencies irrespective of these factors (Lachman, Shaffer, & Hennrikus, 1974; Vitkovitch & Tyrrell, 1995; Lachman, 1973; Lachman, Lachman, Thronesbery, & Sala, 1980).

Picture name agreement, sometimes termed codability (Gilhooly & Gilhooly, 1979; Lachman, 1973; Lachman et al, 1974; 1980) is the empirically-derived measure of the number of different names given to an image by the participants who are asked to name it. Stimuli categorically classified as pictures of high name agreement usually elicit a single target response from the entire linguistic population upon naming. For example, the International Picture Naming Project (IPNP) stimulus "obj128dog" is a high name agreement picture in UK English, because the vast majority - if not all - of the participants will respond with the word "dog" ($p_{\text{DominantName}} = 1$) (Oppenheim, in prep.). Pictures of low name agreement usually elicit various responses amongst a linguistic community: the IPNP stimulus "obj472truck" has low name agreement ($p_{\text{DominantName}} = .59$) because it can also be called a lorry ($p_{\text{SecondaryName}} = .37$) (Oppenheim, in prep.). Sources of variability in naming can be due to the different lexical representations that exist within a community of speakers (i.e., a couch can also be called a sofa or a settee by UK-English speakers), the use of possible abbreviations in naming (i.e., plane instead of aeroplane) or because a picture may trigger irrelevant responses (i.e., naming a tomato as an apple) (Vitkovitch & Tyrrell, 1995).

Pictures with high name agreement generally elicit faster responses than pictures with low name agreement (e.g., Gilhooly & Gilhooly, 1979; Lachman, 1973; Lachman et al., 1974; 1980; Vitkovitch & Tyrrell, 1995). In a large scale normative study in the French language, Alario et al. (2004) found that picture name agreement was, independently of eight other statistical predictors associated with the conceptual factors in picture naming (image agreement, concept familiarity, visual complexity, imagability, age of acquisition, written frequency, number of phonemes, number of syllables), the *strongest* predictor of picture naming speed, validating name agreement as perhaps the most important variable to predict retrieval success and speed. Importantly, this strong association

between picture name agreement and naming latencies is cross-linguistic (Bates et al., 2003): it has been found in American and British English (Snodgrass and Yuditsky, 1996; Ellis and Morrison, 1998; Szekely et al., 2004), Spanish (Cuetos et al., 1999), French (Bonin et al., 2002; Kremin et al., 2000), Italian (Dell'Acqua et al., 2000), Greek (Dimitropoulou, Duñabeitia, Blitsas, & Carreiras, 2009), Japanese (Nishimoto, Ueda, Miyawaki, Une, & Takahashi, 2012) and Persian (Bakhtiar, Nilipour, & Weekes, 2013).

This strong effect which was not language-sensitive elicited the assumption that name agreement, apart from a useful methodological variable, should also be a cognitively meaningful predictor of individuals' internal states. Name agreement was initially used as a measure to evaluate performance in semantic memory tasks and presented a behavioral dissociation between semantic and episodic memory in typically-developed adults: pictures with low name agreement were named with longer naming latencies and were more prone to erroneous responses compared to high name agreement pictures, which was used to measure semantic memory performance, but they yielded better recall performance overall, which was used to measure episodic memory performance (Mitchell, 1989). The opposite dissociation was observed when lexical frequency was used as a measure, since more frequent words were found to be easier to recall for memory (Mandler, Goodman, & Wilkes-Gibbs,) and at the same time faster in naming (Carroll & White, 1973), indexing a functional difference in memory recall for the two most common variables associated with the retrieval of words from the mental lexicon (i.e., name agreement and lexical frequency). Other functional differences based on a picture's codability have also been reported, such as the developmental dissociation between repetition priming and episodic memory in children and young adults, in which name agreement variations persist after repetition but not recall (Lorsbach & Morris, 1991; Mitchell & Brown, 1988).

1.4.1 Picture Name Agreement as an index of lexical competition

In relation to theories of word production, the robust effects of picture name agreement have been interpreted as an index of the processes that occur within the individual speaker in the level of lexical selection. In line with Vitkovitch and Tyrell‘s (1995) sources of name disagreement, researchers began noticing that there is a distinction between the uncertainty of pictures and the existence of alternative names for the depicted objects, factors that both result in name disagreement. The former empirical measure was classified as image agreement and is identified as the match or mismatch between a picture and the canonical representation of its depicted object (Alario et al., 2004), which also predicts naming latencies (Barry et al., 1997) but is assumed to reflect difficulty at the earlier stages of visual and conceptual access and not during formulation of words (e.g., Bonin et al., 2002). Name agreement variations, on the other hand, have not been found to affect object-decision (i.e., the ability to indicate whether the depicted picture is a real object or not) response times (e.g., Vitkovitch & Tyrell, 1995) and it is, thus, assumed that their independent influence begins only after visual identification has taken place (Johnson, Paivio & Clark, 1996). In cases where the sources of disagreement is due to the variability of available names for a picture (i.e., *true name agreement*) this variability in alternative names is thought to affect the conceptual and the lexical stages of word production, mostly linked to lexical co-activation and the subsequent level of lexical selection (Alario et al., 2004; Johnson et al., 1996) and less frequently attributed to carrying over of the word-form activation during phonological encoding (Valente et al. 2014).

Low name agreement pictures evoke more potential candidates for the name of the depicted object than high name agreement pictures, and this induces more effortful processing for the speaker, an effect apparent in response times (e.g., Alario et al., 2004; LaGrone & Spieler, 2006). The default explanation of this chronometric effect is grounded in the hypothesis of lexical selection by competition: name agreement should measure the level of co-activation of alternative lexical nodes (i.e., sofa, settee), which creates ongoing conflict between the activated lexical representations, slowing target word selection due to increased co-activation (i.e., couch) (e.g., Levelt et al., 1999;

Roelofs, 1992; 2003). By assessing the performance difficulty and error patterns in populations with semantic and lexical deficits, as well as the neurophysiological effects of naming low name agreement pictures in healthy adults, researchers have, thus, created an "off-label" use of name agreement as a measure of lexical co-activation, where increased variability in names across subjects is used as a measure of the degree of lexical competition within subjects (e.g., Kan & Thompson-Schill, 2004; Novick et al., 2009; Rodríguez-Ferreiro et al., 2009; Bose & Schafer, 2017).

In the neuropsychological literature, older compared to younger adults show larger effects of name agreement and this difference has been interpreted in the framework of lexical competition: aging decreases the ability to resolve the existing competition among the alternative names that lower name agreement pictures evoke (LaGrone & Spieler, 2006). In picture naming in aphasia, lower name agreement is associated with more error-prone word production (i.e., producing less appropriate responses) in some patients (Kremin et al., 2001; Laiacona et al., 2001; Cameron-Jones & Wilshire, 2007) and presents different error patterns between patients and age-matched controls (Bose & Schafer, 2017). Bose and Shafer (2017), for instance, found that patients with aphasia were less prone to name variability in picture naming: typically-developed adults used significantly more alternative names than patients for both high and low name agreement pictures, while patients produced significantly more omissions and fewer visual errors than typically-developed adults in naming low name agreement pictures. This led to a claim that brain damage created an excessive competition in the lexical network, which in high conflict conditions (i.e., naming low name agreement pictures) significantly increased the patients' chance of word retrieval failures. A similar pattern was observed in patients with dementia: name agreement is a strong predictor of overall naming accuracy (i.e., naming a picture with a reasonably appropriate word) in patients with Alzheimer's disease (Harley & Grant, 2004), as they are much more likely to make semantic errors when naming low name agreement pictures and fostered the interpretation that the degradation of the semantic and lexical systems as a

result of cognitive aging increases the inability to resolve competition at the lexical level (Rodríguez-Ferreiro et al., 2009).

In neuroimaging, picture name agreement has been associated with increased left inferior frontal gyrus (LIFG) activity: high conflict trials (i.e., low name agreement pictures), elicited greater Blood-Oxygen-Level-Dependent (BOLD) activation in the LIFG area (Kan & Thompson-Schill, 2004). LIFG is assumed to be involved in conflict resolution, when more than one semantic representations are competing for a response (Thompson-Schill, et al., 1997). This effect has been replicated with healthy individuals, with lower name agreement pictures eliciting increased activity in the left frontal operculum and propelling a lexical selection-by competition interpretation (e.g., Kan, Kable, Van Scoyoc, Chatterjee & Thompson-Schill, 2006). Complementary to that, Novick et al (2009) investigated name agreement variabilities in a patient with restricted damage in LIFG and found that they had performed significantly poorer when naming pictures with low name agreement compared to patients with lesions in other brain areas and healthy controls. These findings were also interpreted in terms of competition: regions of ventrolateral prefrontal cortex which are more active in naming low name agreement pictures have been assumed to index the resolution of the representational competition that arises when more than one candidate response is available in the production system (Novick et al., 2009).

While high-spatial resolution methods, like fMRI, have allowed to identify the brain regions that are sensitive to name agreement variations, researchers have used time-sensitive techniques, like Event-Related Potentials , to more specifically attribute these effects to the process of lexical selection (e.g., Cheng et al., 2010; Shao et al., 2014). In associating observed effects with word production sub-stages, most of the literature derives estimations of the temporal and spatial markers based on Indefrey and Levelt's (2004) proposed timeline. After reviewing a large number of picture naming studies, they proposed an average timeline in which most word production processes take place, following the serial and discrete processing assumptions by Levelt et al.'s (1999) model. According to their proposal, word

production consists of a series of discrete stages and ERP effects directly index the onsets and conclusions of the stages that generate them: conceptual preparation takes place up to 175 ms post-stimulus onset, lexical selection between 175 ms and 250 ms, phonological encoding between 250ms and 330ms, syllabification between 330 ms and 445 ms and self-monitoring from 355 ms until after articulation (see also Indefrey, 2011 for an updated timeline). Indefrey and Levelt's (2004) proposed timeline, thus, provides a commonly used framework for predicting and interpreting ERP effects of name agreement, but the results have not been very consistent: some studies report both earlier and later effects (e.g., Cheng et al., 2010), some attribute name agreement effects solely to the time window of lexical selection (e.g., Shao et al., 2014, see also Johnson et al., 1996; Alario et al., 2004 for the behavioral counterparts), while others find that name agreement modulates only later time windows (Valente et al., 2014).

To my knowledge, Cheng et al. (2010) was the first study to describe the electrophysiological effects of picture name agreement, using ERPs. In their experiment, participants silently named pictures of objects with high and low name agreement, while the electrical activity of their brain was recorded and because of covert naming, it was not possible to record response times behaviorally. They observed differences in brain amplitude by varying name agreement starting as early as 100-150 ms post-stimulus onset (corresponding to the P1 ERP component), continuing in the 250-350 ms window (known as the N2 range). The N2 peaked at 290 ms post-stimulus onset and differences were also found in a later time window of (between 800-900 ms). In relation to the proposed timeline proposed by Indefrey and Levelt (2004), Cheng et al. (2010) suggested that the P1 differences reflect object recognition difficulty for low name agreement pictures, while the N2 effect relates to the window of phonological encoding. More specifically, they suggested that the competition between alternative lexical items for low name agreement pictures was not resolved during the window of lexical selection according to the assumed timeline and for this reason the effect was observed slightly later than expected, questioning the strict discreteness of Levelt et al.'s (1999) model.

A more explicit competition-based interpretation of name agreement effects in ERPs has claimed that lexical selection involves "selective inhibition" of alternative responses, as a specific mechanistic implementation of Levelt et al.'s (1999) selection process. Shao et al. (2014), investigated the electrophysiological effects of picture name agreement for both object and action naming in an overt picture naming task, in which they familiarized participants with the correct names of the pictures before the experiment, while they also recorded participants' naming latencies. To understand how name agreement modulates lexical selection, they primarily focused their analysis on the corresponding time window in line with Indefrey and Levelt's (2004) suggestions and the previously reported N2 component by Cheng et al (2010). They found that name agreement variations had effects both on naming latencies and ERPs: low name agreement pictures took significantly longer to name than pictures with high name agreement and had a more negative ERP amplitude in the N2 window (170-330 ms), with the N2 effect peaking at around 250 ms post-stimulus onset. Additionally, a negative correlation between participants' naming latencies and N2 amplitude was found for action pictures, further strengthening their competition interpretation. Shao et al. (2014) suggested that differences in name agreement reflect differences in the recruitment of selective inhibition, a conflict resolution mechanism which is engaged to reduce competition before lexical selection. Selective inhibition is generally thought to reflect the suppression of alternative responses which are competing with the target in conflict resolution tasks (e.g., Forstmann et al., 2008), while in word production selective inhibition has been suggested to reflect the suppression of incorrect names that are co-activated while speaking (e.g., Shao, Meyer, & Roelofs, 2013; Shao, Roelofs, Martin, & Meyer, 2015; Vromans & Jongman, 2018). Even though existing competitive models of word production had not explicitly specified an inhibitory mechanism (e.g., Levelt et al., 1999; Roelofs, 1992; 2003; 2018), Shao et al. (2014) suggested that such a mechanism can be easily incorporated into a competitive selection. In competitive models, naming pictures with multiple appropriate candidates makes the difference in activation levels between the target word and the competitors relatively small and

therefore the levels of competition higher (Levelt et al., 1999). In that case, a top-down inhibitory mechanism may be recruited to weaken the activation of the competitors, until successful selection takes place (Shao et al. 2013; 2014). In their interpretation, therefore, name agreement is considered not only a measure of lexical co-activation, but also a measure of the requirement of conflict resolution mechanisms that assist a successful selection.

In a more exploratory investigation of the neural underpinnings of word production, which was not limited to analyzing a pre-designed temporal window in a precise topographic location, Valente et al. (2014) reported strong name agreement effects in later processing stages. The authors examined several variables associated with picture naming using a trial-by-trial multiple regression analysis of electrophysiological activity in the entire spectrum after picture onset until participants' responses, while they also used a pre-experimental familiarization phase with the stimuli. Name agreement differences were found in the time window of phonological encoding: between 380-620 ms and 100 ms before articulation. Age of acquisition also had a strong effect on ERPs in those time windows, possibly reflecting differences in retrieval of the word form for later acquired words (see also Laganaro and Perret, 2011; Laganaro et al., 2012). Valente et al. (2014) attributed the effects of name agreement to the variability of names existing for low name agreement pictures, however this variability did not seem to affect the earlier time window in ERPs that Indefrey and Levelt (2004) had associated with the process of lexical selection. Consequently, the authors suggested that the temporal overlap between name agreement and age of acquisition may suggest an interaction between the two at a later word planning stage, in line with cascaded models of word production (e.g., Dell, 1986) and picture naming (e.g., Humphreys, Riddoch, & Quinlan, 1988).

Thus, despite the consistent behavioral associations between name agreement and naming latencies, the effects are much less consistent in ERP research. In interpreting the name agreement effects according to Indefrey & Levelt's (2004) timeline for production processes, the inconsistency in the different time windows in which name agreement effects are reported (e.g., Shao et al., 2014 vs

Cheng et al.. 2010 & Valente et al., 2014), does not allow to draw concrete conclusions about whether differences in name agreement reflect differences in competition during lexical selection, or processing difficulty during lexical co-activation or phonological encoding.

1.5 Conceptual issues in current name agreement studies

Given this *prima facie* use of name agreement as a measure of lexical competition as well as the inconsistencies in the name agreement ERP literature, there remains a need to re-evaluate whether and how picture name agreement affects word production processes by additionally addressing some of the conceptual and methodological issues in studies that have previously used name agreement as a cognitively meaningful predictor of individuals' internal states.

1.5.1 Population-level name agreement and individual-level competition

Name agreement, both in picture naming norms and in word production studies, is a population-level measure, which derives from several independent word retrievals given by a particular pool of participants (Bates et al., 2003; Szekely et al., 2004). This "on-label" use of name agreement is particularly useful when attempting to predict aggregate group behavior in picture naming or replicate the population's modal responses across different languages (e.g., Bates et al., 2003). In cognitive psychology, though, name agreement is also used as an "off-label" predictor of individual's production systems (e.g., Alario et al., 2004; LaGrone & Spieler, 2006; Bose & Schafer, 2017), whereas there is the default assumption that all the available names for a picture (population-level name agreement) are also available within each participant in every word retrieval (individual-level name agreement).

This interpretation of picture name agreement is mostly based on assuming a Luce choice-inspired (Luce, 1959) selection principle guiding word selection and naming latencies: population name agreement for the dominant name in picture naming norms is assumed to index the probability

that an individual speaker will use that word each time they name the same picture. For instance, for the IPNP "obj128dog" ($p_{\text{DominantName}}$= 1) (Szekely et al., 2004), it is a priori assumed that each time a person will name that picture they should always use the word "dog". But what is the theoretical explanation if a limited number of people from the same linguistic population will use the word "labrador" or "mutt" to name that picture? In the cases of pictures with higher name agreement, this assumption may be less problematic, because the probability of deviating from frequencies in norms is relatively low and the theoretical implications for psychological research are not as substantial. In cases, though, of pictures with lower name agreement, like the IPNP "obj472truck" ($p_{\text{DominantName}}$= .59) which is also called lorry ($p_{\text{SecondaryName}}$= .37) (Oppenheim, in prep.), the main assumption in line with a stochastic selection is that *each* time a person names that picture they should have a 59% chance of calling it a "truck" and a 37% chance of calling it a "lorry".

However, this assumption has never actually been empirically tested: this variability between participants' responses is not established to be indicative of the level of lexical conflict that exists within participants' independent word retrievals. In other words, the stochastic axiom of lexical selection is assumed to guide both selection probability and at the same time explain the chronometric effects observed for pictures that activate multiple words across participants, allowing the assumption that variations in picture name agreement reflect the levels of activation of words within *each* participant for *each* independent word retrieval. The alternative explanation of this variability existing for low name agreement pictures is that it instead reflects the heterogeneity across individual speakers' unique and stable word selections: 59% of speakers have an idiosyncratic preferences for the word "truck" and 37% for the word "lorry" and these are stable across naming episodes. In that case, instead of stochastic selection pattern based on name agreement variations, the probability of a particular word to be selected would only be reflective of the weight of this word within the production system of the individual speaker.

1.5.2 Dominant names only? The need to keep track of alternative responses

In picture naming norms and word production studies, usually two measures of name agreement are assessed or reported: the percentage of subjects giving the most common name, and sometimes the information *h*-statistic (e.g., Bates et al., 2003; Szekely et al., 2004; Alario & Ferrand, 1999; Alario et al., 2004). Percentage agreement captures the proportion of the participants who produced the modal name for a picture, while the *h*-statistic is intended to measure the dispersion of the responses provided. The *h*-statistic was originally adopted by Snodgrass & Vanderwart (1980), in an attempt to capture the distribution of names across participants. While percentage agreement increases with increased consensus, increasing *h* values indicate decreasing name agreement. For example, couch and truck both have 50% percentage agreement, but couch can be also named as "sofa", "settee" or even "armchair" by some participants, while truck has only one other alternative name: "lorry"; both concepts will then have equal percentage agreement scores, but the second has a higher *h* value (i.e., less dispersion). Similarly, when the *h*-statistic has a value of 0 and a percentage agreement of 1, it indicates perfect name agreement. These two measures of name agreement are, thus, very highly correlated and often interchangeably used in experiments.

The *h*-statistic is, therefore, an early attempt to account for the influence of alternative names, in addition to the proportion of individuals who select the modal response. Even though the statistic can be informative as an estimate in the population, because it is a population-level measure, it can become misleading for within-subjects name agreement in the same way as percentage name agreement (see 1.5.1). If unique individual differences are a reflection of modal name agreement (percentage agreement), then the distribution of alternative names (*h*-statistic) should vary within each individual accordingly. For instance, if a subject may never consider "sofa" or "settee", then the within-subject *h*-statistic should be 1, despite the .67 value of *h* in population-level norms (e.g., Bates et al., 2003; Szekely et al., 2004).

An alternative way to account for these differences when using percentage name agreement would be to actually monitor secondary and tertiary name responses in norms (cf.Oppenheim, 2017). The majority of the name agreement literature has primarily focused on the dominant names, by using name agreement measures that specifically focus on dominant responses (e.g., percentage dominant name agreement combined with the *h*-statistic), and either restricting analyses of name agreement effects to dominant names or using pre-experiment corrective familiarization to train participants to use each picture's dominant name (e.g., Alario et al., 2004; Shao et al., 2014; Valente et al., 2014). Tracking alternative responses can be useful in evaluating individual differences in name preferences, and at same time they can provide a valuable source of information to test the lexical competition hypothesis. In the same way that researchers use a picture's modal name use (dominant name agreement) to estimate its strength, the use of the second most common name (secondary name agreement), provides a direct measure of the strongest competitor word in the production system, and can, therefore, be additionally evaluated in relation to behavioral (cf.Oppenheim, 2017) or electrophysiological effects.

In the lexical competition hypothesis specifically, secondary name agreement should be a much more meaningful predictor of naming latencies than the *h*-statistic: the levels of activation of the second-best response should modulate response times, in an opposite fashion than that of the target response: i.e., when the levels of activation of the second best option (secondary name agreement) are increased, then lexical competition should also increase and this effect should be evident both behaviorally and electrophysiologically. For instance, a surprising recent finding regarding secondary name agreement came to challenge the basic empirical evidence in favour of competition. In a multiple regression analysis of a large picture naming study which calculated population-level secondary name agreement for the first time, Oppenheim (2017) replicated the dominant name agreement RT effect, but additionally showed that naming latencies for the dominant name (e.g., "couch") were faster when the probability of the secondary name (e.g., "sofa") increased, which contradicts the default

interpretations of name agreement variations as evidence of competitive selection. However, because that study relied on population-level measures of dominant and secondary name agreement, its interpretation is vulnerable to one of the same criticisms that I already noted for studies that interpret dominant name agreement effects as evidence for competition: if name agreement measures only index variability between speakers, instead of conflict within speakers, then lexical selection may not actually require deciding between two strong candidate responses. Thus, to interpret effects of dominant or secondary name agreement, it is essential to show that variability between speakers actually predicts variability (and hence conflict) within speakers.

1.5.3 Pre-experimental familiarization is problematic

In picture naming studies-even in studies which independently evaluate the effects of picture name agreement-it has become very common to include a pre-experiment procedure to visually familiarize the participants with the stimuli and/or assign the "correct" labels for each picture (e.g., Alario et al., 2004; Shao et al., 2014; Valente et al., 2014). The familiarization phase can either be a separate naming session, as in Alario et al. (2004), or an unrecorded pre-experimental phase, as in Shao et al. (2014) and Valente et al. (2014), and it is mostly intended to eliminate variability in naming latencies for subsequent data analysis or in early visual processing differences in ERP studies. While there appears to be consistency between the behavioral effects reported during the familiarization phase and the actual experiment (Alario et al., 2004), such familiarization also introduces an additional task demand that could change or confound subsequent investigations. In picture-word interference studies, for instance, it has been reported that the typical semantic interference effect appeared only after familiarization (Gauvin, Jonen, Choi, McMahon, & de Zubicaray, 2018) or that familiarization can even reverse the polarity of the interference effect to facilitation (Collina, Tabossi, & De Simone, 2013). Although there is not sufficient evidence that familiarization significantly modulates the semantic interference effect in general, it is noteworthy that the vast majority of picture-word

interference studies include familiarization in their design (143 out of the 161 as reviewed in Bürki et al., 2020).

However, the strong reasoning to argue against the use of a pre-experimental familiarization phase derives from its conventional use in picture name agreement studies, because familiarization can violently interfere with lexical preferences. For instance, the norming study of Alario et al. (2004) consisted of two phases: in Experiment 1 participants named each picture using a label of their choice and were later given the preferred label to be memorized for Experiment 2, which repeated the naming task with the instruction to only use the labels the experimenter assigned as correct responses. In addition, the two overt ERP name agreement studies, by Shao et al. (2014) and Valente et al. (2014), both used an unrecorded familiarization phase during which they introduced the correct names for the pictures. As with Alario et al., in Shao et al. (2014) familiarization was corrective: participants initially named the pictures with the label of their choice and were then immediately corrected by the experimenter in cases where they used non-dominant names. However, in line with the alternative hypothesis that population-level name agreement may also reflect stable individual preferences even for non-dominant names, it is reasonable to assume that any instruction to use specific names could potentially overwrite such preferences. In that case, it is unclear whether some of the reported effects in Shao et al. (2014) could stem from actual differences in competition levels for low and high name agreement items or an externally-induced experimental confound introduced during familiarization, which disproportionally affects non-dominant responses and induces an additional conflict for pictures of low name agreement.

The discrepancy between the timeline in which name agreement effects are reported in Shao et al. (2014) and Valente et al. (2014) could also be explained by the different procedures they followed during the familiarization phase. Valente et al. (2014) did not include a production task during familiarization and the instruction to use the preferred names was given to participants via a booklet. In cases where participants produced the label of their choice offline and were immediately corrected

for them, (as in Shao et al., 2014) this could be considered an initial, separate naming session (similar to the repeated naming used in Alario et al., 2004). Since the behavioral and electrophysiological effects during the initial naming were not recorded in Shao et al. (2014), I could only remain speculative about the expected findings in free naming and their relation to those reported by Valente et al. (2014). Nevertheless, it is reasonable to assume at least some priming effects emerging from repetition and visual exposure to the stimuli (see Cave, 1997; Alario et al., 2004), as well as an additional level of conflict arising as a result of the task demands (see Nozari & Hepner, 2019), which should be carefully evaluated in relation to hypothesis testing.

1.5.4 The lack of consistent electrophysiological markers of picture name agreement

The majority of the behavioral studies which report name agreement effects assume that they reflect increased difficulty during lexical selection, which according to Indefrey and Levelt (2004) takes place between 175 ms and 250 ms post-picture onset. However, out of the three ERP naming studies, only Shao et al.'s effects (2014) were clearly fit within the expected time window for lexical selection: after scaling their estimated naming latencies according to the timeline, they claimed that lexical selection in their experiment took place 250–344 ms, while name agreement effects peaked at 290 ms. Cheng et al. (2010) used covert naming and could not propose a temporal window of selection, while Valente et al. (2014) did not explicitly associate name agreement effects with lexical selection. It is, therefore, possible that either name agreement effects do not solely occur within the window of lexical selection in a traditionally serial framework, that the strict seriality of the timeline by Indefrey and Levelt (2004) cannot account for the variability observed in most production studies (Nozari & Pinet, 2020), or both.

The discrepancies of consistent ERP effects could also stem from the inconsistency in the methodologies used: in traditional categorical (Cheng et al., 2010; Shao et al., 2014) uses of the variable, the effects are analyzed in respect to their peaks in a pre-specified time windows, while a

continuous use of the measure is usually combined with state-of the art statistical methodologies in ERPs (Valente et al. 2014). While the traditional use of analysis of variance in ERPs is mostly used when attempting to associate observed effects with cognitive processes, the use of mixed effects modeling has allowed researchers to account for stimulus and participant effects, as well as analyze the entire temporal and spatial spectrum of the variance (see Bürki, Frossard & Renaud, 2018 for a review of the methodology in ERPs and 2.2.4 for a detailed statistical description of the approach), while at the same time evaluate parameter estimations as in regression, instead of simple p-values, which can become misleading and over-interpreted (Hubbard & Lindsay, 2008). Valente et al. (2014) adopted this approach by additionally analyzing both stimulus and response-locked ERPs and found name agreement effects arising at a later temporal window, not previously reported. However, despite the conservative and disciplined approach they adopted in their analysis, the inclusion of familiarization could have potentially unwillingly obscured some of the effects as discussed above (see 1.5.3.).

All things considered, it is clear that there is a lack of a stable electrophysiological pattern of name agreement effects, which could help guide interpretations in relation to within-subjects lexical co-activation or competition. Even though behavioral studies consistently show name agreement to be a strong predictor of naming latencies, it is also necessary to investigate how behavioral effects would change by omitting the process of familiarization or by adopting more naturalistic designs (e.g., Oppenheim, 2017; cf. Oppenheim & Balatsou, 2019), like collecting picture naming norms. Nevertheless, by distinguishing between what name agreement variations actually estimate and the attempts to particularly try to associate these effects with the lexical competition hypothesis, it is possible to rightfully use this measure as both an empirically validated variable in picture naming tasks and a cognitively meaningful predictor of individuals' internal states.

1.6 Thesis aims

The overall aim of the current dissertation is to evaluate the use of *picture name agreement* as a predictor of lexical co-activation in word production, while addressing the aforementioned conceptual issues and methodological inconsistencies in the previous literature. The broader scope of the present research is to understand how communicative language production works and present empirical findings that contribute to the debate on the nature of the lexical selection mechanism (i.e., competitive versus non-competitive hypotheses), using picture naming studies with behavioral and electrophysiological measurements. The experimental approach followed here is based on simple picture naming tasks, which I believe most closely resemble the processes that engage in typical word production. In an attempt to avoid using "infrequent derailments of the process", which Levelt et al., (1999; p.2) used to characterize the speech errors literature, I instead suggest that the widely-used picture-word interference studies may be the "infrequent derailments" of word production and adopt a simplified and purified approach in my investigations.

Two of the three empirical chapters in the current thesis (Chapters 4 & 5) combine traditional behavioral measures (i.e., analyzing responses and naming latencies) with ERPs. The main motivation for the use of ERPs in word production is to further understand the cognitive processes and mechanisms that take place before speakers articulate a single word. The temporal resolution of ERPs is very high, on a par with the time-course of neural firing rates in the brain and they can thus shed light onto phases of information processing from the onset of a stimulus all the way to response execution and even after that. ERP research is in general technically challenging and especially in the field of language production, due to the motor movements that introduce large artefacts, significantly limiting its use in the field compared to other disciplines in cognitive science, such as language comprehension. The current work aims to bridge this gap with the use of advanced methodological techniques, which are not limited to specific temporal windows or spatial localization of the effects (see Chapter 5) and at the same time address crucial theoretical questions in more naturalistic experimental designs.

In evaluating picture name agreement, I try to bridge the gap between what it actually measures (i.e., population-level responses) and how it is typically interpreted in the framework of a stochastic competitive lexical selection (i.e., as a measure of within-speaker lexical competition). The first objective of the current thesis is to resolve this discrepancy by assessing, for the first time, the psychological reality of picture name agreement within the individual speaker. By evaluating whether individuals' choices in naming are more accurately predicted by their previous word selections (a tendency termed as *idiolects*), or by the word selections of other speakers in their linguistic communities, I wish to examine whether name agreement measures this co-activation of words.

A secondary objective of the current work is to evaluate the flexibility of the lexical selection mechanism in handling exogenous and endogenous lexical conflict, in reference and contra to individuals' idiolects. In an attempt to bridge strictly competitive and non-competitive models and driven by the hypothesis that lexical competition can situationally emerge as competitive according to task demands (Nozari & Hepner, 2019), I aim to study whether experimentally-induced competition directly impacts the integral part of the selection mechanism or instead affects ad hoc processing stages that are not inherent to selection. By studying the distinct effects of name agreement (endogenous lexical conflict) and name change (exogenous lexical conflict), I hope to distinguish between the processes that are integral to word production (like that of lexical co-activation) and those that are not obligatory (like that of response competition).

A third objective of the experimental work is to explore the temporal and spatial distribution of name agreement effects in simple picture naming and the extent to which these variations are indexing increased co-activation or difficulty of selection. While previous work has reported name agreement effects in naturalistic designs (e.g., Valente et al., 2014), the extent to which variation of names affect selection of words has not been investigated in reference to response consistency. By investigating both dominant and secondary name agreement variations that emerge in repeated simple picture naming and how they are modulated by repetition priming for items with consistency in naming

(i.e., the same name was used in both sessions), I aim to evaluate how and when lexical co-activation is modulated by distinct properties of words in the mental lexicon, indexed by name agreement.

In sum, the current thesis aims to contribute a substantial body of experimental work into one of the most widely used and yet often misinterpreted variables in word production research: *picture name agreement*. By evaluating the availability of responses that speakers have for production and consequently their influence during lexical selection, I aim to contribute to our understanding of the processes that underlie single word retrieval.

CHAPTER 2-Methodological Considerations

In this thesis, I examine the behavioral and neural effects of name agreement in simple picture naming tasks, in which I measure the participants' responses, naming latencies and electrophysiological activity. In this chapter, I describe the principal methods used in the following experiments, including techniques for measuring and analyzing responses, naming latencies and ERPs.

2.1 Mixed-effects modeling in the analysis of behavioral measures

In psycholinguistic experiments one of the most common dependent measures is response time: the interval between the onset of a stimulus and the onset of the participants' response. The influence of the independent variables, which can either be a property of the experimental design (e.g., stimulus repetition, experimental condition, etc.), or the stimulus (e.g., a picture's name agreement or visual complexity, etc.), is usually estimated by aggregating each participant's mean response time per condition and conducting a statistical test (e.g., t-test, ANOVA, regression) to assess significance. By convention, significance permits researchers to generalize a result to a population of participants, but only considering the stimuli used in the study (F1 ANOVA). A complementary tactic is by-stimulus analysis, in which response times are aggregated over participants for each stimulus (F2 ANOVA), but although this allows generalization of the findings to the "population" of the kind of stimuli used in the study, it fails to simultaneously offer generalization to the population of participants (Clark, 1973).

This analytical approach is usually followed in experimental psychology because the group of stimuli and participants used are selected amongst many other alternative samples as a representative subset of a population of interested, i.e., some undergraduate students of the Psychology department are chosen as a subset of the entire population of undergraduate students of Psychology. By using statistical tests in which the response times are averaged either across participants or stimuli, it is uncertain whether the observed effects are attributed to the influence of the independent variables on

response times, or to the unique properties of the subsets of the participants (e.g., a subset of participants could be significantly slower than the larger population) or stimuli (e.g., a subset of stimuli could be significantly harder to process than others) used in experiments. Such uncertainties increase the likelihood to falsely reject the null hypothesis (Type I error) and can have wider implications when considering the general replicability crisis in psychology (e.g., Pashler & Harris, 2012).

Although the use of ANOVAs has been the default statistical approach in psychological research, in recent years it has been widely advocated to replace it with simple and linear regressions. Some main advantages of regression over ANOVAS are: (1) they include parameter estimation in addition to p-values, (2) they can include continuous predictors instead of simple factors, (3) they are able to incorporate crossed random effects, 4) they can account for the statistical variance of binary variables as well (McCulloch, Searle, Neuhaus, 2008; Boisgontier & Cheval, 2016). Thus, a way to empirically distinguish between the influence of the properties of participants and stimuli and the actual influence of the dependent variables on response times is to use regression, and in particular mixed-effects models, in which participants and stimuli are treated as random effects, while the dependent variables are considered non-random, i.e., fixed effects. While fixed effects estimate how the mean of the independent variable is influenced by one or more dependent variables, with random effects it is possible to estimate how the statistical variability for both the dependent variable (random intercepts) and the independent variable (random slopes) affect the statistical result, by estimating the association of the repeated measures. This is particularly helpful for response time analyses, because as with every continuous measure, some participants tend to be faster than others and some stimuli in different conditions elicit faster responses than others.

Treating stimuli or participants as random effects is widely used in response time analysis, because it allows the inclusion of the random slopes into the statistical model and increases the robustness of inference, even with missing data. In response times, researchers often exclude non-correct responses or data above or below a threshold as not being realistically expected from the design.

Especially when missing data are not missing at random (i.e., some stimuli elicit faster processing time), fitting variables that predict the distribution of the data into the model increases the accuracy of its results (Baayen, Davidson, & Bates, 2008). An advantage of using mixed effects models in response time analysis is that, unlike with simple regression or ANCOVA, it is possible to additionally estimate the influence of covariates in the statistical result, which can either be stimulus or participant-dependent, thus strengthening the predictability of the model. Apart from continuous data analysis (i.e., reaction times), it is also possible to fit random effects in the analysis of binary data (i.e. responses), by using mixed-effects logistic regression.

In the current research, mixed effects-effects models are used to analyze both continuous and binary data. The analyses are performed in R Studio (R Studio Team, 2020) using the lme4 v.12 library (Bates, Maechler, Bolker, & Walker, 2016). In Chapter 3, I use generalized mixed-effects logistic regression to analyze participants' responses (binary data) in a picture name experiment using the glmer::binomial function in lme4. In Chapters 4 & 5, mixed effects linear regressions are used to analyze the participants' naming latencies in the respective picture naming experiments using the lmer::gaussian function in lme4. All confirmatory analyses (i.e., testing pre-planned hypotheses with significance estimations) include a maximal random effects structure for stimuli and participants in accordance with each experimental design, by additionally respecting the levels of predictor variables, such as experimental Session. Maximal random effects are used conservatively and with the awareness that the models may suffer minimal power loss, but eventually incorporated to generalize their results better (Barr, Levy, Scheepers & Tily, 2013).

2.2 The Event-Related Potentials (ERPs) technique

Electroencephalography (EEG) is a physiological method that records the electrical activity of the brain from a number of electrodes (usually 32, 64, or 128) placed on the scalp. The first human EEG was recorded by Hans Berger in 1924 and today this technique is widely used by clinicians,

mostly in epilepsy diagnosis (Tatum, 2014), and researchers, primarily in the form of Event-Related potentials (ERPs). EEG records the summation of the excitatory and inhibitory synchronous activity of groups of neurons that have a similar spatial orientation close to the cortical surface (Niedermeyer, & da Silva, 2005). For reasons to do with cancellation of currents oriented in the opposite direction and current diffusion through various layers of biological tissue such as dura, bone and skin, EEG has a poor spatial resolution compared to other techniques that measure indices of brain activity, such as fMRI. However, the temporal resolution of EEG is very high, allowing to track variations in current density at a sub-millisecond timescale.

2.2.1 Principles of Electroencephalography (EEG)

The electrodes used for EEG recording are usually set in an elastic cap and placed over the scalp according to the 10-20 convention (Sharbrough, 1991), an internationally recognized method of placement, in order to ensure standardized methods procedures. EEG electrodes can either be passive or active. In active systems, the electrodes contain a pre-amplification module which amplifies the signal between the skin and electrode. In passive systems, after placing a cap on the participant's head, each electrode site is degreased using 70 degrees alcohol and conductive gel is applied to serve as a conductive bridge between the scalp and each electrode (impedance reduction procedure). An amplifying device records the difference in electrical potential between each electrode and the electrode of reference, which ideally should have zero potential and can either be a virtual (i.e., average) reference, deriving from the linearly combined signal of the recordings from all the electrodes, or real (i.e., mastoid) electrodes. The recording system includes an amplifier which is responsible for converting the analog signal (continuous over time) as recorded from the electrodes into a digital signal (discrete in time) that can be processed by a computer. Before it is processed further, the EEG signal needs to be filtered to attenuate artefacts produced by biological movement and organs different from the brain (e.g., a heartbeat) and electrical artefacts from the environment of

testing. Filtering of the EEG signal can either be performed online or offline and the most common types include high pass (keeping high frequencies), low pass (keeping low frequencies), band pass (keeping frequencies within a certain range) and notch filtering (eliminating a particular frequency, such as 50Hz), which is generally not advised to use, because it can potentially severely distort the signal (e.g., Leske & Dalal, 2019). The EEG signal recorded does not only reflect brain activity, but also a number of other sources of bioelectrical current such as eye movements, muscle and heart activity, which have to be reduced or discarded. Such artefacts are usually corrected using the regression analysis (e.g., Di Flumeri, Aricó, Borghini, Colosimo & Babiloni, 2016) or the Independent Component Analysis (ICA; e.g., Jung et al.1998), or by completely removing the contaminated EEG epochs, either manually or automatically.

2.2.2 From EEG to ERPs

In the field of cognitive neuroscience, the raw EEG signal recorded form the scalp is not usually directly informative of cognitive processing. One of the most common ways in which EEG datasets are analyzed is through computation of Event-Related Potentials (ERPs) (see Figure 1). An ERP is the

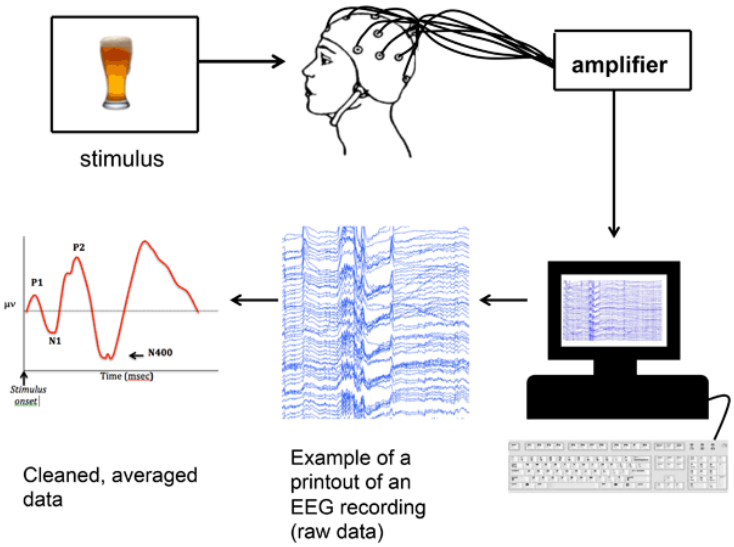

Figure 1. Schematic representation of a typical ERP recording setting. Reprinted from Beres (2017). Time is of the essence: A review of electroencephalography (EEG) and event-related brain potentials (ERPs) in language research. *Applied psychophysiology and biofeedback*, 42(4), 247-255)

average electrophysiological response of the brain in relation to a series of cognitively meaningful visual, auditory or tactile stimuli or in relation to a spontaneous response of the participants to it (Luck, 2005). An advantage of ERPs is that, unlike behavioral measures that represent information processing at one point in time (that of response registration for instance), they allow us to observe brain activity variations accompanying the stream of cognitive operations that take place from the onset of stimulus exposure until even after a behavioral response is registered. Participants' brain responses to a stimulus of interest are aggregated over multiple trials and waveforms corresponding to different conditions are computed and compared, consisting of peaks with specific latencies and shape. The necessary steps to transform a raw electrophysiological signal into an ERP include (not necessarily in linear order):

- Time-locking: before any data processing, the presentation of each stimulus must be locked to particular markers in the EEG signal (stimulus-locked ERP). It is also possible to simultaneously or independently mark the behavioral response that is generated by the participant (response-locked ERP).
- Filtering: The main goal of filtering the EEG data is to attenuate the deviant frequency of signal that is most likely due to external noise or artefact contamination. For instance, the meaningful frequencies in a typical ERP waveform are between 1 and 10 Hz, and by applying a low-pass filter to remove frequencies above 30 Hz it is ensured that most movement artefacts or line noise will not significantly contaminate the ERP waveform. In contrast, by filtering frequencies lower than approximately 0.1 Hz researchers ensure that most slow waves, due to static electricity in the participants' body surface, will be removed.
- Artefact detection, correction or rejection: Artefacts such as those caused by eyeblinks, eye movements, muscle activity and skin potentials can be either manually dismissed when examining single trials (artefact rejection) or automatically corrected in the entire raw EEG signal (artefact correction). The most common type of artefacts is that relating to eye blinks, which are usually monitored in the EEG recording by placing electrodes around

the eye (above, below or to the side), can be automatically reduced with the use of a regression analysis. In regression analysis, ocular artefacts are modelled over a number of occurrences and when variation between events is low enough (*SD* < .005 at all electrode sites), the artefact is corrected using a propagation factor that is scaled according to the average electrical activity of the scalp (e.g., Hoffmann & Falkenstein, 2008). A different way to deal with ocular artefacts is to perform Independent Component Analysis (ICA), in which the pure eye activity in the signal is isolated following decomposition of the signal into independent components and discarded before recomposition (e.g., Vigário, 1997).

• Epoching and baseline correction: The continuous data are segmented into single-trial EEG epochs, usually starting 200 ms before stimulus presentation. During baseline correction, the voltage measured in the pre-stimulus window (e.g., -200 to 0 ms for stimulus-locked ERPs) is subtracted from the ERP waveform, in order to ensure that the signal reflects the differential evoked potential elicited relative to the pre-stimulus activity of the brain.

• Re-referencing: During re-referencing, the signal of the new reference is subtracted from each EEG channel recorded. Re-referencing can be done with respect to mastoid, average or nasal references. Although all different types of re-referencing should not theoretically impact the quality of the data, it has been shown that there are notable differences for each referencing strategy used (Lei & Liao, 2017), and therefore the re-referencing type should be chosen with caution and in relation to practices established in the literature.

• Averaging: Single-trial EEG epochs with clean and preprocessed data are averaged together to create ERP waveforms for each subject in every experimental condition, which are later aggregated over the entire pool of participants to create grand averages in each condition. Difference waves can be computed between conditions of interest and peak amplitudes or peak latencies can be measured and statistically analyzed as in the case of other quantitative data.

ERP peaks indexing specific cognitive functions are often called ERP components and are defined by their polarity (positive or negative going voltage), timing, scalp distribution, and sensitivity to task manipulations. Some of the most common ERP components in language research include the N400 (Kutas and Hillyard, 1980), The N200 (Schmitt, Münte, & Kutas, 2000), the Error-related Negativity (Falkenstein Gehring et al., 2018), the P600 (Osterhout & Holcomb, 1992) and the Left Anterior Negativity (Friederici, Hahne, & Mecklinger, 1996). ERP components are assumed to index cognitive processes (Luck 2014), but their utilities and interpretation have been historically challenged (e.g., Donchin, Ritter & McCallum, 1978), since components observed at the group-level not do not always hold at individual subject or trial levels (e.g., Gaspar, Rousselet, & Pernet, 2011; Rousselet et al., 2011; Luck, 2014). ERP peaks can be easily over-interpreted as components, so it is advised to be conservative when describing and interpreting components, unless the experimental task is explicitly manipulating a cognitive function which has been previously associated with a component in the literature (e.g., as with the N400 in semantic violations). Current research questions the conventional focus on ERP components and instead advocates the use of more advanced methodologies in order to link electrophysiological responses with cognitive functions (Bridwell et al., 2018).

2.2.3 Independent Component Analysis (ICA) of ERPs

Because ERPs are extremely sensitive to motor artefacts relating to speech (Luck, 2005) the literature on word production does not abound with ERP studies investigating the cognitive processes immediately preceding articulation. Most studies usually involve indirect measures, such as naming latencies, to understand the processes that take place before articulation, because experimental work that combines ERPs with speaking needs to correct or reject the most common artefacts in ERPs (such as eye blinks and line noise) in addition to those induced by motor movements related to speech.

In order to successfully isolate and remove complex artefacts, recent EEG studies tend to rely on Independent Component Analysis (ICA; Makeig et al., 1999; Jung et al., 1998; 2000) instead of the

traditional regression methods (e.g., Gratton, Coles & Donchin, 1983), because the latter can only hope to reduce artefacts that manifest in a repetitive fashion (such as eyeblinks, jaw movements or electrocardiogram). ICA is a computational method that separates the subcomponents of a multivariate signal and can reconstruct it after excluding some of them (Hyvärinen, 2013). ICA results in a set of independent components (ICs) which contain qualitatively similar activity to EEG, but which are separated based on their idiosyncratic frequency properties. Once deconstructed into ICs, components relating to artefacts or unrelated to brain activity can be identified and removed before the signal is re-constructed into electrode-array data (see Figure 2).

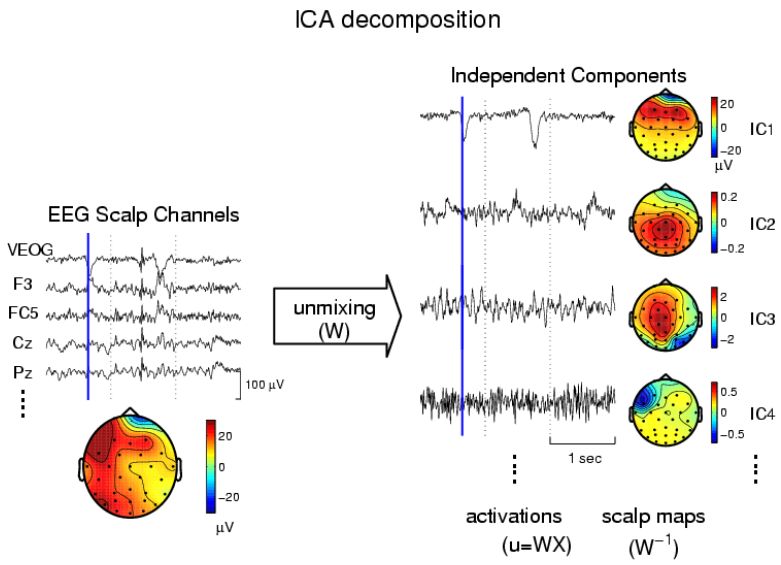

Figure 2. Illustration of the Independent Component Analysis (ICA). Figure reprinted from https://sccn.ucsd.edu/~jung/Site/EEG_artifact_removal.html. The left side of the figure represents the raw EEG recording data and the right side of the figure the ICs corresponding to ocular and noise artefacts (IC 1 & 4 respectively) and the actual electrophysiological activity of the brain (IC 2 & 3).

The analysis follows the principles of linear decomposition and could be characterized as a rotated Principle Component Analysis (PCA) (Wold, Esbensen & Geladi, 1987) that maximizes the statistical independence of the components instead of detecting them according to their maximal variance. ICA is based on the assumptions that (1) the temporal activity recorded consists of independent sources, (2) the propagation delays of the sources are negligible, (3) the summation of the signal from different sources is linear at all electrodes, and (4) the number of independent signal

sources is the same as the number of electrodes (i.e., if the signal is recorded from 64 electrodes, then the Independent Components should also be 64; Makeig, Bell, Jung & Sejnowski, 1996; see Figure 3.). Because the data are decomposed into spatially fixed and temporally distinct information sources, the ICs projected on scalp maps indicate the synchronous activity of both the brain and the artefacts as participants perform a cognitive task. In ERP studies, therefore, the components most likely resemble the neural activities in the areas where they are generated (Ku et al., 2007).

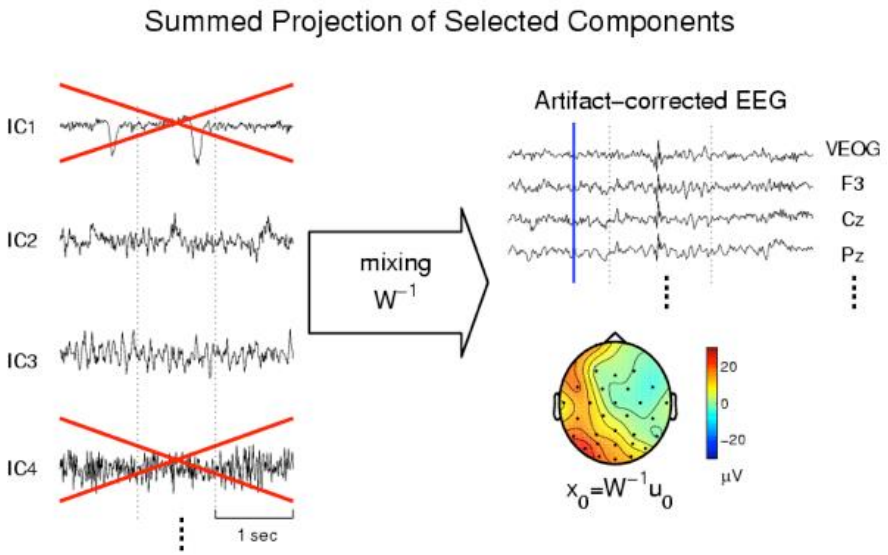

Figure 3. Illustration of the Independent Component Analysis (ICA). Figure reprinted from https://sccn.ucsd.edu/~jung/Site/EEG_artifact_removal.html. On the left side of the figure the ICs corresponding to ocular artefacts and noisy data are removed and the signal is reconstructed without them in the artifact-corrected EEG (right side).

2.2.4 Mass univariate analysis of ERPs

Given the nature of ERPs (i.e., averaged EEG epochs over trials by participant per condition), the traditional approach for statistical analysis is to select peaks or ERP mean amplitudes across trials and perform an analysis of variance or a t-test in the different conditions of interest (e.g., Barkley, Kluender, & Kutas, 2015; Strijkers, Costa, & Thierry, 2010). This widely used approach, though, has three important limitations: (1) it limits ERP analyses across temporal and spatial dimensions that are usually focused on a given component of interest, and determining such dimensions appropriately can be challenging (e.g., Alonso-Prieto et al., 2015), (2) as with classic ANOVA-type analysis of behavioral data (see 2.1.) this approach fails to take both stimulus and participant variances into

account, since the analysis is performed on averaged data (Bürki et al., 2018), and (3) an additional challenge in electrophysiological data analysis is that it involves multi-dimensional data (i.e., a large number of electrodes each providing a large number of individual measurements per unit of time, in the range of 1000 per second across many trials) and therefore researchers are more likely to face statistical issues in relation to false positives and the multiple comparisons problem (Miller, 1981). These issues, apart from eliciting or raising significant challenges in the statistical analysis of ERP studies, may additionally contribute to the replicability crisis in science (Open Science Collaboration, 2015), due to the already complex nature of ERP data analysis (Bürki et al., 2018).

These issues can partially be addressed by implementing a more sophisticated statistical approach that does not primarily rely on the evaluation of average trends. Pernet, Chauveau, Gaspar, ans Rousselet (2011) proposed a two-level mass univariate analysis approach with cluster mass permutation tests, implemented in their LIMO EEG toolbox. In a first level analysis in LIMO, each individual dataset is independently analyzed using a general linear model approach across all time and space dimensions of the EEG signal, producing a beta coefficient. In a second level analysis, the estimated parameters are compared across participants using robust statistical tests (i.e., t-test, ANOVA, regression) and significance is estimated with the cluster mass method (see Figure 4.) Cluster-based analysis in ERPs is a nonparametric statistical test which is performed by grouping together the neighboring variables (*t* or *F* values of electrodes for instance) into clusters (see Maris & Oostenveld, 2007 for a detailed description of this approach). Although LIMO does not directly model stimulus and participants as independent random effects (i.e., stimuli remain nested under participants), it is a powerful statistical tool which emphasizes effect size whilst correcting for multiple comparisons. Some of the robust statistical tests in LIMO in the second level of analysis are the following:

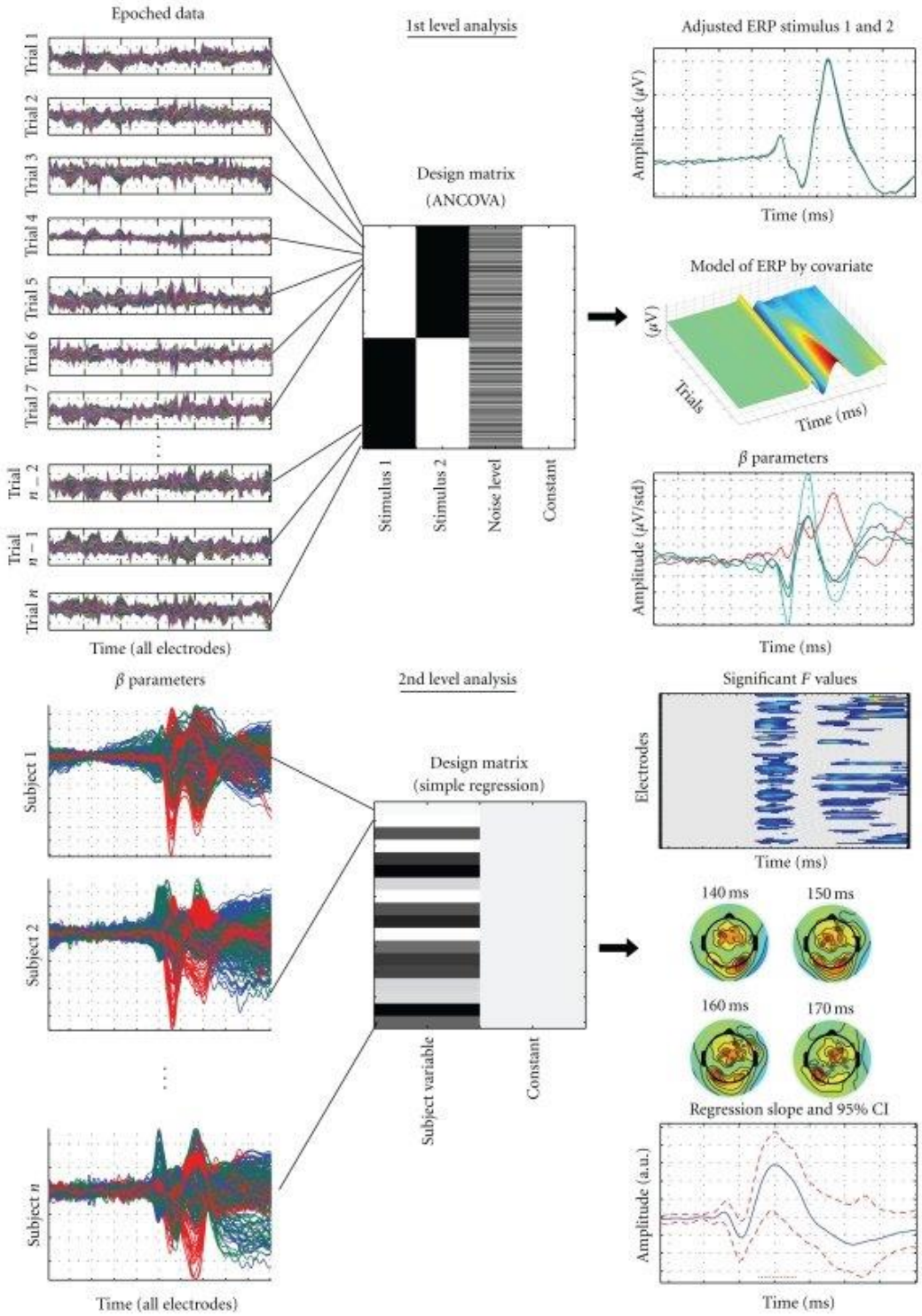

Figure 4. Illustration of the LIMO analysis of ERPs. Reprinted from Pernet, Latinus, Nichols & Rousselet, 2015. Cluster-based computational methods for mass univariate analyses of event-related brain potentials/fields: A simulation study. *Journal of neuroscience methods*, 250, 85-9.

• One sample t-test, which examines the covariation of single-trial ERPs with the use of the bootstrap-*t* approach, in which participants are selected randomly with replacement.

Using cluster statistics, the data is centered, and participants are resampled, while each *t* value provides an approximation of the *t* distribution under the null hypothesis and is used to estimate the quantiles. The *p* values are calculated according to the average times the *t* values fall below or above the quantiles.

• Paired t-tests in LIMO are performed based on a percentile bootstrap in which subjects from each group are sampled with replacement, to maintain within-subject variance. Contrasts are tested under the alternative hypothesis and the *p* value is the smallest value obtained by averaged times the difference between groups is above zero or one minus this average.

• Repeated measures ANOVAs are estimated from *F*-values for all conditions which are centered independently, and the *F* distribution is estimated under the null hypothesis. Correction for multiple comparisons is performed using cluster mass statistics with the use of bootstrapped *F* values. *P* values are obtained by observing how many times the observed *F* values are above the *F* statistic.

2.2.5 ERPs and language production

Due to the technical challenges associated with motor artefacts induced by speaking, ERPs are not as widely used in the study of language production as in other areas of neurolinguistics, such as language comprehension. However, an advantage to use ERPs in production is that, unlike comprehension, the effects can be further evaluated in relation to the quality and quantity of the speakers' linguistic output: instead of associating components and cognitive functions based on participants' performance in passive tasks (i.e. button-pressing responses in sentence processing), language production ERP research allows to relate the observed electrophysiological patterns with the actual linguistic response generated by speakers. Despite the relatively smaller amount electrophysiological studies in the field, the ERP language production literature has provided significant insights into the architectures and mechanisms that underlie human speech. For instance,

ERPs in production have revealed the ultra-fast processing of the brain before articulation: the most common variables in typical word production studies (e.g., visual complexity, object familiarity, name agreement, lexical frequency) have been found to affect electrophysiological activity as early as 100 ms post picture onset and until 100 ms before articulation (Valente et al., 2014). Similarly, variance in non-linguistic cues (e.g., pictures) affects the earliest time windows in picture naming tasks, indexed by the P1/N1 component modulations (e.g., Abdel Rahman & Sommer, 2008; Cheng et al., 2010), as in any other task that recruits the visual system (e.g., Luck et al., 1994).

Most ERP studies in word production involve a picture naming task with experimental manipulations targeting various cognitive processes, such as lexical selection or phonological encoding. Some of the most frequently mentioned ERP components affected by such manipulations in word production studies are (but are not limited to): the N200 (e.g., Costa et al., 2009; Shao et al., 2014), the P200 (e.g., Aristei et al., 2011; Strijkers et al., 2010; Python, Fargier & Laganaro, 2018), the P300 (e.g., Costa et al., 2009), the N400 (Blackford, Holcomb, Grainger & Kuperberg, 2012; Shitova, Roelofs, Schriefers, Bastiaansen, & Schoffelen, 2017; Piai, Roelofs & van der Meij, 2012), and later effects closer to the time range of articulation (Rose & Abdel Rahman, 2017; Rose, Aristei, Melinger & Abdel Rahman, 2019; Janssen, Hernández-Cabrera, van der Meij & Barber, 2015; Valente et al., 2014; for a recent review on lexical selection see Nozari & Pinet, 2020).

In terms of the mechanisms underlying semantic-to-lexical access, phonological form retrieval, or articulatory preparation, most ERP studies of picture naming have manipulated naming difficulty and reported modulations in time windows ranging from 200 ms post stimulus onset until later stages of pre-articulation period (e.g., Costa et al., 2009; Valente et al., 2014). For instance, increased semantic interference in the cumulative interference task has been shown to produce increased ERP amplitudes in the window of the P300 component preceded by a decrease in N200 amplitude (Costa et al., 2009) or an increase in P200 amplitude (Python et al., 2018) followed by decreased positivity in later time windows (Janssen et al., 2015). Other semantic manipulations, such as those in picture-word

interference studies, generally elicit increased N400 amplitude (Piai et al., 2012; Shitova et al., 2017; Blackford et al.,2012; Dell'Acqua et al., 2010; Rose et al., 2019; Wong et al., 2017), as well as later modulations in the time window immediately preceding articulation (e.g., Dhooge, De Baene & Hartsuiker, 2013; Rose et al., 2019). These effects have often been interpreted in the framework of Indefrey and Levelt's (2004) proposed timeline of word production processes and particularly tied to naming difficulty associated with increased competition, however the timeline and its normative use as a framework for interpretation have since been questioned in the literature (e.g., Nozari & Pinet, 2020; Munding et al., 2016).

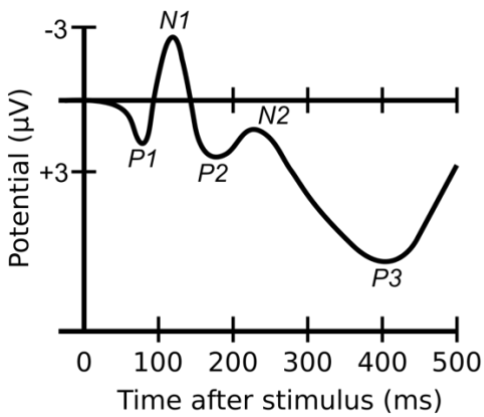

Figure 5. Illustration of Typical ERP peaks: P1, N2, P2, N2. Reprinted from Luck, 2005. *An Introduction to the Event-related Potential Technique*. MIT Press.

The field is less clear for the electrophysiological components reported in picture name agreement studies, since the number of published studies is limited and the methodological and conceptual inconsistencies between studies have rendered the identification of a stable ERP marker of simple production processes quite challenging (see 1.5.4). An additional challenge is that some studies have attempted to identify and label specific components (Cheng et al., 2010; Shao et al., 2014), while others have analyzed effects over a continuous time-window without specifically indexing components (Valente et al., 2014). Studies investigating differences elicited by high and low picture name agreement often report both early and late effects (Cheng et al., 2010; Valente et al., 2014), but the

majority of ERP modulations are observed in windows associated with semantic-to-lexical mapping processes according to the estimates of Indefrey and Levelt (2004): lower name agreement pictures are associated with increased N200 amplitude between 250 and 350 ms (Cheng et al., 2010; Shao et al., 2014), or later effects starting at around 400 ms post picture onset and lasting until the pre-articulation period (Valente et al., 2014).

2.2.6 A brief overview of the N200 component

The N200 is a negative ERP component usually salient between 200 and 350 ms post-stimulus onset over anterior regions of the scalp (see Figure 5.), which has been associated with the recruitment of cognitive control mechanisms (Luck, 2005). The N200 was first reported by Sutton, Braren, and Zubin (1965) in a stimulus uncertainty task, where it was followed by an enhanced P300 for cues with reduced sensory modality in a visual processing (i.e., switching of paired cues with either clicks or light flashes). The N200 component is generally enhanced for infrequent trials or deviant trials and, as an early peak, it is thought to be indicative of the stimulus categorization process (Luck, 2005). The N200 is commonly reported in tasks that involve response inhibition or inhibitory control, like the Eriksen flanker task, in which participants are asked to suppress an incongruent response (Heil, Osman, Wiegelmann, Rolke & Henninghausen, 2000), or the go/no-go task, in which participants are asked to abort response activity for specific trials, even after response initiation. In those tasks, increased N200 amplitude for the deviant conditions is assumed to index the inhibition of a response.

In neurolinguistics, N200 modulations have been reported in the case of language switching (e.g., Jackson et al, 2004; Verhoef, Roelofs & Chwilla, 2009), object naming (e.g., Schmitt, Münte & Kutas, 2000; Abdel Rahman, van Turennout & Levelt, 2003; Shao et al., 2014; Cheng et al., 2010) and selective access to language representations in bilinguals (e.g., Rodriguez-Fornells et al., 2002; 2005; 2006; Balaguer, Sebastián-Gallés, & Rodríguez-Fornells, 2005), most of which have also been attributed to suppression of a response. In language production specifically, the N200 has been reported both in simple overt picture naming (Abdel Rahman et al., 2003; Costa et al., 2009; Shao et

al., 2014) and silent picture naming (Schmitt et al., 2000; Cheng et al., 2010). With the exception of Costa et al. (2009) who found that increased semantic interference led to less negative N200 amplitude in cumulative semantic interference, in simple picture naming, interpretations of the N200 have also been associated with naming difficulty due to inhibition of non-target responses (e.g., Schmitt et al., 2000; Abdel Rahman et al., 2003; Cheng et al., 2010; Shao et al., 2013; 2014; 2015; Vromans & Jongman, 2018). Although simple production should not require any explicit need to suppress a response, in the ERP name agreement literature, increased N200 amplitudes for lower name agreement pictures were suggested to index increased competition between the existing labels before the production of the target word (Cheng et al., 2010; Shao et al., 2014).

2.2.7 A brief overview of the N400 component

One of the most widely studied ERP components in language research is the N400, a negative-going deflection that peaks at around 400 ms post-stimulus onset and is widely regarded as an index of semantic violation or semantic processing depth (Kutas & Federmeier, 2011). The N400 was first reported by Kutas and Hillyard (1980) in the processing of semantically incongruent sentences, eliciting more enhanced negativity for sentences like: “He took a sip from the transmitter” compared to “He took a sip from the waterfall”. Since then, the N400 has become a mainstream semantic index in neurolinguistic studies, as well as in studies investigating semantic memory (e.g., Holcomb et al., 1999), recognition memory (e.g., Smith & Guster, 1993), and attention (e.g., Deacon & Shelley-Tremblay 2000, for a review see Kutas & Federmeier, 2011). The N400 has also been found to be sensitive to the lexical properties of single words, since it is increased for the processing of pseudowords (e.g., Friedrich, Eulitz & Lahiri 2006), low-frequency words (e.g., Barber et al., 2004) and words with a smaller orthographic neighborhood (e.g., Holcomb, Grainger, & O'rourke, 2002).

In language production in general, the N400 is seldom explicitly reported, possibly superseded or masked by large potential deflection elicited by articulatory preparation, while in word production

specifically, it has mainly been reported in studies of picture-word interference (Blackford et al.2012; Piai et al., 2012; Shitova et al., 2017). In the cumulative semantic interference study of Costa et al. (2009), who also reported N200 modulations, there was a significant effect in the N400 window associated with increased semantic similarity, but the effect was not cumulative as it was for the N200 (i.e., the N400 did not increase with increased ordinal position within each block, as the N200 did). In picture word interference studies, semantically related distractors elicit enhanced N400 amplitude compared to unrelated distractors, which has been interpreted either as increased lexical co-activation and consequently as evidence for the competition hypothesis (Piai et al., 2012; Shitova et al., 2017) or merely as increased lexical co-activation (Nozari & Pinet, 2020). In contrast, in word-target picture naming, the N400 has been reduced for pictures preceded by semantically related versus unrelated words, while this priming effect was dissociated with the behavioral interference effect (Blackford et al., 2012). A possible explanation is that the N400 reflects spreading activation between two different sources of information, that is pictures and words, and therefore facilitates the production based on semantic similarity. Although these effects likely originate from semantic-level processes, it remains uncertain whether the N400 in language production is the counterpart of the N400 observed in language comprehension (Nozari & Pinet, 2020), due to the relatively small presence of the component in the relevant literature.

CHAPTER 3- The psychological reality of picture name agreement

This chapter investigates the psychological validity of picture name agreement within the individual speaker. Using repeated naming, this is the first study to date to evaluate response consistency within individuals in relation to their previous word choices and name switching according to population-level norms for the dominant and secondary responses. We demonstrate the validity of picture name agreement as a cognitively meaningful predictor of individuals' states, but additionally show that speakers' previous selections also guide their future word preferences.

This chapter is accepted under revision for publication from: Balatsou, E., Fischer Baum, S., & Oppenheim, G. M. (in revision). The psychological reality of picture name agreement. *Cognition.*

Abstract
Picture name agreement is one of the most commonly used measures in language production. Beyond measuring population-level tendencies, researchers often assume that name agreement also indexes cognitive processes that occur within individuals. For instance, if picture naming norms show that 50% of speakers name a picture as couch, then each time a person tries to name the picture, they should have a 50% chance of selecting couch. An alternative, however, is that name agreement may simply reflect population-level sampling of more stable individual preferences, developed through experience (i.e., 50% of speakers prefer the name couch). One way to distinguish between these possibilities – and assess the psychological reality of name agreement – is simply to re-norm pictures with the same individuals. In this study, we therefore collected timed naming norms for a large set of line drawings from the same 25 native British English speakers twice, 1-2 weeks apart. Results show participants' name choices in Session 2 are jointly predicted by population-level name agreement, from our previous norms, and individuals' own productions in Session 1. This is the first direct demonstration that picture name agreement has some psychological validity, but also reveals that it does not directly index within-subject lexical competition as previously assumed.

Keywords: name agreement, picture naming, word production, lexical competition, idiolects

3. 1 Introduction

Picture naming is one for the simplest and most commonly used tasks in the study of language production, and one of the strongest and most consistent predictors of picture naming speed and success is a picture's name agreement (e.g., Lachman et al., 1974; Vitkovitch & Tyrell, 1995; Alario et al., 2004). Name agreement is an empirically derived measure of the proportion of speakers who independently produce the picture's modal name when asked to name it. When most participants in a norming study give the same name for a picture, it is said to have high name agreement; when few produce even the most common name, it is said to have low name agreement. Thus, name agreement estimates from picture naming norms naturally extend to predicting how new participants from the same population should name the same stimuli: if 50 out of 50 participants named a picture as "dog" in previous norms, then the picture will most likely elicit "dog" responses from the next 50 participants. When selecting materials for new experiments, responsible researchers therefore consult norms to ensure that most participants will generate their desired names of their own volition; this is the classic 'on-label' use of name agreement.

3.2. Name agreement as a predictor of individual-level cognitive processes

However, in recent decades, an "off-label" use of name agreement has also become quite common. From early on, researchers noted that pictures with high name agreement tended to be named faster than those with low agreement, independent of other word-level attributes, such as word frequency or image familiarity (Lachman et al., 1974; Lachman & Lachman, 1980; Vitkovitch & Tyrell, 1995; Alario et al., 2004). Early studies of picture naming latencies reported robust effects of age of acquisition and lexical frequency (e.g., Butterfield & Butterfield, 1977, Carroll & White, 1973; Oldfield & Wingfield, 1965), but population-level name agreement, sometimes described as codability, soon proved an even stronger predictor (Gilhooly & Gilhooly, 1979; Lachman, 1973; Lachman & Lachman, 1980; Lachman, Shaffer, & Hennrikus, 1974). This basic chronometric effect

has been replicated in many languages (Bates et al., 2003), including American and British English (Snodgrass and Yuditsky, 1996; Ellis and Morrison, 1998; Szekely et al., 2004), Spanish (Cuetos et al., 1999), French (Bonin et al., 2002), Italian (Dell'Acqua et al., 2000), Greek (Dimitropoulou et al., 2009), Japanese (Nishimoto et al., 2012) and Persian (Bakhtiar et al., 2013), inviting speculation about cognitive processes that might underlie it. The most common explanation is that low name agreement pictures induce some form of challenge within the individual speaker, since they must decide which of the available names to use for that picture, with this indeterminacy resulting in longer naming latencies (Barry et al., 1997; Bates et al., 2003; Lachman et al., 1974; Paivio et al., 1989; Snodgrass & Yuditsky, 1996; Vitkovitch & Tyrrell, 1995; Weekes et al., 2007). Such speculation marks a subtle but important shift from the "on-label" use of name agreement to predict aggregate group behavior to an "off-label" use of predicting within-individual cognitive processes.

Perhaps inspired by such robust effects in norms, researchers have stopped merely controlling for name agreement and instead begun specifically manipulating it as a way to investigate a range of cognitive functions, directly related to language production or not. For instance, picture name agreement has been associated with dissociations between semantic and episodic memory performance (Lachman & Lachman, 1980; Mitchell 1989), phonological encoding (LaGrone & Spieler, 2006) and repetition priming in picture naming tasks in both children and adults (Lorsbach & Morris, 1991; Mitchell & Brown, 1988). Similarly, studies that manipulated picture name agreement in clinical populations have associated higher name agreement with greater naming accuracy. For instance, studies with Alzheimer's disease patients have found that name agreement is one of the strongest predictors in their naming performance (Harley & Grant, 2004; Rodríguez-Ferreiro et al., 2009), and patients with aphasia appear especially error-prone when naming low name agreement pictures, compared to matched controls, prompting an interpretation that they have greater difficulty selecting among alternatives (Laiacona et al., 2001; Kremin et al., 2001; Cameron-Jones & Wilshire, 2007; Bose & Schafer, 2017).

Name agreement manipulations have also been used assess the cognitive processes and neural substrates of word production (e.g., Indefrey and Levelt, 2004; Indefrey, 2011). For example, greater Left Inferior Frontal Gyrus (LIFG) activity when naming low compared to high name agreement pictures has been interpreted as evidence that LIFG mediates selection among competing alternatives in production (Kan & Thompson-Schill, 2004; Thompson-Schill, et al., 1997). Similarly, electrophysiological differences between high and low name agreement pictures in the N200 range have been interpreted as reflecting the recruitment of response inhibition, as a mechanism to suppress the competing alternative names for low name agreement pictures (Cheng et al., 2010; Shao et al., 2014).

The main theoretical premise behind such interpretations is that name agreement is specifically tied to lexical selection, reflecting the extent to which individual speakers consider alternative lexical responses before selecting a word (e.g., Indefrey & Levelt, 2004; LaGrone & Spieler, 2006; Bose & Schafer, 2017). Competition in production refers to the idea that the co-activation of alternative words (i.e., sofa) slows the selection of a target word (i.e., couch) as a result of ongoing conflict between the activated lexical representations (Levelt et al.,1999; Roelofs, 1992; 2003; Howard et al., 2006). Although the effects of name agreement in word production appear robust, there remains active debate about whether many effects cited as support for lexical competitive actually require a competitive mechanism for lexical selection[1].

Thus, researchers typically interpret name agreement as evidence for competitive lexical selection specifically, and more generally assume that a picture's name agreement describes the distribution of options available to each individual when attempting to name a picture. Therefore, effects of name agreement are assumed to directly reflect the processes that occur within each

[1] Non-competitive models of word production argue that empirical evidence for lexical competition can be explained in other ways, such as postlexically at a prearticulatory response buffer stage (Mahon et al., 2007), or as 'competitive' incremental learning (Oppenheim et al., 2010).

individual (e.g., Bates et al., 2003). According to the competition narrative, for instance, naming a picture of a dog imposes no difficulty, because no other names exist or compete for selection. In contrast, naming a picture of a couch, which can also be named as sofa or settee, is assumed to impose great difficulty, because each individual should consider the additional names identified by picture naming norms from other members of their linguistic community. The basic problem with this narrative, which the current study aims to address, is that name agreement is an empirical measure of group-level tendencies, prima facie unsuited for use as a predictor of individual-level cognitive processes. Thus, such uses and interpretations of name agreement make four major assumptions about the nature of individual-level lexical selections:

1. An individual's likelihood of choosing any word is a stochastic function of its activation in their mind when they try to choose. As illustrated in the Luce Choice rule (Luce, 1959), the probability of selecting a word is assumed to be determined by the ratio of its activation to that of any alternatives (e.g., Levelt et al., 1999). Such a stochastic word selection function is common to most models of production (e.g., Oppenheim et al., 2010), and in competitive production models it is further used to explain the time required to select a word as a function of the level of its activation and that of its competitors (Levelt et al., 1999; Roelofs, 1992; 2003; Roelofs & Piai, 2015).

2. Each individual considers the range of possible responses observed in their larger linguistic community. If picture naming norms show that speakers use both 'couch' and 'sofa' to name a picture of an upholstered multi-person seating object, then each time an individual speaker tries to name the picture, they should sample from these responses. Similarly, if norms indicate a range of 15 possible responses to a picture of an electric can opener, then a competitive interpretation of this 'number of names' effect (e.g., Szekely et al., 2003) must assume that each speaker considers the full range of observed responses, or at least a representative subset.

3. Group-level norms index the relative activation, and therefore retrieval probability, of each option within each individual. Population-level norms identify not only the range of options that each

individual will consider but also the probability of an individual selecting each option. If relevant norms indicate that half of all participants named a given picture as 'couch', then Speaker A should have a 50% probability of selecting 'couch', Speaker B should have a 50% probability of selecting 'couch', and so on.

4. Each retrieval is independent of previous retrievals. Although not obvious, this point follows from the assumption that group-level norms predict individual-level cognitive processes, especially when assessing name agreement effects in repeated naming paradigms or those where researchers pre-train participants to use particular names (Mitchell & Brown, 1988; Alario et al., 2004; Valente et al., 2014; Piai & Roelofs, 2013). Moreover, relaxing this assumption quickly erodes the assumed links between population-derived norms and individuals' cognitive processes.

While most of these assumptions seem quite plausible, it is worth asking what other factors or cognitive processes might give rise to name agreement measures and thus name agreement effects. Returning to the actual method of estimating name agreement – asking n individuals to name the same picture – one possibility is that name agreement measures simply reflect a process of sampling stable individual preferences. In the couch/sofa example, it is easy to imagine how an individual speaker might develop a persistent bias to choose one option, never actually considering the alternative. For instance, researchers have detected repetition priming in picture naming up to 48 weeks after initial exposure (Cave,1997), shown that repetition priming is stronger for lower name agreement pictures (Park & Gabrieli, 1995), and confirmed that word-specific aspects of such priming persist for at least one week (Francis & Sáenz, 2007; see Francis, 2014, for a review). Although such persistent priming has typically been assessed in terms of decreases in naming latencies, rather than increases in the likelihood of selecting a particular name, a recent model of word production argues that both outcomes can result from continual, experience-driven adjustments in semantic-to-lexical mappings (Oppenheim et al., 2010): each time a speaker retrieves a word for production, an incremental learning process adjusts that mapping, increasing the ease and likelihood of retrieving the target again and decreasing

the ease and likelihood of retrieving activated alternatives, thus providing momentum to select and reinforce the same target in the future. All else equal, such adjustments should accumulate into speaker-specific tendencies to use particular words: *idiolects*. Low name agreement in picture naming norms, then, may simply reflect heterogeneity in individual speakers' word preferences or idiolects, not the extent to which individuals consider alternatives. Under this alternative proposal, the best predictor of whether an individual will choose "couch" or "sofa" should not be name agreement estimates from population-level norms, but instead their own past behavior.

3.3 The current study

Thus, it is not obvious that name agreement should predict individual level competition, because it actually measures between-subjects linguistic variation. More generally, as a (between subjects) population-level measure, it is unclear whether name agreement is even a psychologically valid predictor of (within subjects) cognitive processes associated with name uncertainty. Although it is possible that the between-subjects variation that is measured by picture naming norms does indeed reflect the range and relative strengths of the names that each individual considers (we term this "the Luce choice account"), it is also possible that the between-subjects variation that is measured by picture naming norms simply reflects between-subjects variation (we term this "the idiolect account").

Because traditional norming studies ask individuals to name a set of pictures just once, they cannot distinguish between these possibilities.[2] We can however distinguish between them -and finally assess the psychological reality of name agreement- by simply examining individuals' name selection consistency across two naming sessions. If population-level name agreement effectively predicts the options available to each individual, in line with our Luce choice account and the way that the researchers typically use name agreement, then whether a person uses a particular name to describe a

[2] Although Alario et al. (2004) reported a broadly similar two-session norming task, they did not and could not examine within-speaker name consistency because they followed each Session 1 response with a desired name for participants to use in Session 2.

picture (i.e., couch) in the second session should depend on its population-level contingent probability, regardless of their selection in the previous session. In the couch and sofa example, a speaker should have a 50% chance to select couch each time they name the picture, regardless of whether they previously selected sofa. However, if name agreement instead reflects more stable between-subject variation, in line with our alternative 'idiolect' account, then a person should simply repeat their initial word selection when renaming a picture, regardless of its contingent probability in the population-level norms.

3.4. Methods

3.4.1 Summary

The basic methodology followed the standard IPNP norming procedures (Szekely et al., 2003), except that each participant named the full picture set twice, one to two weeks apart (Mean: 8.6 days, *SD* = 3.3).

3.4.2 Participants

Twenty-five Bangor University students (18 female, Mean age : 21.3 years, *SD* = 5.1) participated in exchange for course credit. One participant was replaced due to technical problems. All reported British English as their native language, normal or corrected-to-normal vision and hearing, and no known language disorders. None had participated in Oppenheim's (in prep.) previous norming study. The study was approved by Bangor University Ethics Committee and participants received course credit or cash compensation.

3.4.3 Materials, apparatus and procedure

Pictures for the naming task were the 525 black-and-white line drawings of common objects from the International Picture Naming Project (Bates et al.,2003). As in previous applications, these

were grouped into 5 blocks of 105 pictures each, including one filler at the beginning of each block, followed by 104 experimental items. Twenty-five unique sequences approximately counterbalanced stimulus orders across sessions and subjects. Pictures were presented via PsychoPy2 (v1.83.01) on a 17" CRT in a soundproof testing booth at the Bangor Language Production Laboratory. Responses were recorded via a headmounted microphone, feeding into both a digital recorder and a custom-built delayed- threshold voicekey. In each approximately 30-minute session, the participant was seated in front of the computer monitor and asked to quickly and accurately name each picture as it appeared. Each trial began with a small black fixation cross at the center of the screen for 50 ms. Next, a picture (422 x 422 pixels) appeared at the center of the screen for 3000 ms or until the voicekey triggered. Short self-paced rests followed each 105-trial block. One to two weeks later, the participant returned to repeat the full procedure.

3.4.4 Analytical approach

Responses were initially transcribed on-line, and later confirmed via audio recordings. Oppenheim's (in prep.) recent norms from the same population provided dominant and secondary names for each picture. Following those norms, responses that deviated from an expected name only in plurality or the addition of an article (e.g., "toe"/"toes", "boat"/"a boat") were accepted as tokens of that name; possible abbreviated forms (e.g., plane and aeroplane), however, were considered distinct responses. In cases where a participant produced two or more codable responses in a single trial (e.g., "dog... cat"), we analyzed the first.

Statistical analyses apply confirmatory logistic mixed effects regression, via the glmer::binomial function in the lme4 v1.12 library (Bates, Maechler, Bolker, & Walker, 2016) in R v5.5.1 (R Development Core Team, 2016). All fixed effects are centered and contrast coded. All models also include maximal random effects structures (Barr, Levy, Scheepers, & Tily, 2013) for

participants and items, omitting correlations between random effects to facilitate convergence. P-value estimations use the Wald approximation method.

3.5. Results

Excluding 899 trials (3.4%) in which a voicekey error ended the trial early (< 300ms post stimulus onset) leaves 25351 total picture naming attempts for our analyses (12644 in the first session and 12707 in the second session), summarized in Table 1.

Table 1: Response frequencies and mean naming latencies for each session. Note that naming latencies are calculated after excluding any trials with audible hesitations for the dominant and secondary responses. Mean RTs dominant and secondary names are provided, for comparison with Oppenheim's (in prep.) recent norms.

	Current Experiment								**Oppenheim's (in prep) norms**			
	Responses				**Latencies**				**Responses**		**Latencies**	
	Session 1		**Session 2**		**Session 1**		**Session 2**					
	Mean	**N**	**Mean**	**N**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**N**	**Mean**	**SD**
Dominant	.81	10249	.82	10443	988	354	961	328	.78	-	978	217
Secondary	.10	1293	.10	1273	1165	433	1149	421	.10	-	1125	399
Other	.08	1018	.07	950	-	-	-	-	-	-	-	-
Omissions	.006	84	.003	41	-	-	-	-	-	-	-	-
Total	12644		12707		1045	404	1012	376				

3.5.1 Population-level name agreement

Population-level name agreement for this experiment was directly compared to that of Oppenheim's (in prep.) To set the stage, we can consider correspondence between the frequencies of dominant names in the current experiment and those reported in recent norms from the same population (Oppenheim, in prep). By-item response frequencies in Session 1 corresponded well to recent estimates of both their dominant name agreement (by-item Pearson's correlation between dominant name frequency in Oppenheim, in prep, and Session 1 of the current experiment: $r = .90$, $p < .001$) and secondary name agreement (excluding 65 items without a secondary name: $r = .86$; $p < .001$). Such by-item correspondences also remained in Session 2, for both the dominant name ($r = .88$, $p < .001$) and the secondary name ($r = .83$, $p < .001$). By-item response frequencies also correlated well between

Session 1 and 2 within this experiment, for both dominant (r= .90, p < .001) and secondary (r= .86, p < .001) name agreement. Thus, considered at the population level, name selections were consistent with previous norms and appear relatively stable across sessions.

3.5.2 Individual-level name agreement

But we can also ask whether the same individuals tended to use the same names across sessions. For instance, Table 1 indicates that 81% of participants named items using their dominant names in Session 1. If this proportion simply reflects a sampling of individuals and their preferred names -81% of our participants happened to prefer these pictures' dominant names, as described in our "idiolect" account -then we would expect that *the same* 81% should use these dominant names in the second session. Thus, the probability of a person using the dominant name in both sessions would be, simply, .81. On the other hand, if they were stochastically selecting among responses each time, as described in our 'Luce choice' account, then only 81% of the original 81% should use the dominant name in both sessions. Thus, the probability of a person using the dominant name in both sessions would be .81ˆ2=.66.

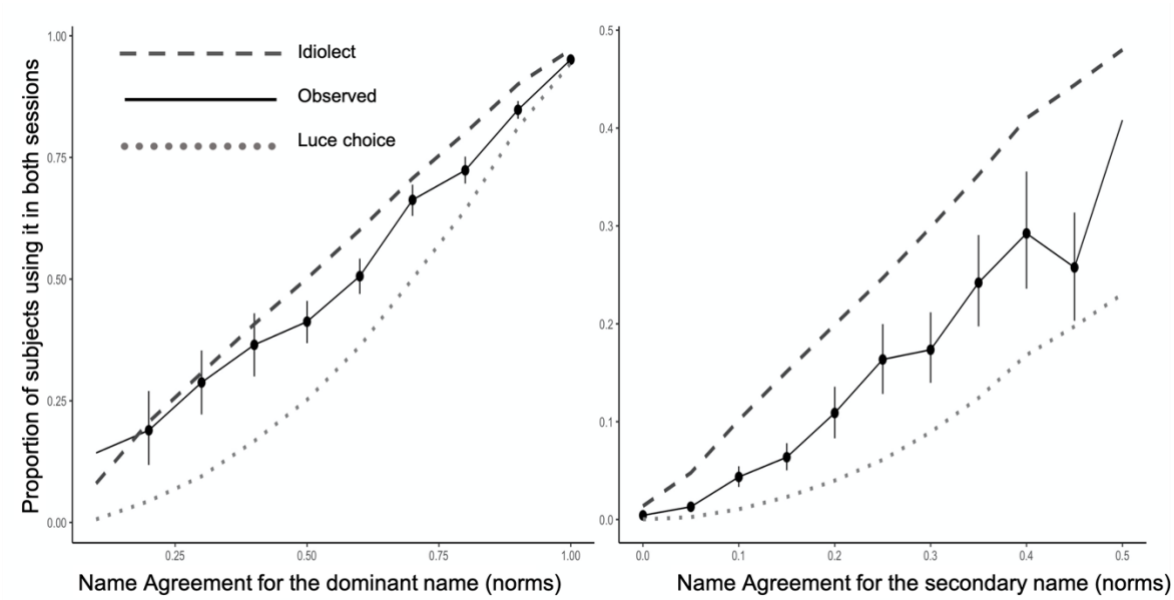

Figure 1: logistic mixed effects regressions modeling the likelihood of participants using the same dominant name 1b) and secondary name (1b) in both sessions. The y axis represents the proportion of subjects who used the dominant (1a) and

secondary name (1b) twice and the x axis represents the population-level name agreement for the dominant (1a) and secondary names (1b) from Oppenheim's (in prep.) Bangor ipnp norms.

As described in the Methods section, we used maximal logistic mixed effects regression, to predict participants' likelihood of producing a picture's dominant name in Session 2 as a function of (1) its population-level name agreement from Oppenheim's (in prep) recent Bangor norming study (a continuous measure from 0:1, centered); and (2) whether the individual participant produced the dominant name in Session 1 (a binary measure {0,1}, centered). To estimate random slopes within items, we excluded the 117 items for which every participant produced the dominant name in Session 1, leaving 408 items and 9806 trials for this analysis.

First considering our Luce choice account, if between-subjects measures of dominant name agreement predict the within-subjects strength of a dominant response, as researchers typically assume, then participants should be more likely to produce the dominant name for a picture with higher name agreement, compared to one with lower name agreement, independent of their prior behavior. Confirming this prediction, participants in our experiment were significantly more likely to use the dominant name in Session 2 for high name agreement pictures than for low name agreement pictures, regardless of whether they themselves had produced that name previously (odds ratio: 68.75:1, $\beta_{\text{DominantNameAgreement}} = 4.23$, $SE = .21$, $p < .001$).

Now considering our alternative idiolect account, if participants develop and maintain persistent name preferences, then their likelihood of producing the dominant name for a picture should specifically depend on their having chosen the dominant name in the past. Confirming this prediction, participants here were also significantly more likely to name a picture in Session 2 using its dominant name if they had previously done so in Session 1 than if they had previously given another name instead (odds ratio: 10.84:1 , $\beta_{\text{UsedDominantInSession1}} = 2.38$, $SE = .11$, $p < .001$). Thus, we find support for both for the traditional Luce choice account of name agreement measures, and also for our novel

idiolect account: population-level name agreement and individual's previous word selections jointly predict their likelihood of selecting a dominant name in the second session (see Figure 1a).

Until now, our narrative has focused on name stability, but a stronger test of the idea that name agreement predicts within-speaker response conflict may come from specifically examining cases where a speaker switched responses across sessions. Assuming that a picture can elicit multiple acceptable responses, the Luce Choice account predicts that speakers should be more likely to switch to a stronger dominant name than to a weaker dominant name. Confirming this prediction, fitting the above model to a relevant subset of the data showed that participants were significantly more likely to switch from a secondary name in Session 1 to a dominant name in Session 2 for pictures with high name agreement than for those with lower name agreement (odds ratio: 41.35:1 , $\beta_{\text{DominantNameAgreement}}$= 3.72, SE = .44, p < .001).

According to both accounts, these effects should also hold for non-dominant names. If the distribution of responses across the population predicts the strength of these options within each individual, then speakers should be also more likely to select stronger secondary names. Similarly, if speakers develop preferences even for non-dominant names in the first naming session are they more likely to select the same secondary responses when naming again later? To address this question, we repeated the previous logistic regression analysis but instead focused on secondary names, thus estimating the likelihood a participant producing a picture's secondary name in Session 2 as a function of (1) its population-level *secondary* name agreement from Oppenheim's (in prep) recent Bangor norming study (a continuous measure from 0:1, centered); and (2) whether the individual participant produced the secondary name in Session 1. To estimate effects within items, we further excluded 216 items that no participant had named using the secondary name in Session 1; this leaves 309 items and 7435 trials for the current analysis. Replicating our results for dominant name use, whether participants selected the secondary name during the Session 2 was predicted by both the population's frequency of using the secondary name from our previous norms (odds ratio: 645:1, $\beta_{\text{SecondaryNameAgreement}}$= 2.38 ,

SE= .15, $p < .001$) and participants' own productions in Session 1 (odds ratio: 10:1 , $\beta_{\text{UsedSecondaryInSession1}}$= 6.46, *SE*= .48 $p < .001$). Thus, we can broaden the scope of our previous conclusion: speakers are also more likely to produce more commonly used secondary names and secondary names that they themselves have chosen in the past (see Figure 1b).

As considered previously, the Luce Choice account makes particularly strong predictions about the likelihood of name switches. If a population's use of a secondary name predicts its strength within the individual speaker, then individuals should be more likely to switch from a dominant to a stronger secondary response in the second session. This is a particularly important prediction to test with secondary names because one interpretation of the pattern in switches to dominant names is simply that speakers gradually switch to more appropriate or 'correct' responses. Confirming this prediction, participants were also significantly more likely to switch from a dominant name in Session 1 to a stronger than a weaker secondary name in Session 2 (odds ratio: 8.22:1, $\beta_{\text{SecondaryNameAgreement}}$= 9.01, *SE*= .62, $p < .001$). Thus, this finding strengthens our claim that population-level name agreement can predict response conflict within individuals, even in cases where people switch away from dominant names.

3.5.3 Monte Carlo analysis of name consistency across sessions

The analyses so far consider only the two most common names for each picture. Although these account for 87% of all responses in the experiment, it is worth considering whether the name stability trend that we have observed might extend to other responses as well. For instance, in recent norms, "stove" accounted for 14% of all responses to a picture of an oven ("oven": 43%; "cooker": 34%; "stove": 14%; Oppenheim, in prep). If such minority name selections reflect robust individual differences, rather than transient noise, then we should expect participants to repeat such names at rates above chance. Thus, to more generally assess how well participants' names corresponded across the two sessions, we also ran a Monte Carlo simulation to incorporate *all* codable responses.

For this analysis, we assessed how often participants' names corresponded between Session 1 and Session 2 and compared that proportion to what would be expected by chance, that is, under the assumptions of the Luce Choice account. We focused this analysis on trials in which a codable response was produced in both Session 1 and Session 2, for a total of 12.524 trials across all participants. Of these trials, we found that that participants in the current experiment produced an identical response in Session 1 and Session 2 10.593 times .

We then used Monte Carlo techniques to simulate the production of a codable response in Session 2, using the norms from Oppenheim (in prep), that is, from an independent sample of participants. For example, for a picture of an oven, the Monte Carlo procedure randomly selected a Session 2 response from among the codable responses produced in Oppenheim (in prep), 43% of the time selecting the word "oven", 34% of the time selecting the word "cooker" and 14% of the time selecting word "stove." In a single run of the Monte Carlo analysis, this random selection of a Session 2 response was carried out for all 12.524 trials, and the randomly selected Session 2 response was compared to the actual Session 1 response, to estimate the proportion of trials in which the name would be expected to correspond by chance. This Monte Carlo procedure was carried out 1.000 times, to provide a distribution of chance values that could be compared to the observed name consistency between Session 1 and Session 2. The results of this analysis are shown in Figure 2.

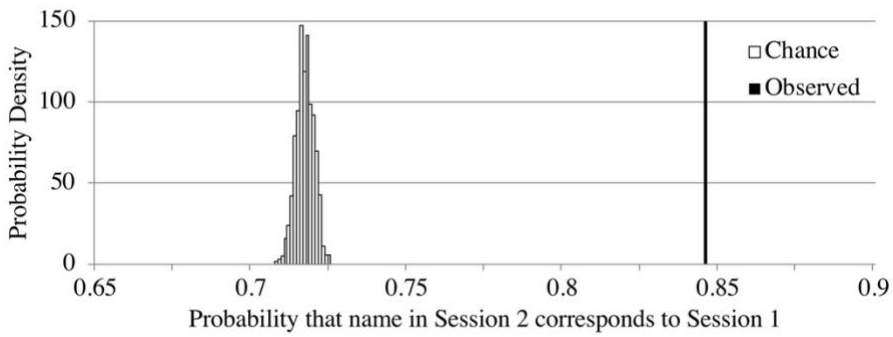

Figure 2: Results of the Monte Carlo analysis comparing the observed proportion of trials in which the name in Session 2 corresponded to name produced in Session 1 (black bar), compared to a distribution of the results of the Monte Carlo analysis that estimated the proportion expected by chance (white bars).

On average, our analysis estimated that only 71.7% of Session 2 responses would be expected to correspond to the Session 1 name by chance. The distribution of these chance values was narrow, with 95% of the runs of the Monte Carlo analysis falling between 71.1% and 72.7%. None of these values came near the observed proportion of the 84.6%, indicating that participants were repeating the same responses more often than would be expected by chance ($p < .001$). Note that although many of these repetitions were cases in which the response in both Session 1 and Session 2 was the dominant name (e.g., "zebra"-"zebra"), the sample also includes cases in which the response in both Session 1 and Session 2 was less probable (e.g., saying "zeppelin" for a picture of a blimp twice, despite the fact that only 11% of the independent sample produced that response to that picture). These results provide further evidence against a strong version of the Luce Choice account, as they suggest that participants are more likely to produce the same idiosyncratic responses across sessions than would be expected by chance.

3.6 Discussion

Picture name agreement has long been associated with robust behavioral, neuroimaging, and electrophysiological effects in word production tasks. Current literature generally assumes that measures of picture name agreement reflect the distributions of possible names that are active and available to each individual speaker each time they attempt to name a picture. However, this interpretation is primarily based on untested speculations about the processes that underlie word production tasks, rather than direct empirical evidence. In this study we used a repeated picture naming task to assess the within-subjects psychological reality of picture name agreement, specifically examining whether population-derived norms predict naming behavior beyond what one would expect from a sampling of idiolects.

A first point is that our group-level correlations support the validity of picture name agreement for its on-label use, i.e., predicting variance in item names for a population as a whole. The frequencies

of the most commonly and the second most commonly used names in both Session 1 and Session 2 corresponded well with those reported in recent norms with the same population (Oppenheim, in prep.), and with each other. Thus, norms have a predictive power for estimating the distribution of names with different subjects of the same linguistic community, which supports their use in selecting materials for other experiments.

However, the main purpose of our investigation was to assess picture name consistency within individuals, and its relationship with norms from their larger linguistic community. To our knowledge, this is the first systematic investigation of picture name consistency in typical adults. There has been recent interest in response stability in the neuropsychological literature (van Scherpenberg et al., 2019), but without comparison to neurally intact populations. The current study thus provides a useful baseline.

3.6.1 Population-level norms predict within-speaker variability

By providing the first examination of consistency over repeated naming of the same stimuli in the same task, our results both support and extend the robust effects of picture name agreement reported in previous studies. Logistic regressions of within-subjects name consistency demonstrated that individual's word selections in Session 2 were jointly predicted by the distribution of names in their linguistic community and their own previous responses in Session 1, both for dominant and secondary name use. This suggests that individuals consider the linguistic tendencies observed in a wider population, but they also maintain their stable word preferences across naming episodes. Logistic regressions of within-subjects name switching also demonstrate that population- level name agreement can predict the conflict within the individual, or at least the availability of multiple names, predicting both switches from secondary to dominant names and, more remarkably, from dominant to secondary names. This switching behavior is important for two reasons. First, in line with the Luce choice-inspired stochastic selection account (i.e., that name agreement reflects the availability of all

names within the individual), it confirms that individuals tend to switch to the names that other speakers use more frequently to describe the same stimuli. Secondly, population-level norms also predict speakers' likelihood of spontaneously switching to non-dominant names, thus demonstrating that switching behavior cannot simply be explained as an inevitable move toward more normative responses (i.e., modal names).

Our stability and switching measures indicate that picture name distributions from norms at least predict co-availability of responses within individual speakers, which is a crucial precondition for the interpretation of name agreement effects as reflecting response competition. They cannot directly show that speakers necessarily coactivate multiple labels within the same trial, but that is an assumption that is common to both competitive (Howard et al, 2006; Roelofs, 2018) and noncompetitive (Oppenheim et al, 2010) accounts of word production effects. On the assumption that switching across trials implies coactivation within trials, our results therefore provide necessary preconditions for competition- or conflict-based effects to emerge (as assumed by, e.g., Indefrey & Levelt, 2004; LaGrone & Spieler, 2006; Bose & Schafer, 2017).

If speakers do in fact consider multiple names for the same picture, then recent empirical findings seem to challenge the idea that these names are competing for selection (in the sense of, e.g., Levelt, et al., 1999). For instance, in picture naming norms, after accounting for dominant name agreement, pictures with stronger secondary names appear to be named faster than those with only weaker alternatives (Oppenheim, 2017; in prep). Under a strict competition model, the opposite pattern should emerge. One possible resolution would be to suggest that competitive selection only comes online when a particular task demands it (Nozari and Hepner, 2019), such as an instruction to name a picture while ignoring a superimposed word (picture-word interference; but see e.g., Dylman & Barry, 2018). In that case, however the question arises as to whether online competition is a necessary feature of word production, as opposed to an accommodation to particular experimental tasks (e.g., Oppenheim & Balatsou, 2019).

3.6.2 Population-level norms overestimate within-speaker variability

Although our results provide strong support for a core prediction of the Luce choice account, they also demonstrate that name agreement estimates from norms systematically overestimate within-speaker variability. Within-subjects regressions and within-items Monte Carlo analyses demonstrate that individuals maintain word preferences that are more stable than population-level norms would suggest. These robust individual differences imply that population-level name agreement also reflects individuals' stable word preferences that are likely formed through previous experience. Of course, we cannot rule out the possibility that this apparent stability may have resulted from some kind of long-lasting priming from Session 1 to Session 2. In fact, such persistent priming forms the basis of our alternative idiolect account (cf. Oppenheim, et al., 2010). If it is persistent enough to affect word selection one week later, then it is also plausible to assume that it should affect word selection a week after that and a week after that (e.g., the power law of forgetting: Wixted & Ebbesen, 1991). Thus, even if speaker initially settles on couch by chance, a simple rich-get-richer effect should increase their likelihood of choosing it again in the future, resulting in the development of individual linguistic tendencies over time. Incrementally approximating a one-concept-one-word rule should limit lexical coactivation, and therefore activation error and competition, making production faster and more efficient.

However, any such idiolect account must also address the question of why speakers clearly do maintain synonyms in their productive vocabularies. As our participants' switching behavior demonstrates, speakers who choose couch can also choose sofa, implying that they have not completely eliminated the latter from their vocabularies. One possible explanation for this maintained flexibility comes from the needs of interacting with a larger linguistic community that includes other speakers with different word preferences. In comprehension, it is thus beneficial to maintain many-to-one word-to-concept mappings, and listeners, much like speakers, appear to continually update them for efficient communication (Rodd et al., 2013). There is also direct evidence for lexical alignment

between interlocutors (Garrod & Anderson, 1987) -a tendency for conversation partners to adopt a one-concept-one-word rule for their shared communication- providing a basis for assuming transfer between the comprehension and production systems. Although it may be efficient for a speaker to maintain a single word for a concept, in terms of their own production needs, communication requires and provides flexibility.

3.7 Conclusion

This study provides the first demonstration that picture name agreement has a psychological reality within individual speakers, comparing predictions from a stochastic account of the phenomenon to those from an idiolect-based account. There is some evidence that name agreement, as measured in the traditional way, relates to within-individual lexical co-activation, and by extension possible lexical competition. Norms from a speaker's linguistic community do predict their likelihood of using particular names, and even their likelihood of switching to alternative names when retested, suggesting that speakers consider the range of names observed in their larger linguistic community. But we also have evidence for more stable differences between individuals' semantic-to-lexical mappings: speakers are far more consistent in their naming preferences than would be expected by chance alone. It is unclear from the current study whether that consistency reflects within-experiment priming effects, in line with an incremental word learning framework (Oppenheim et al., 2010), or pre-existing differences in how speakers map from concepts to words. Given this heterogeneity among speakers, it is remarkable that name agreement measures do such a good job of predicting naming performance and show such consistent neurophysiological effects. This efficacy is somewhat surprising, but not too surprising, because it is still probably the case that pictures that have multiple names have more lexical co-activation, even if population measures of name agreement are not the perfect way to measure that co-activation.

In general, there are certain challenges when assuming static properties of a processing system, such as language, that continually changes through experience; we cannot assess current performance without affecting future performance. Thus, in language production, as elsewhere, population-level norms usefully supplement the data that we can collect from individuals. But we need to exercise caution when assuming that things that are true on a population level must also be true within an individual. This concern is emblematic of a wider concern that we see elsewhere, such as in the debate between group-level and case-study approaches in the neuropsychological literature: although trends may hold when collapsing across individuals, accurate psychological interpretation of a pattern crucially depends on sufficiently powered evidence from within individuals.

CHAPTER 4- Endogenous conflict and exogenous competition in word production: an ERP study of name agreement in overt picture naming

In this chapter, I report how we investigated the flexibility of the lexical selection mechanism by evaluating behavioral and electrophysiological name agreement effects in picture naming before and after corrective familiarization. The inherent conflict that originates from picture name agreement variations is later replaced by the demand to switch to undesired responses for production. This suggests that the selection mechanism handles differently endogenous lexical co-activation and exogenous response competition.

This chapter is submitted for publication as an article: Balatsou, E., Thierry, G., & Oppenheim, G. M. (submitted). Endogenous conflict and exogenous competition in word production: an ERP study of name agreement in overt picture naming. *Cognitive Neuropsychology*.

Abstract
How many mechanisms do we use for word selection in language production? Differences in picture name agreement—an empirical measure of how often people produce a picture's modal name—are associated with robust behavioral and electrophysiological correlates that researchers often interpret as evidence that a competitive lexical selection process is part of normal language production. Complicating this interpretation, previous electrophysiological studies of name agreement have typically begun with an unrecorded 'familiarization' procedure, specifying the precise words that participants should use to name each picture and thus requiring them to suppress their preferred names when they conflict. Here, our theoretical question is whether such task demands merely amplify conflict within a core lexical selection mechanism, or instead engage adjunct, task-induced control processes. We measured naming latencies and electrophysiological activity as participants named high- and low-agreement pictures before and after a corrective familiarization procedure. Critically, our familiarization procedure introduced name changes for high and low agreement pictures equally often. Naming latency analyses indicate that the name agreement effects that emerged before corrective familiarization (that is, when participants were simply asked to name pictures, as in timed norming studies) were supplanted by name change effects after familiarization. Prior to corrective familiarization, event-related brain potentials linked name agreement to modulations in the N200 and N400 time windows. After corrective familiarization, weak forms of those effects were accompanied by two new effects of directed name change: an independent modulation in the N200 window and a later anterior positivity. Taken together, these results suggest that, although speakers can successfully exert deliberate control over the endpoint of word production, doing so primarily involves processes distinct from those that typically support lexical selection in language production.

Keywords: name agreement; picture naming; event-related potentials; selective inhibition; lexical competition

4.1. Introduction

How do speakers select a word for speech? Most models broadly agree on two major levels of representation that underlie single word production: a semantic level and a form-based level (Dell, 1986; Levelt et al., 1999). As a stimulus (e.g., the picture of a cat) triggers the activation of several corresponding semantic features (e.g., [PET], [FOUR-LEGGED], [FURRY]), these activate representations of both the target word (e.g., "cat") and other words that correspond to similar meanings (e.g., "dog"). One of these lexical representations can then be selected for further form-based processing, eventually culminating in its overt articulation. While models agree on the existence of these levels, they continue to debate the functional properties of the lexical selection process that bridges them. The selection process must implement some kind of winner-take-all function, but a key question is how a mechanism could correctly select a desired word without already knowing which word it desires. One general approach is to assume that the system gradually accumulates evidence (or activation), selecting the first word whose evidence exceeds some simple threshold or decision criterion (e.g., Mahon et al., 2007; Oppenheim et al., 2010, Simulation 6; cf. Anders, Ries, van Maanen, & Alario, 2015). For instance, the system would select cat if cat reached the threshold first, and dog if dog reached the threshold first, but because cat's progress toward its threshold is independent of dog's, this kind of selection process is typically described as "non-competitive". By contrast, "competitive" accounts posit that lexical selection necessarily considers the relative evidence for one candidate over any others, thus selecting the first word whose evidence exceeds a relative threshold (Levelt, Roelofs, & Meyer, 1999; Nozari & Hepner, 2019; cf. Roelofs, 1992). For instance, if cat and dog were strongly co-activated while a speaker attempted to retrieve cat, this co-activation would make cat's selection not only less likely (as in the non-competitive account) but also more effortful and time-consuming. Thus, the core theoretical distinction between the accounts is whether the lexical selection process that bridges the semantic and form-based levels considers evidence for alternatives as evidence against a

target, and therefore makes target selection more difficult (not merely less likely) when one has a strong alternative.

4.1.1 Evidence for exogenous lexical competition

The majority of evidence in favor of a competitive account of this lexical selection process derives from a task called picture-word interference (e.g., Schriefers, Meyer, & Levelt, 1990). In picture-word interference experiments, participants are generally instructed to name a picture (e.g., a picture of a cat), while attempting to ignore the simultaneous visual or auditory presentation of a 'distractor' word (e.g., the word "dog"). Naming latencies are typically greater when the distractor word is semantically related to the target (e.g., "cat") than when it is unrelated (e.g., "couch"; for a recent review, see Bürki et al., 2020). Although the competition in this task comes from obvious sources outside the production system—a superimposed distractor and an explicit task demand to suppress responses to it—proponents of the competitive selection account (e.g., Levelt et al., 1999; Roelofs, 2018) typically assume that (1) these sources increase competition within the core lexical selection process of the production system, and (2) to resolve this competition, speakers engage the same core mechanisms that they would normally use to select words in communicative language production. That is, researchers characterize this exogenous manipulation as amplifying endogenous competition, and therefore interpret the associated naming latency effect as revealing competitive endogenous mechanisms for lexical selection. But alternative accounts of picture-word interference argue that the empirical results do not actually provide compelling evidence for a competitive model of lexical selection (see Oppenheim & Balatsou, 2019, for recent discussion). For instance, the Response Exclusion Hypothesis posits that speakers suppress distractors post-lexically, at a later pre-articulatory or response-buffer stage, thereby obviating the need for competition in the earlier lexical selection process (Mahon et al., 2007; Janssen, Schirm, Mahon & Caramazza. 2008; Dhooge et al., 2013; Dhooge & Hartsuiker, 2010; 2011). Thus, although the effects of semantic manipulations in

picture-word interference tasks are empirically robust, interpretations differ as to whether it is assumed that speakers integrate the task demands into their earlier lexical selection process or address them by recruiting additional mechanisms, such as later, post-lexical monitoring.

In paradigms with less obvious task demands, the evidence that has been claimed to support competition can also be explained without it. For instance, in blocked cyclic picture naming studies, where participants repeatedly name a small set of pictures, stimuli in semantically related blocks (e.g., cat, dog, and horse in a single block) elicit longer naming latencies than stimuli in semantically unrelated blocks (e.g., cat, flower, and couch; e.g., Vigliocco, Vinson, Damian, & Levelt, 2002). Similarly, when speakers name pictures in semantically heterogenous sequences, their naming latencies increase as a function of the target picture's ordinal position within its semantic category (Howard et al., 2006). Like semantic picture-word interference, both manifestations of cumulative semantic interference are empirically robust and were long claimed to support the competitive lexical selection account (e.g., Howard, Nickels, Coltheart, & Cole-Virtue, 2006; Vigliocco et al., 2002; Wheeldon & Monsell, 1994), for instance under the assumption that lateral inhibition between activated candidate words (e.g., Harley, 1993; Howard et al., 2006; Stemberger, 1985) implements a relative decision criterion (e.g., Levelt et al., 1999). However, a more recent model of word production demonstrated that persistent, experience-driven adjustments of semantic-to-lexical mappings are sufficient to explain the effects, obviating the need for competition in the selection mechanism itself (Oppenheim et al., 2010): Each time a speaker retrieves a word for production, an incremental learning process adjusts the mapping, both increasing the ease and likelihood of retrieving the target again (e.g., "dog") and decreasing the ease and likelihood of retrieving co-activated alternatives (e.g., "cat"; see also, e.g., Damian & Als, 2005; Navarrete, Del Prato, Peressotti, & Mahon, 2014). Although many researchers still use the assumption of competitive lexical selection to describe such cumulative semantic interference effects, they now generally acknowledge that the effects alone do not compel that interpretation (e.g., Nozari & Hepner, 2019). Moreover, even though these paradigms present

subtler task demands than picture-word interference, there remains concern about factors such as strategic response preparation in the blocked cyclic naming paradigm (e.g., Belke, 2017; cf. Schnur, 2014) and distinguishing their possible contributions from the more typical functioning of the language production system.

4.1.2 Evidence against endogenous lexical competition

One way to reduce the contribution of task demands is to consider the competition that is naturally present in the production system (as suggested by Oppenheim and Balatsou, 2019). Even when presented on their own, in norming studies, pictures vary in how reliably they elicit particular names, a property called picture name agreement. By definition, pictures with higher name agreement elicit their dominant (i.e., modal) names more reliably than those with lower name agreement, and this characterization holds for individual speakers as well as their linguistic communities (Balatsou, Fischer-Baum, & Oppenheim, in revision). Speakers are also quicker to name pictures that have higher name agreement (e.g., "cat") than those with lower name agreement (e.g., "couch", which is also commonly named as "sofa" in American English; e.g., Lachman, 1973; Gilhooly & Gilhooly, 1979; Vitkovitch & Tyrell, 1995; Bates et al., 2003; Alario et al., 2004). In brain imaging studies, naming pictures with lower name agreement is associated with greater activation of the left inferior frontal gyrus, claimed to resolve or 'bias' the assumed competition (Kan & Thompson-Schill, 2004). In electrophysiological studies, naming pictures with lower name agreement has been claimed to increase amplitudes in a stimulus-locked N200 time window (Shao et al., 2014: 170-330 ms; Cheng et al., 2010: 250-350 ms, but cf. Valente et al., 2014) that associated Indefrey & Levelt's (2004) 200-350ms claim for the process of "lemma" retrieval. Such name agreement effects are typically interpreted as evidence for endogenous competition during lexical selection: if norms reveal the range and strength of names that individual speakers consider each time they name a picture, then pictures with lower

name agreement should require more control to suppress the co-activation of alternative names and thereby choose the very best.

But we have recently argued that this competitive interpretation of name agreement effects may be circumstantial (Balatsou, Fischer-Baum, & Oppenheim, in revision; Oppenheim, 2017; in prep). Target accessibility is sufficient to link response speed with response likelihood without recourse to competitive selection, so a more specific test of the competitive account should instead consider the strength of the strongest alternative (cf. Oppenheim et al., 2010). Ceteris paribus, a name that faces focused competition from a single strong alternative should be harder to select than one that faces distributed competition from an array of weak alternatives. Remarkably, our norms that quantified competition in this manner demonstrated that, all else being equal, speakers were actually quicker to produce the dominant names for pictures (e.g., "couch") that have strong alternatives (e.g., "sofa") than those without. That is, the study replicated the classic RT facilitation effect of "dominant" name agreement, but further showed that production in these "high-conflict" situations (with a strong "secondary" name) was actually faster rather than slower, a result that is more consistent with a non-competitive model where speakers simply choose whichever word reaches threshold first.

4.1.3 Reconciling the accounts: how many mechanisms can speakers use for lexical selection?

We see two possible ways to reconcile the apparent evidence for exogenous competition from tasks like picture-word interference with the evidence against endogenous competition from tasks like picture naming norms. The first way, which we call a multi-factor account, is to assume that manipulations of exogenous competition engage selection mechanisms that are distinct from those that resolve endogenous conflict. For instance, if a speaker is given the goal of naming a picture instead of a distractor, or otherwise wishes to use or avoid certain words, they might implement ad hoc constraints via a generate-then-test procedure (cf. Barsalou, 1983), choosing responses and then evaluating, disrupting, and correcting them as necessary. Mahon and colleagues' (2007) Response Exclusion

Hypothesis is one example of such an account, as is Nozari and colleagues' (2016) distinction between selection control and post-monitoring control, but Levelt's (1983) perceptual loop hypothesis fits equally well; the key attribute is that a multi-factor account separates the 'smart' imposition of ad hoc constraints from a more limited process of initial lexical selection (cf. Fodor, 1983).

The second way, which we call a single-factor account, is to assume that the core process for lexical selection directly incorporates such ad hoc constraints. The most compelling example of a single-factor account is Nozari and Hepner's (2019) recent proposal for a flexible selection criterion that scales a relative threshold according to task goals. According to their model, speakers always use the same competitive process to select their words, but when task demands require them to name a target instead of a distractor, or use a particular 'best' word to name a picture instead of a reasonable alternative (e.g., "couch" instead of "sofa"), they increase their selection criterion, implementing a speed-accuracy trade-off that exaggerates competitive effects. When a task allows any reasonable response—more typical of picture naming norms—speakers reduce this criterion, so response times depend more on the time course of lexical activation than the time course of lexical selection, and, consequently, lexical selection may appear non-competitive. Similar core assumptions—though without the ability to reconcile contradictory findings—underlie the long-time use of picture-word interference as the basis for models of typical language production (e.g., Levelt et al, 1999), as well as the assumption that name agreement effects are robust to corrective familiarization (e.g., Alario et al., 2004; Shao et al, 2014; Valente et al., 2014; more on this below), most clearly illustrated in Roelofs (1992) claim that the lexical selection process in picture-word interference directly incorporates a response-set constraint. According to recent versions of such accounts (e.g., Shao et al., 2015), the core process underlying all competitive selection effects (name agreement, cumulative semantic interference, and semantic picture-word interference) is "selective inhibition": even without explicit distractors, naming a picture as "couch" involves specifically, effortfully, suppressing the alternative, "sofa". Although such single-factor accounts should, in principle, be compatible with the same

secondary monitoring processes that we described for multi-factor accounts, their crucial distinguishing feature is that they allow task goals to directly modulate the selection process, effectively reaching inside the selection module, and argue that any late monitoring or adjunct control processes are merely incidental to observed "competitive" effects.

4.1.4 Using electrophysiology to compare correlates of endogenous and exogenous lexical conflict

In light of these two ways of incorporating task goals, it is worth considering the role of corrective familiarization in picture naming experiments. For the purely practical goal of avoiding data loss, researchers often[3] precede picture naming experiments by introducing participants to each of the stimuli and providing the specific names that they should subsequently use to name them. For instance, before Valente and colleagues (2014) assessed ERP correlates of picture name agreement, their participants first read a booklet that introduced each picture and its 'correct' name; Shao and colleagues' (2014) followed their initial booklet-based training phase with a second training phase where participants attempted to produce the specified names in response to the specified images, receiving corrective feedback whenever their responses deviated. Such corrective "familiarization" would seem reasonable if one assumed that each picture had exactly one correct name, and participants just needed to be pointed to it. In fact, in a recent two-session norming study Balatsou, Fischer-Baum, & Oppenheim (in revision) showed that speakers were more likely to spontaneously switch from non-dominant names to dominant names that were more commonly used by other members of their linguistic community (i.e., across-subject name agreement predicted within-subject voluntary name change), a pattern that seems consistent with the idea that dominant names are generally "more correct" than alternatives. But we also found evidence that speakers often consider their alternative names to be correct as well: those who used non-dominant names in the first session were more likely to re-use

[3] For instance, Bürki et al (2020) note that 143 of the 161 picture-word interference experiments in their meta-analysis began with such a familiarization phase, precipitating recent debate about its possible role in establishing picture-word interference effects (e.g., Collina, Tabossi, & De Simone, 2013; Gauvin, Jonen, Choi, McMahon, & de Zubicaray, 2018).

them later than one would expect by chance. If speakers do maintain stable preferences for alternatives despite their subordinate status, then corrective familiarization seems more problematic: directing speakers to use particular names—even dominant names—may actually introduce a novel task demand to override own preferences, suppressing one name to recall another instead, thus creating a form of exogenous competition (cf. Nozari et al., 2016). Though long-term incremental learning processes (e.g., Oppenheim et al., 2010) may produce some quick adaptation to such corrective feedback, we assume that maintaining the specific task goal and name target should depend more on adjunct memory and control processes[4]. The key theoretical question is whether the language production system incorporates this task goal into the same (early) process that it employs to resolve endogenous lexical conflict (e.g., Nozari & Hepner, 2019) or instead applies ad hoc executive control processes (e.g., Oppenheim & Balatsou, 2019).

In this paper, we use event-related potentials to adjudicate between the single- and multi-factor accounts, by assessing whether endogenous conflict and exogenous competition affect the same aspects of the picture naming process. In doing so, we also provide the first ERP-based assessment of picture name agreement effects in overt speech that is not confounded by corrective familiarization. This study takes the form of a single-session, three-phase, overt picture naming task that integrates the free-response approach that researchers have typically used to assess behavioral correlates of picture name agreement (e.g., Bates et al., 2003; but cf. Alario et al., 2004) with the corrective-familiarization approach that researchers have typically used to assess electrophysiological correlates of name agreement (e.g., Shao et al, 2014; Valente et al, 2014; but cf. Cheng et al, 2010). In Phase 1, we recorded response times and electrophysiological activity as participants named 68 high name agreement and 68 low name agreement pictures, randomly interleaved; this provides measures of pre-familiarization correlates of name agreement, and thus selection from among acceptable alternatives.

[4] Working with a patient with encephalitis-based left hippocampal damage, we have previously noticed typical implicit learning combined with particular difficulty incorporating corrective familiarization (Oppenheim, Tainturier, & Barr, 2015; Oppenheim, Barr, & Tainturier, 2016).

Note that we selected low-agreement pictures that had at least one strong alternative name and avoided those that had low agreement due to low-level visual or conceptual problems (e.g., Cheng et al., 2010; Vitkovitch & Tyrell, 1995). In Phase 2 we provided corrective familiarization, changing or confirming the names for high- and low-agreement pictures with equal likelihood (i.e., 50% in each case; e.g., frog → toad, truck → lorry); we did not collect data during this phase, both because we lacked relevant predictions and because presenting multiple stimuli in each trial (i.e., the picture and its desired name) is problematic for ERP methods. In Phase 3, we recorded response times and electrophysiological signals as participants named the pictures again, this time using the specified names, thus providing the first direct measure of the effects of 'correction' in corrective familiarization and allowing us to empirically assess whether it modulates the same ERP components associated with simple name agreement. We can thus ask 1) when and how name agreement effects manifest in the initial free-naming phase, 2) when and how a task goal to use a particular name affects the naming process, and 3) whether the two sources of lexical conflict interact.

We can identify several predictions for the effects of name agreement in Phase 1, before corrective familiarization. From many previous behavioral studies (e.g., Bates et al., 2003), we expect lower name agreement to be associated with longer naming latencies. But a major question for our electrophysiological investigation is whether name agreement modulates N200 amplitudes without prior corrective familiarization. Recent reviews argue that lexical access in picture naming typically begins within 200ms of stimulus onset (Strijkers & Costa, 2011; Indefrey, 2011). Though Shao et al.'s (2014) report of a 170-330 ms right-anterior N2 modulation fits well within this window, the plausibility of their claim that it indexes selective inhibition—and thus that name agreement effects in general index lexical competition—seems to hinge on their prior use of corrective familiarization. Evidence for such name agreement-based N200 modulations in other studies is less convincing. Although Cheng et al. (2010) did report a name agreement-based 'N2' modulation in a task without prior corrective familiarization, their use of covert naming, their later time window (250-350 ms), and

the left parietal distribution of their effect raise some doubt as to whether their modulation actually indexed the same processes. And Valente et al's (2014) post-familiarization study did not report any significant name agreement-based modulations until much later, 380 ms after stimulus onset[5]. Thus, detecting a name agreement-based modulation in the N200 range would generally increase our confidence that name agreement specifically affects early lexical access processes. And specifically detecting a right-anterior N200 modulation without prior corrective familiarization would cast doubt on its characterization as indexing selective inhibition, as opposed to mere lexical co-activation or selection processes more broadly. Moreover, if N200 and RT name agreement effects both specifically index lexical selection as opposed to mere (co)activation, then we might expect the magnitudes of those effects to correlate. Considering the reports of later name agreement-based modulations, and the uncharacteristic topography of Cheng et al.'s N2, it also seems plausible that agreement could modulate amplitudes in the N400 range, typically measured over central electrodes, which has been associated with increased naming difficulty in other tasks (e.g., Schendan & Kutas, 2007; Costa et al, 2009; Schendan & Maher, 2009).

We consider the influence of more constrained production goals and task-mandated lexical control in Phase 3, when participants attempt to name the pictures using our corrected or confirmed names. Because this will be the third time participants have seen these pictures (as in Shao et al., 2014) and the second time they have named them (as in Shao et al., 2014, and Alario et al., 2004), one would typically expect memory and implicit learning to produce some overall reduction in both naming latencies and ERP amplitudes, but an increase in response caution may counteract that trend. If name agreement effects are robust to the imposition of task-mandated lexical control (Alario et al., 2004; Shao et al., 2014; Valente et al., 2014) then we should see the same name agreement effects as in Phase 1. In addition, if the goal of choosing a particular word merely amplifies the endogenous competition

[5] The unfortunate practice of omitting statistics for non-significant contrasts prevents any further attempt at reconciling the results.

that normally underlies name agreement effects (per the single-factor account), then we would expect corrective familiarization to exaggerate any name agreement effects that were already present in Phase 1 (i.e., producing superadditive Phase X Name Agreement interactions for both naming latencies and relevant ERP effects).

There is also the more specific question of how speakers implement a task demand to override a preferred name, recalling and producing a less-preferred name instead. Behaviorally, implementing such a change should generally increase naming latencies. If speakers incorporate such task demands into the early lexical activation and selection process-whether by selectively suppressing a preferred name (Shao et al., 2014), suppressing responses in general (e.g., by increasing a simple or relative selection threshold), and/or incorporating short-term memory as a secondary source of lexical activation (e.g., simply increasing the activation of the task-mandated option)-then the name change manipulation should affect the same early processes as name agreement. Most accounts would also predict a particularly large name change cost for high-agreement items, on the assumption that it would be especially difficult to suppress the preferred name for high-agreement stimulus (but cf. Mahon et al., 2007), and/or that it would be especially difficult to activate an otherwise-weaker specified alternative. Observing such an interaction in naming latencies would do little to distinguish between accounts, but for ERPs the single-factor account would specifically require an interaction in the early (N200) time window. On the other hand, if speakers simply recruit additional mechanisms to fit their naming behavior to specific task goals, then we would broadly expect little or no convergence between the ERP modulations associated with name agreement and name change predictors, and the emergence of distinct components that were not present during the initial picture naming task. For instance, if targeted production engages a select-but-verify approach, where speakers initially select a preferred name as usual, and then monitor and repair it to fit the new task goal (e.g., Nozari et al, 2016; cf. Mahon et al, 2007), then we might expect normal early name agreement effects to be followed by later, additional modulations, independently associated with the name change manipulation, though the

variety of possible mechanisms (Postma, 2000) prevents a more precise prediction. While a single-mechanism account would not necessarily be incompatible with such additional differences, it would predict that the most important components would be common to both free naming and targeted selection tasks.

4.2. Materials and Methods

4.2.1 Participants

Twenty-six healthy Bangor University students (16 females; Mean Age =20.9 years, *SD* = 3.20; twenty-two right-handed and 4 left-handed) were recruited from a participant panel and took part in a three-phase picture naming task in a single session. Data from nine participants were discarded because they failed to produce at least 30 valid, artefact-free trials per condition: two failed to switch to the desired names on more than 30% of the trials, and seven had excessive alpha contamination or electrode drifts. Therefore, the current analysis is restricted to 17 datasets (13 females, 4 males; 14 right-handed, 3 left-handed; Mean Age = 19.6 years, *SD* = 1.96). All participants were native English speakers and had normal or corrected-to-normal vision, no neurological impairment and no self-reported symptoms of developmental dyslexia. Participants that signed-up for the experiment were given an information sheet and informed consent which they signed before taking part in the study. The study was approved by Bangor University Ethics Committee and participants received course credit or cash compensation for their participation.

4.2.2 Stimuli and Design

As stimuli for the naming task, we selected 176 black-and-white line drawings of common objects (68 High agreement, 68 Low agreement; see Table 1) from the International Picture Naming Project (Bates et al., 2003). To identify their local names and name agreement, we used Oppenheim's (in prep.) recent picture naming norms, gathered from 100 different individuals from the same

population. When selecting the low name agreement stimuli, we ensured that the pictures had at least one strong alternative name rather than having just low identifiability. Given other constraints, we also attempted to match word lengths (High agreement = 5.26 letters; Low agreement = 5.65 letters) and frequency (in Zipf values from SUBTLEX-UK: van Heuven, Mandera, Keuleers, & Brysbaert, 2014).

Table 1. Mean Name agreement (Oppenheim, in prep) and Lexical Frequency (van Heuven et al, 2014) Zipf values for the stimuli. Standard deviations are included in the parentheses.

Name Agreement	**Name**	**Lexical Frequency Mean (SD)**	**Name use proportion Mean (SD)**
High (N = 68)	Dominant	4.39 (0.54)	0.93 (0.12)
	Secondary	3.73 (0.98)	0.05 (0.09)
Low (N = 68)	Dominant	3.81 (1.10)	0.59 (0.12)
	Secondary	3.88 (1.06)	0.23 (0.11)

Because the experimental task required participants to change between dominant and 'secondary' alternative names in both the high and low name agreement conditions, we took care to identify plausible alternatives. For both high and low agreement pictures, we therefore selected as the alternative the second most commonly used name from Oppenheim's (in prep.) norming study. Thus, all of these alternatives emerged from the same norming process as the dominant names. In the few cases where the norms provided no alternative names-images with perfect name agreement-we chose alternative names from the WordNet online database (Princeton University, 2010). Whenever we could choose between similarly plausible alternative names, we also attempted to match the lexical frequency of dominant and alternative names, $F(1, 134) = 2.893$, $p = .09$ (Table 1).

The experimental task consisted of a three-phase design that participants completed in a single session on the same day (Figure 1): (1) an initial overt picture naming task (free naming), (2) corrective familiarization, and (3) a final overt picture naming (post-correction target naming). During Phase 1 (free naming), participants freely named each picture and we coded their responses online as matching the expected dominant name, secondary name, or neither. Based on the names that each participant produced in this block, an algorithm quickly selected their target names for use in Phase 2

(familiarization), changing the targets for 50% of all dominant- and secondary-named items; other responses were randomly changed to either name with equal frequency, but excluded from subsequent analyses.

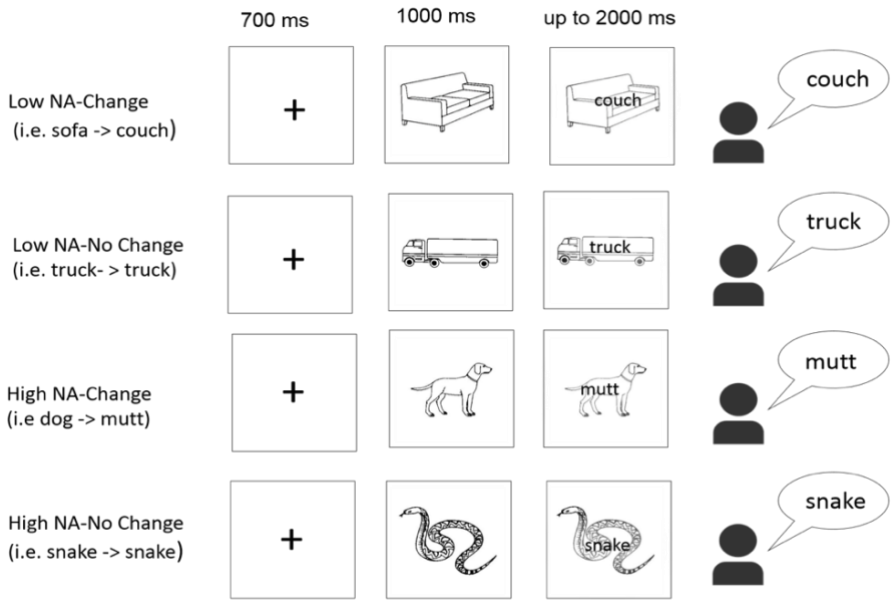

Figure 1. Familiarization design and experimental conditions

This Name Change manipulation similarly affected all items, counterbalanced across participants, and all conditions within each participant (i.e., providing approximately the same number of valid trials in the Name Agreement x Name Change conditions for each subject), so any difference between the ERPs elicited in the different conditions should not be due to any uncontrolled differences between the stimuli or the familiarization process.

4.2.3 Experimental Procedure

Prior to the experiment, each participant degreased their scalp by washing with baby shampoo and water. Then, they were seated in a comfortable chair, within a soundproofed room with dimmed lights, 100 cm away from a 19-inch computer monitor. A cap with 64 electrodes was fitted to their head. Small amounts of alcohol were used to further degrease their scalp before small amounts of gel

(consisting of electrolytes, sand and water) were deposited at each electrode site to increase electrical conductivity between scalp and electrodes.

Participants were instructed to relax and refrain from blinking or moving unnecessarily, and to name each picture as accurately and fast as possible, in a clear and loud voice. Pictures were presented via E-Prime 1.2 (v. 1.2.1.847) on a 17" CRT. Responses and naming latencies were recorded via a 'shotgun' small diaphragm condenser microphone, placed approximately 15 cm away from the participants' mouth, feeding into both a digital recorder and a delayed-threshold voicekey. An experimenter in an adjoining monitoring room manually recorded the participants' productions. Each block in each phase began with a set of five high name agreement filler pictures (name agreement: M=0.87, *SD* = .14; frequency: M= 4.47, *SD* = .57) followed by a block of 68 experimental trials. Short self-paced breaks followed each 68-trial block and each phase.

During Phase 1 ("free naming"), participants were simply instructed to name each picture as quickly and accurately as they could, while refraining from making any other noises (i.e., producing any appropriate name, per standard norming instructions). Each trial began with a small black fixation cross at the center of the screen which appeared for 150 ms, followed by a picture (567 x 567 pixels) appearing at the center of the screen for 3000 ms or until the participant's response triggered the voicekey. In Phase 2 (corrective familiarization), participants were instructed not to name each picture until the correct name appeared atop it, because they would need to use that correct name later in the experiment. Each trial began with a small black fixation cross appearing for 700 ms, then the picture appeared alone for 1000 ms and then the desired name was superimposed on the picture for an additional 2000 ms, or until the participant's voice triggered the voicekey. Each trial therefore lasted ~1500 ms after the onset of the picture display. Phase 3 (post-correction target naming) followed the same procedure as Phase 1, except that participants were instead instructed to name each picture using the words that we had provided in Phase 2.

4.2.5 Electrophysiological recording and data analyses.

Electrophysiological data were recorded in reference to Cz at a rate of 1 kHz from 64 Ag/AgCl electrodes placed according to the 10-20 convention (Jasper,1958). Impedances were kept below 5 kΩ. Ocular artefacts were monitored using vertical electro-oculogram channels (VEOG) set above and below the left eye. All pre-processing steps and analyses were performed using the EEGLAB Toolbox (version 14.1.2b; Delorme & Makeig, 2004) in MATLAB (version R2018a, Mathworks Inc.) and Neuroscan (Scan 4.5, Compumedics). EEG data were filtered bandpass using zero-phrase shift digital filtering (0.1 Hz, 24 dB/oct – 20 Hz, 48 dB/oct). All data were visually inspected for abnormalities and sections of continuous data containing major muscle artefacts or recordings taken during pauses were dismissed.

Ocular artefacts were mathematically corrected using independent component analysis (ICA) in EEGLAB. Independent components (ICs) were inspected by plotting component activations as well as component spectra and maps to see which ICs contributed the most at 5 Hz and 20 Hz frequencies. ICs containing ocular and muscle artefacts as well as electrode pops were removed. Prior to accepting ICA correction, we plotted the EEG data before and after ICA correction to make sure that rejecting ICs led to ocular artefact correction rather than spurious data changes. On average, 2.52 ICs (SD = .71; min = 1, max = 3) were rejected per participant. EEG files were then visually inspected for remaining artefacts and EEG periods contaminated by remaining artefacts were manually excluded. Continuous recordings were cut into epochs starting 200 ms before and ending 1000 ms after picture onset. Baseline correction was performed in reference to pre-stimulus activity (-200–0 ms) for stimulus-locked analyses. Individual averages were digitally re-referenced to the global average reference and Individual ERPs were then averaged together in each condition to obtain grand averages. ERP components were defined based on inflection points in the mean global field power (MGFP) measured across the scalp, which summarizes the contribution of all electrodes in the form of a single vector norm.

4.2.6 Analytical Approach

Responses were initially transcribed on-line and were recorded and manually checked offline for accuracy. To ensure that we were only comparing valid trials, we excluded 621 (13%) trials where a speaker failed to produce either the dominant or secondary name in the Phase 1 or failed to switch to the desired name in Phase 3. This left 4160 picture naming attempts for our analyses in total: 2055 in Phase 1 and 1938 in Phase 3. We also excluded 78 trials with recorded naming latencies shorter than 400 ms from both the behavioral and ERP analyses, on the assumption that such quick responses could not be valid in this paradigm, thus restricting the current dataset to 3925 trials in total, 2026 in Phase 1 and 1899 in Phase 3.

Inverse-transformed naming latencies (-10000/RT) were analyzed with confirmatory linear mixed effects regression, via the lmer function in the lme4 v1.12 library (Bates et al.,2016) in R (v5.5.1, R Development Core Team, 2016). All fixed effects were sum-coded (centered around their condition-weighted midpoint), and all models included maximal random effects structures (Barr et al., 2013) for participants and items, omitting correlations between random effects to facilitate convergence. P-value estimations use the Satterthwaite approximation.

The aim of the ERP analyses was to determine whether the same effects would emerge for name agreement in simple picture naming and in naming with explicit task demands to switch names. For this reason, we compared mean ERP amplitudes in time windows that we had selected based on previous reports of picture name agreement effects in Phase 1 (i.e., Shao et al., 2014; Cheng et al., 2010; Valente et al., 2014), and also compared the change and no change conditions in these same time windows in Phase 3.

ERP analyses used linear mixed effects regression of mean amplitudes for each condition for each participant, for each specified electrode within the specified time windows, including per-participant maximal random effects structures, as above, and per-item random intercepts; the main

advantage of this approach over traditional ANOVA for ERPs is its ability to directly provide meaningful, directional effect estimates in the form of β-values and standard errors for each contrast. All predictors are sum-coded binomials (Name Agreement: {Low = -0.5, High = 0.5}; Name Change: {No Change = -0.5, Change = 0.5}), and p-value estimations use the Satterthwaite approximation. Phase 1 analyses assessed name agreement effects in the N200 and N400 ranges separately. Phase 3 analyses assessed name agreement, name change, and their interaction in the N200 and N400 ranges separately. Phase 3 analyses also assess name change effects for a later (450-600 ms) positive modulation observed in the current study, but since this effect was not specifically predicted nor previously empirically reported, our analysis and interpretation are more speculative.

4.3 Results

4.3.1 Behavioral results

Our 17 participants provide 4624 total trials for our analyses (2312 in Phase 1 and 2312 in Phase 3). We restrict our analyses to the 3925 (84%, summarized in Table 1), to focus on trials where participants produced either the dominant or secondary name in Phase 1 and successfully switched to the targeted dominant and secondary name in Phase 3. Out of the 2312 total trials in Phase 1, we focus our analyses on the 2026 (88%) where participants' responses corresponded to either the picture's dominant name (1669 trials, 72% of all Phase 1 trials) or its secondary name (357, 15%); tertiary names or other responses consisted of 286 (12%) trials in Phase 1 and were excluded from the analysis. These name frequencies and their associated naming latencies in Phase 1 correspond well to Oppenheim's (in prep.) norms for the same items' dominant (Frequency = 76%, Mean RT = 900 ms, *SD* = 186 ms) and secondary (Frequency = 14%, Mean RT =1046 ms, *SD* = 318 ms) names, and by-item response frequencies correlate well between Phase 1 of the current experiment and Oppenheim's recent norms for both dominant ($r = .75$, $p < .01$) and secondary ($r = .60$, $p < .001$) name agreement (see Appendix A), thus validating our use of these norms in the design of this study.

From the 2312 total trials in Phase 3, we focus our analyses to the 1899 (82%, also summarized in Table 2) trials where, after having used either the dominant or secondary name in Phase 1, participants successfully named pictures using the dominant (974 trials, 42% of all Phase 3 trials) or secondary (925, 40%) name that we specified during Phase 2. We excluded 413 (17%) of the trials from the analysis that did not meet these criteria: 58 (3%) trials in the high Name Agreement-No Change condition, 97 (5%) trials in the high Name Agreement-Change condition, 116 (6%) trials in the low Name Agreement-No Change condition and 142 (7%) trials in the low Name Agreement-Change condition. These restrictions allow us to assess the effects of our corrective familiarization procedure. The relatively similar frequencies of the dominant and secondary responses in Phase 3 (mean difference in Phase 3: .029, $t(16) = 0.90$, $p = .38$), contrasting with the corresponding frequencies from Phase 1 (mean difference in Phase 1: .77, $t(16) = 22.76$, $p < .001$; mean difference between Phase 1 and 3 differences: $t(16) = 18.87$, $p < .001$), indicate that participants were generally successful in overriding their preferred names to use the names that we had specified in Phase 2.

Table 2. Mean picture naming latencies (ms) in Phases 1 and 3. Standard deviations are included in the parenthesis. Subject-weighted mean Ex-Gaussian component estimates follow. To allow direct comparison between Change and No Change conditions, items depicted in Phase 1 are back-sorted according to their Phase 3 status (Change-No change).

Name Agreement	**Phase 1 (Free naming)**					**Phase 2 (Familiarization Feedback)**	**Phase 3 (Post-correction target naming)**				
	N	**Mean RT (*SD*) ms**	***μ***	***σ***	***τ***		***N***	**Mean RT (*SD*) ms**	***μ***	***σ***	***τ***
High	529	902 (223)	733	99	171	No Change	520	932 (204)	778	77	152
	549	913 (249)	712	85	202	Change	481	1095 (285)	904	145	205
Low	484	1111 (352)	850	185	265	No Change	462	1003 (293)	768	124	240
	464	1084 (349)	819	149	275	Change	436	1118 (331)	849	127	277

To assess how name agreement effects on response time are modulated by corrective familiarization, we used linear mixed effects regression (Table 3) to predict inverse-transformed naming latencies as a function of three centered predictors and their interactions: (1) Phase (i.e., naming phase) (a binomial predictor, {Phase 1 = -0.5, Phase 3 = 0.5}), (2) Name Agreement from Oppenheim's (in prep) recent Bangor norming study (a continuous predictor, {0:1}, centered at 0.76, its mean value for all items in the study) and (3) Name Change (a binomial predictor, {No change = -0.5, Change = 0.5}). Collinearity was not an issue here because the experimental conditions were not correlated. We also report two sub-models, restricted to the data from Phase 1 and Phase 3 respectively (see Table 3).

Table 3. Summary of LMM analyses of inverse-transformed naming latencies.

Both Phases

	β	*SE*	*t*	*p*
Intercept	-10.43	0.23	-45.32	<.001
Phase	0.49	0.27	1.84	.082
Name Agreement	-2.62	0.35	-7.44	<.001
Name Change	0.65	0.09	6.95	<.001
Phase*Name Agreement	3.57	0.51	6.98	<.001
Phase*Name Change	1.39	0.16	8.75	<.001
Name Agreement*Name Change	0.94	0.38	2.46	.022
Phase*Name Agreement* Name Change	0.31	0.66	0.48	.64

Phase 1 only

	β	*SE*	*t*	*p*
Intercept	-10.67	0.27	-40.28	<.001
Name Agreement	-4.42	0.52	-8.50	<.001
Name Change	-0.03	0.10	-0.35	.73
Name Agreement* Name Change	0.74	0.48	1.55	.12

Phase 3 only

	β	*SE*	*t*	*p*
Intercept	-10.19	0.27	-38.20	<.001
Name Agreement	-0.82	0.33	-2.52	.015
Name Change	1.34	0.15	8.95	<.001
Name Agreement* Name Change	1.16	0.50	2.30	.023

As detailed in Table 3 and illustrated in Figure 2, significant main effects of Name Agreement and Name Change indicate that speakers were in general faster to name pictures with higher name agreement ($\beta_{NameAgreement}$ = -2.62, *SE* = 0.35, $p < .001$) and pictures with confirmed rather than

corrected names ($\beta_{\text{NameChange}}$ = -0.65, *SE* = 0.09, $p < .001$). A marginally significant main effect of experimental Phase suggests that speakers were, overall, somewhat slower after the corrective familiarization procedure than before (β_{Phase} = 0.49, *SE* = 0.27, $p = .082$). A significant interaction between Name Change and experimental Phase ($\beta_{Phase*\text{NameChange}}$ = 1.39, *SE* = 0.16 $p < .001$) confirms that the slowing associated with the Name Change predictor specifically emerged after the corrective familiarization procedure (Phase 1: $\beta_{\text{NameChange}}$ = -0.03, *SE* = 0.10, $p = .73$; Phase 3: $\beta_{\text{NameChange}}$ = 1.34, *SE* = 0.15, $p < .015$). In contrast, a significant interaction between Name Agreement and experimental Phase ($\beta_{\text{Phase*NameAgreement}}$ = 3.57, *SE* = 0.51, $p < .001$) indicates that Name Agreement effects were much stronger before corrective familiarization (Phase 1: $\beta_{\text{NameAgreement}}$ = -4.42, *SE* = 0.52, $p < .01$) than after (Phase 3: $\beta_{\text{NameAgreement}}$= -0.82, *SE* = 0.33, $p = .015$). Although a significant two-way interaction between Name Agreement and Name Change ($\beta_{\text{NameAgreement*NameChange}}$ = -0.94, *SE* = 0.38, $p = .022$) would be consistent with the idea that the Name Change instruction caused greater slowing for pictures with higher name agreement than those with lower name agreement, the corresponding three-way interaction between Phase, Name Agreement, and Name Change—which provides a more specific test of that claim—did not approach significance ($\beta_{\text{Phase*NameAgreement*NameChange}}$ = 0.31, *SE* = 0.66, $p = .64$), thus hindering that interpretation.

To consider the sources of the patterns in further detail, we estimated ex-Gaussian components for each subject in each condition (Table 2.), and then submitted the resulting components to linear mixed regressions that were analogous to those above but with the random effects structure appropriately simplified to include only per-subject random intercepts (full results are provided in Appendix B). Instead of transforming the RT data to approximate a normal distribution, an ex-Gaussian analysis attempts to describe an RT distribution as a convolution of a normal (i.e., Gaussian) distribution and an exponential distribution. It thus estimates, for each set of observations, a μ and σ, describing the mean and standard deviations, respectively, of the normal distribution, and a τ, describing the contribution of the exponential distribution. Factors that affect most trials in the same

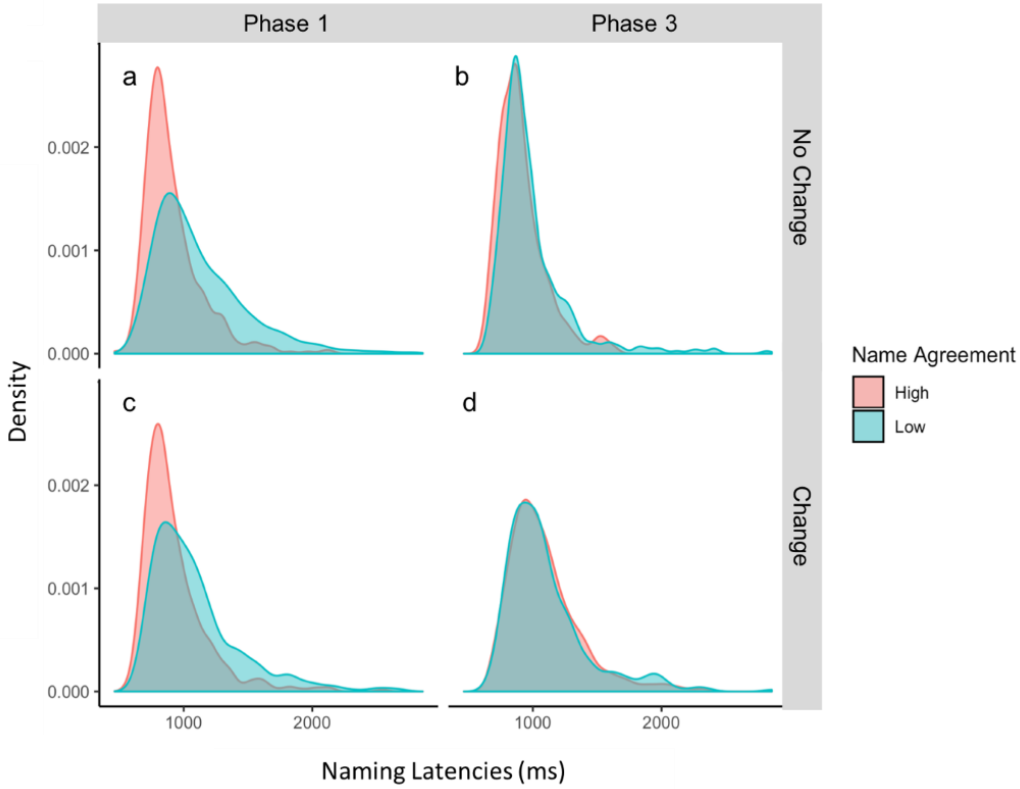

Figure 2. Density plots of naming latencies for high and low Name Agreement (categorical) pictures in free naming (Phase 1) and post-correction target naming (Phase 3) (vertical) and Name Change and No Change (horizontal). To allow direct comparison between Change and No Change conditions, items depicted in Phase 1 are back-sorted according to their Phase 3 status (Change-No Change).

way-and are thus most crucial for most theories of typical processing-can be expected to affect the μ component. Results of the analyses for the μ component (Table B1) largely concur with the results from the inverse-transformed method, except that in the Phase 3 sub-analysis only the Name Change predictor reaches significance ($\beta_{\text{NameChange}}$ = 103.80, SE = 21.41, p < .001) and the Name Agreement predictor trends in the direction of faster responses for low-agreement pictures ($\beta_{\text{NameAgreement}}$ = 83.00, SE = 54.30, p = .13; other ps > .30). A Phase 3 Name Agreement x Name Change interaction reached significance only for the σ component (Table B2; $\beta_{\text{NameAgreement*NameChange}}$ = 166.69, SE = 59.77, p = .008; other ps > .30), though still without the relevant three-way interaction in the overall analysis ($\beta_{\text{Phase*NameAgreement*NameChange}}$ = 113.46, SE = 108.59, p = .30), while a Phase 3 main effect of Name Agreement reached significance in the expected direction only for the τ component (Table B3; $\beta_{\text{NameAgreement}}$ = -201.73, SE = 48.35, p < .001; other ps > .20).

In sum, the response time data suggest that, after corrective familiarization, the initially robust effect of picture name agreement is largely replaced by a name change cost, with only limited support for the hypothesis that suppressing a high-agreement name should be especially difficult.

4.3.2 ERP results

ERP analyses used linear mixed effects regressions to predict, for the same trials as the naming latency analyses, the mean amplitudes in specified time windows, per participant, per condition, as a function of the centered binomial predictors and interactions of Name Agreement {Low = -0.5, High = 0.5}, Name Change {No Change = -0.5, Change = 0.5}, and Phase {Phase 1 = -0.5, Phase 3 = 0.5}. This approach follows traditional ANOVA ERP analyses but provides meaningful estimates of the magnitude of each effect.

4.3.2.1 Free Naming (Phase 1)

ERPs elicited by pictures in the initial "free naming" phase (Figure 3) showed a P1-N1-P2 complex that is typical of responses to visual stimuli. This included a P1 that peaked at 100 ms followed by an N1 peaking at 150 ms, and a P2 peaking at 200 ms over parieto-occipital areas of the scalp. This P1-N1-P2 complex was followed by an N400 peaking at around 400 ms over somewhat more central areas of the scalp. As expected, the posterior P2 was reversed in polarity over frontocentral regions of the scalp, manifesting as an N200, which peaked at around 210 ms.

In the N200 range, over a set of frontocentral electrodes that have previously been shown sensitive to this factor (FCZ, FC2, FC4, FZ, F4; Figure 3), linear mixed effect effects regressions confirmed that pictures with low Name Agreement evoked significantly more negative mean N200 amplitudes than those with high Name Agreement ($\beta_{\text{NameAgreement}}$ in Phase 1 = 0.61 μV, SE = 0.23, p = .016; see Appendix C for full regression tables). Similarly, in the N400 range, pictures with low Name Agreement evoked significantly more negative ERP amplitudes than those with high Name Agreement

($\beta_{\text{NameAgreement}}$ in Phase 1 = 0.94 µV, *SE* = 0.20, $p < .001$), assessed over 6 frontocentral electrodes that are the typical foci of N400 effects in the literature (FCZ, CZ, FC1, C1, FC2, C2; Figure 3). The modulations in these ranges appeared to have similar scalp distributions (Figure 3) and the magnitudes of participants' effects were strongly correlated (r = .68, p = .003).

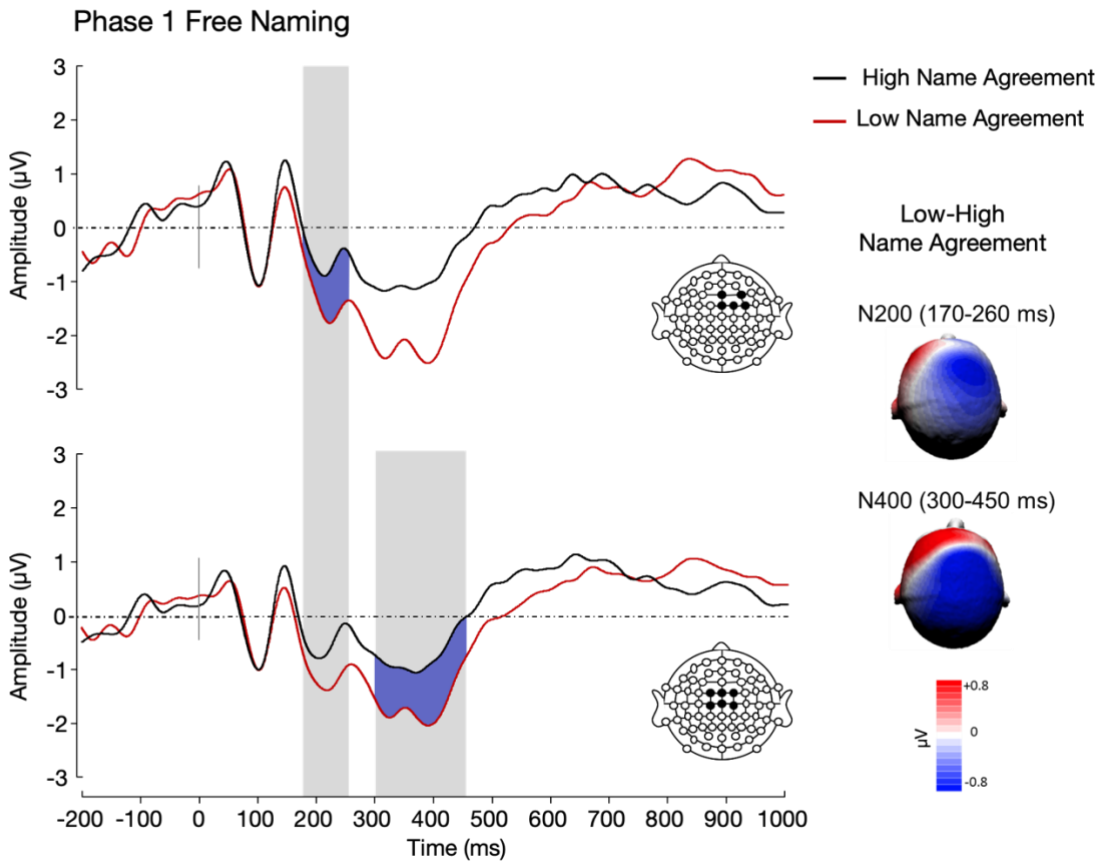

Figure 3. Event-related brain potentials elicited by high and low Name Agreement pictures in Phase 1 (Free Naming), including linear derivation of the electrodes FCZ, FC2, FC4, FZ, F4 for the N200 and linear derivation of electrodes FCZ, CZ, FC1, C1, FC2, C2 in the N400 range. Shaded areas highlight the time windows of analysis for the N200 and ERP amplitudes in the N400 range, respectively. Topographies show the scalp distribution of differences in potential between low and high Name Agreement conditions.

4.3.2.2 Post-correction target naming (Phase 3)

Following corrective familiarization, name agreement-related differences in Phase 3 appeared to persist in the N200 and N400 time windows, though compared to those in Phase 1 both effects were numerically weaker and associated with different, more frontal topography (Figure 4). The main effect

of Name Agreement on mean N200 amplitude trended in the same direction as in the previous 'free naming' phase, over the same frontocentral electrodes, but did not reach significance ($\beta_{\text{NameAgreement}}$ in Phase 3 = 0.41 μV, *SE* = 0.25, *p* = .12). A model comparing the N200-window effects in the two phases suggested a nonsignificant reduction ($\beta_{\text{NameAgreement * Phase}}$ = -0.20 μV, *SE* = 0.34, *p* = .56).

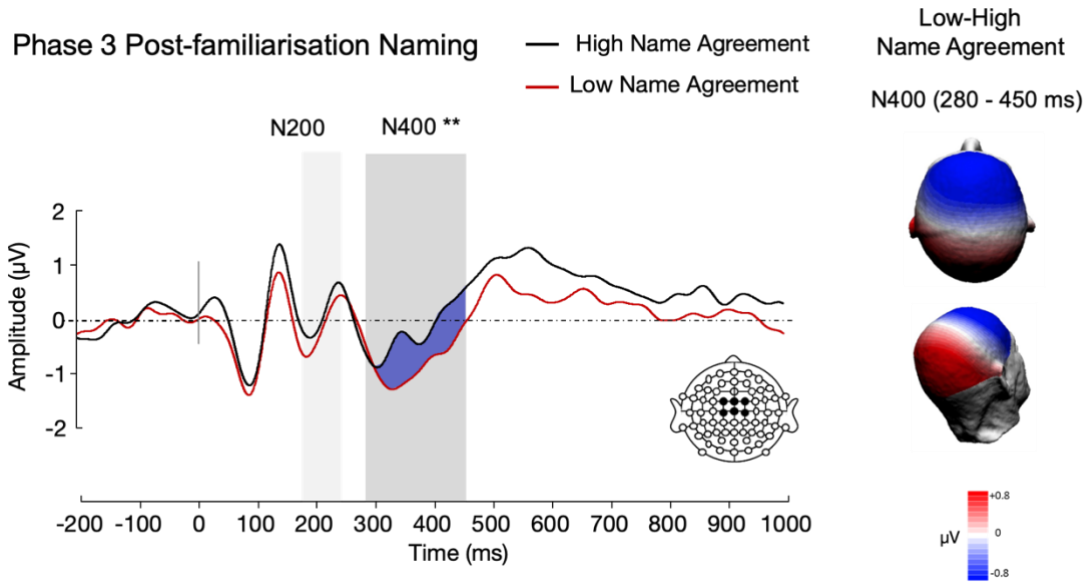

Figure 4 - Event-related brain potentials elicited by high and low Name Agreement pictures in Phase 3 (post-correction target naming), including linear derivation of the electrodes FC1, FCZ, FC2, C1, CZ, C2. Shaded areas highlight the time windows of analysis for the N200, the N400 and the later positive wave, respectively. Dark grey areas highlight the time window were differences between conditions were statistically significant. Topographies show the scalp distribution of differences in potential between low and high Name Agreement conditions in Phase 3.

The Name Agreement effect in the N400 window also trended in the same direction as in Phase 1, and though the effect was also numerically weaker than that in Phase 1 it did reach significance in Phase 3 ($\beta_{\text{NameAgreement}}$ in Phase 3 = 0.54 μV, *SE* = 0.24, *p* = .035; Figure 4); a model comparing the N400-window effects in the two phases suggested a nonsignificant reduction ($\beta_{\text{NameAgreement * Phase}}$ = -0.39 μV, *SE* = 0.29, *p* = .17). Unlike in Phase 1,the magnitudes of participants' Phase 3 effects in these ranges were not significantly correlated (*r* = -.05, *p* = .84).

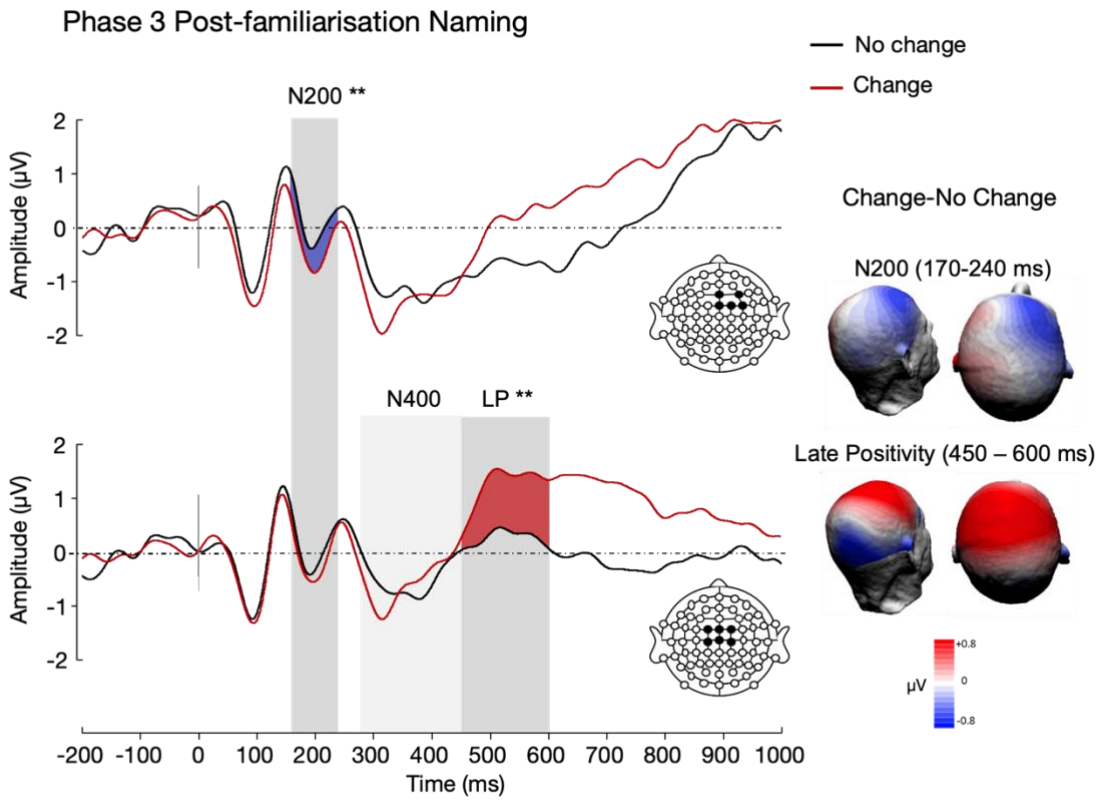

Figure 5 - Event-related brain potentials elicited by Name Change in Phase 3 (post-correction target naming). Top: linear derivation of the electrodes FCZ, FC2, FC4, FZ, F4. Bottom: linear derivation of electrodes FC1, FCZ, FC2, C1, CZ, C2. Shaded areas highlight the time windows of analysis for the N200 and N400 and the later positive wave, respectively. Note that dark grey areas highlight the time window were differences between conditions were statistically significant. Topographies show scalp distributions of differences in potential between Change and No Change conditions in Phase 3.

The Phase 3 ERP analysis also identified a marginally significant effect of Name Change in the N200 time window, assessed over the same frontocentral electrodes as the N200 Name Agreement effect, such that the process of retrieving a coerced name evoked more negativity than that of retrieving a previously volunteered name ($\beta_{\text{NameChange}}$ in Phase 3 = -0.44 µV, *SE* = 0.22, *p* = .061; Figure 5). Compared to the N200 Name Agreement effect in Phase 1 and trend in Phase 3, the topography of this Name Change effect appeared to be more frontal and more lateralized to the right. Name Change did not appear to modulate activity in the N400 window ($\beta_{\text{NameChange}}$ in Phase 3 = -0.02 µV, *SE* = 0.27, *p* = .93; Figure 5), nor did we detect significant interactions between Name Agreement and Name Change in either the N200 range ($\beta_{\text{NameChange*NameAgreement}}$ in Phase 3 = -0.30 µV, *SE* = 0.54, *p* = .58) or the N400 range,($\beta_{\text{NameChange*NameAgreement}}$ in Phase 3 = 0.54 µV, *SE* = 0.39, *p* = .19).

We did, however, identify a later positive modulation associated with Name Change, assessed over the same electrodes as for ERP modulations in the N400 window, from 450-600 ms post picture onset ($\beta_{\text{NameChange}}$ = 1.01 μV, *SE* = 0.41, *p* = .027; Figure 5). The amplitude of this later positive modulation was strongly positively correlated with those in both the N200 (*r* = -.66, *p* = .004) and N400 windows (*r* = .77, *p* < .001)—such that stronger negative modulations in the earlier peaks were associated with attenuation of the later positive modulation—but notably those earlier modulations were not strongly correlated themselves (*r* = .34, *p* = .18).

In sum, before corrective familiarization, Name Agreement was associated with significant and correlated modulations in the N200 and N400 windows. There was some indication that these Name Agreement effects persisted after corrective familiarization, though numerically weaker, no longer correlated, and associated with a different topography than they were earlier. Name Change did not significantly interact with Name Agreement in either window, but was instead associated with an independent, marginally significant modulation in the N200 window and a later frontocentral positive effect.

4.3.2.3 Association between behavioral and electrophysiological effects

In Phase 1, the magnitudes of participants' Name Agreement effects for naming latencies (low minus high) did not significantly correlate with the corresponding differences in either their N200 effects (*r* = -.18, *p* = .49) nor their N400 effects. (*r* = -.28, *p* = .27). In Phase 3, only the correlation between the magnitude of the naming latency effect for Name Agreement and that of the corresponding N200 effect approached significance (*r* = -.48, *p* = .054); the analogous correlation for Name Change was not significant (*r* = -.03, *p* = .91). No correlation between naming latency effects and either N400 modulations or the late positive approached significance (all |*r*| < .3, all *p* > .25).

4.4. Discussion

When and how do lexically specific goals modulate the word production process? Do they affect early processes associated with a word's initial activation and selection, or later processes such as re-evaluation and post-retrieval monitoring? Previous studies have associated differences in name agreement, an empirical measure of response variability in picture naming, with robust differences in naming latencies and ERP effects, suggesting that they index endogenous lexical-level competition (e.g., Alario et al., 2004; Shao et al., 2014; Valente et al., 2014), that is, the same kinds of conflict and decision processes that underlie word selection in typical communicative speech. However, many recent studies with name agreement manipulations have included a pre-experiment familiarization phase, instructing participants to use norm-assessed dominant names before naming each picture (ibid), introducing a confound by contradicting speakers' previously established preferences for low-agreement items (Balatsou et al., in revision). Thus, instructing participants to use particular names for each picture likely introduces a form of exogenous competition that is not inherent to lexical selection per se.

Here, we have sought to determine how the word production system handles such exogenous competition: Do speakers engage an additional mechanism, separate from that of lexical selection, to comply with such task-induced demands (multi-factor account) or do they resolve both endogenous lexical conflict and exogenous lexical competition via a single flexible mechanism (single-factor account)?

4.4.1 Behavioral and electrophysiological effects of endogenous conflict in picture naming

Phase 1 of this experiment assessed the behavioral and electrophysiological effects of name agreement in a simple picture naming task without feedback or prior familiarization; every response was implicitly accepted as correct, thus avoiding any exogenous interference with speakers' pre-established word preferences. A regression analysis of the resulting naming latencies replicated the

classic behavioral observation that speakers name pictures with high name agreement faster than those with low name agreement (e.g., Bates et al., 2003; Snodgrass and Yuditsky, 1996; Ellis and Morrison, 1998). Such name agreement RT effects have often been claimed as evidence for endogenous lexical competition (e.g., Alario et al., 2004), on the assumption that the co-activation of alternative names for pictures with low name agreement (e.g., "sofa", "settee") would delay the retrieval of the target (e.g., "couch"). This competitive account has been empirically challenged, however, by studies focusing on secondary name agreement. Oppenheim (in prep.), for instance, replicated the classic dominant name agreement effect on naming latencies but further demonstrated a facilitation rather than an inhibition of dominant name retrieval latencies for pictures with higher secondary name agreement—that is those with stronger 'competitors'. In any case, we can assume that any name agreement effects in this first phase reflect a simple baseline use of the production system, with any conflict attributed to endogenous rather than exogenous sources.

In this baseline setting, our ERP data showed larger amplitudes in the in the N200 and N400 windows when participants named pictures with low name agreement than pictures with high name agreement. Though the modulation in the N400 range was less expected, N200 effects of name agreement have been reported previously in a simple covert picture naming task (Cheng et al., 2010) and in an overt picture naming study that included prior familiarization (Shao et al., 2014; but cf. Valente et al, 2014). N200 effects are commonly associated with lexical activation and selection in word production (e.g., Indefrey & Levelt, 2004), and have been argued to reflect conflict monitoring in general, regardless of whether this involves selecting or suppressing a particular response (Nieuwenhuis et al., 2003; Donkers & van Boxtel, 2004). In experiments where naming follows corrective familiarization, stronger claims about N200 effects specifically indexing inhibitory processes—either general (e.g., withholding response in a go/go no task, Sanoudaki and Thierry, 2015) or selective (e.g., attempting to retrieve "couch" instead of "sofa", Shao et al. 2014)-might seem plausible: experimenters have demanded that participants avoid using certain words. But without

corrective familiarization (i.e., in Phase 1 of the current study), there is little reason to expect that speakers would need to inhibit specific responses (why would a speaker specifically need to retrieve 'couch' instead of 'sofa'?). It therefore seems more plausible that this N200 modulation in simple picture naming more broadly reflects early differences in the process of lexical access—such as simple co-activation, selection processes (competitive or noncompetitive), or conflict detection—an interpretation that is consistent with other ERP studies of picture naming that have pointed to lexical access as starting approximately 200 ms after picture onset (Strijkers et al., 2010; Costa et al., 2009).

We also detected a name agreement-based difference in the N400 range. While its correlation with and topographical similarity to the earlier N200 modulation suggest some caution in its interpretation, its numerically greater magnitude suggests a distinct but tightly linked process that it is not reducible to spillover from the earlier effect. To our knowledge, this would be the first report of N400 amplitude modulation by name agreement (but cf. Valente et al., 2014, discussed below). One might be skeptical that name agreement should modulate N400 amplitude, given the traditional interpretation of the associated component as reflecting difficulty in resolving semantic incongruities between a word and its presentation context (Kutas and Hillyard, 1980). In line with broader characterization of the N400 component (Kutas and Federmeier, 2011), however, we interpret this difference as reflecting an automatic spreading of lexical-semantic activation, under the assumption that pictures that elicit a variety of names (and therefore presumably have more semantic-lexical associations) should lead to greater activation spreading in the semantic system than pictures that reliably elicit a single name. This interpretation is consistent with N400 modulations that have been reported when naming difficulty increases as a consequence of deeper semantic processing in other word production tasks (Schendan & Kutas, 2007; Schendan & Maher, 2009) and in picture-word interference studies (Blackford et al., 2012; Piai et al., 2012; Shitova et al., 2017; Wong et al., 2017; Python et al. 2018).

In sum, in our baseline "endogenous conflict" condition, lower name agreement was associated with longer naming latencies, and correlated modulations in N200 and N400 windows, in the absence of significant correlations between ERP amplitude modulations and difference in naming latencies between conditions.

4.4.2 Exogenous competition and its possible interactions with endogenous conflict

So how does the production system cope with the imposed goal to produce a specific word, and to what extent does this additional demand affect the endogenous processes of lexical co-activation and selection? We predicted that, with an integrated lexical selection mechanism responsible for resolving both endogenous and exogenous lexical co-activation (single-factor account), the new task goal should modulate and even amplify similar behavioral and ERP patterns after corrective familiarization, whereas the recruitment of a secondary mechanism to resolve a task-induced conflict (multi-factor account) should primarily engage distinct processes, possibly including later control mechanisms.

Following corrective familiarization, the naming latency effects of name agreement were largely replaced by those of name change. A significant Phase x Name Agreement interaction for naming latencies indicates that the effects of name agreement were substantially diminished after corrective familiarization (cf. Mitchell, 1989): High name agreement pictures were still named faster than low name agreement pictures on average, but analyses of response time distributions suggest that this effect was linked to differences in the tail (τ) while the contrast in modal response times (μ) trended in the reverse direction (cf. Roelofs & Piai, 2017; Scaltritti, Navarrete, & Peressotti, 2015). At the same time, speakers were slower to name pictures using 'coerced' names (per the name change manipulation) than names that followed their pre-existing preferences. Although the apparent Name Agreement x Name Change interaction that emerged in Phase 3 would seem consistent with the idea that shifting towards a more dominant name is easier and less time consuming for speakers than using

an undesired non-dominant competitor (cf. Balatsou, Fischer-Baum, & Oppenheim; in revision)-potentially supporting the Single-factor account via additive factors logic-the lack of a crucial Phase x Name Agreement x Name Change interaction cautions against such a strong interpretation. Thus, in Phase 3, the exogenous competition imposed by name change dominated participants' behavioral patterns and essentially replaced (or overshadowed) the effects of the existing endogenous lexical conflict, clearly contrasting with the previous claim that name agreement effects are robust to any effects of corrective familiarization (Alario et al, 2004). More theoretically, if the same lexical selection mechanism were responsible for both resolving endogenous co-activation and ensuring the selection of a task-appropriate word, then we would expect name agreement effects to be stronger in Phase 3 than in Phase 1 (following Nozari & Hepner's claim that task goals amplify name agreement effects), but instead a Phase x Name Agreement interaction showed that they were significantly weaker. Thus, we suggest that the naming latency results are more compatible with the view that speakers adapt to such task goals by adopting an ad hoc strategy of re-evaluating and possibly correcting their responses after at least a substantial chunk of lexical selection has already occurred.

After corrective familiarization, electrophysiological correlates of name agreement differences appeared somewhat weaker, with more frontal topographies, but did not disappear. Reduced N200 amplitude modulation could be attributed to differential repetition priming and error proportional learning, which may also explain the apparent absence of name agreement effects in this range in Valente et al's (2014) post-familiarization study. Also, note that we did detect a marginally significant association between the per-subject magnitudes of the N200 name agreement effect and the corresponding naming latency effect, with stronger N200 modulations associated with stronger naming latency effects. A tighter link between processing in the N200 window and observable behavior, in response to changing task demands, would seem consistent with the spirit of Nozari & Hepner's (2019) specific single-factor proposal (viz, that an increased decision criterion strengthens the link between internal lexical conflict and observable behavioral effects).

The N400 modulation by name agreement that we had observed before corrective familiarization remained thereafter, albeit weaker in magnitude and no longer correlated with the preceding N200 modulation. Assuming a common locus for the Phase 1 and Phase 3 effects, observing this modulation in both phases implies that it was not resolved by the introduction of coerced names, so it is surprising that previous ERP studies of name agreement have not reported similar effects. Considering the paucity of ERP studies of the phenomenon, and the numerous differences between both their tasks and analyses (e.g., Shao et al., 2014; Cheng et al., 2010) , it is possible that their N400-like modulations were either masked by other reported effects or simply not assessed. For instance, Shao et al.'s (2014, Figure 3a) waveform plot for object naming suggests a second modulation beginning around 300ms post-stimulus onset. It is also worth noting that Cheng et al.'s (2010) used a rather late window for their N2 analyses (250-350ms), so it seems possible that their study may have included the N400 in their N2 analysis; supporting this claim, we note that Cheng et al.'s described their N2 as maximal in a left parietal cluster, which would be more typical of an N400. Note also that Valente et al. (2014), using a different analytical approach involving spatiotemporal map segmentation, did identify a negativity associated with picture name agreement beginning at around 380 ms after picture onset that may relate to modulations in the N400 range.

Finally, the requirement to change names modulated ERPs independently from name agreement in Phase 3, such that N200 amplitudes were significantly more negative for coerced names and amplitudes of a late anterior positivity were significantly more positive. We assume that suppressing one's preferred name to produce an alternative—as corrective familiarization demands—should strongly engage the selective inhibition processes that some have suggested underlie name agreement effects, so the N200 peak should reflect the involvement of this process. The lack of a significant interaction between Name Change and Name Agreement predictors in this window, however, suggests that their N200-range modulations index distinct processes. For directed name change, this early modulation was followed by and correlated with a late anterior positive modulation.

That correlation is notable because it emerged in the absence of a strong correlation between the modulations in the N200 and subsequent N400 range, suggesting that the late effect indexes a linked but separate process. Such late anterior effects are often associated with stimulus re-evaluation (e.g., P600), cued recall, domain-general conflict monitoring (error-related negativity or positivity e.g., Nozari, Dell, & Schwartz, 2011), or other applications of executive control. We suggest that such re-evaluation reflects the recruitment of adjunct cognitive resources and response monitoring or correction processes, which may be central to many laboratory tasks but less relevant for normal communicative word production.

4.4.3 Summary evaluation of the single factor and multifactor accounts

Decades of research have assumed that the challenges that task-driven exogenous competition imposes on the production system can reveal aspects of its generic properties. This view assumes that that core aspects of lexical selection can be directly adjusted to incorporate task demands, such as using a particular word instead of its synonym or naming a picture instead of a superimposed distractor, essentially assuming cognitive penetration of the core (early) lexical selection mechanism (single-factor account). Despite the usefulness of such proposals, the consistent dissociations between endogenous conflict and exogenous competition in this study (i.e., the direct name change manipulation largely replaced name agreement RT effects and simply added separate early and late ERP modulations) seem more consistent with the idea that imposing such task-driven constraints requires the contributions of at least two distinct early processes, in addition to a later re-evaluation or monitoring process (multi-factor account).

The strongest evidence for the single-factor account comes from the fact that both endogenous conflict and exogenous competition modulated early processes, manifesting through N200 differences, though they did so quite independently. That is, the predictors did not significantly interact to predict early modulations, and contra a specific prediction from Nozari & Hepner (2019) imposing a task goal

of producing a specific word, via corrective familiarization, quashed the name agreement response time effect instead of exaggerating it. It may however be possible to explain away some of these results. For instance, although it contradicts an explicit prediction, one might expect some decrease in name agreement effects after corrective familiarization as a simple consequence of differential repetition priming (e.g., Mitchell, 1989) and error-proportional learning (e.g., Oppenheim et al., 2010), as well as a general attenuation of N200 amplitude (Llorens et al, 2014). Additionally, the fact that a correlation between per-subject N200 and naming latency effects of name agreement approached significance only after corrective familiarization-and perhaps even the reduction in these effects-could be taken as circumstantial evidence for some connection between processes involved in resolving exogenous and endogenous lexical conflict. Moreover, single-factor accounts typically embrace the inclusion of monitoring and even repair mechanisms in the production process, but simply argue that they are not the primary locus of competitive effects. In Nozari et al.'s (2011) production-internal monitoring proposal, for instance, a late error-related negativity is argued to reflect conflict detection, but the source conflict is assumed to originate earlier, and may serve as a signal for a further repair process (e.g., Nozari et al, 2016).

Conversely, the strongest evidence supporting the multi-factor account is the broad dissociation and lack of interaction between endogenous conflict and exogenous competition in early ERP modulations and summative response time measures. These dissociations suggest that introducing an undesired competitor engaged an additional mechanism, qualitatively different from that involved in typical lexical selection to resolve the ongoing conflict. Although both name agreement and name change were associated with modulations in the N200 time window, without interactions it is difficult to claim that the associated early cognitive processes of lexical activation and selection and selective inhibition or competitive re-thresholding were directly integrated. As discussed above, the later anterior component that emerged for name change after corrective familiarization can broadly be characterized as indexing a late control process-whether or not it plays an essential role in this task-

and under a multifactor account we might associate it more specifically with the process of ensuring the application of ad hoc constraints. Such monitoring could be accomplished by a conflict detection mechanism or may even take the form of a "watchful little homunculus" (Levelt, 1989) or a post-monitoring pre-articulatory control mechanism that is functionally distinct from typical lexical selection (Finkbeiner & Caramazza, 2006, Mahon et al., 2007; Nozari et al, 2016). A key question then is what role, if any, this monitoring or correction process plays in typical communicative speech (e.g., Oppenheim & Balatsou, 2019). Considering the distinction between the response profiles of simple and targeted picture naming, we suggest that, while everyday communication potentially incorporates multiple constraints, exogenous experimental manipulations that strongly engage adjunct control processes are less likely to be informative about the core processes involved in typical word production.

Before concluding, we should address a possible objection to our presentation of these results as a test of single factor accounts, namely that our name-change manipulation changes the goal of the Phase 3 picture naming task. That is, orthogonally manipulating name agreement and name change means that, in the case of high-NA pictures, one might worry that our task forced participants to use sub-optimal names, fundamentally changing the task compared to an experiment where familiarization always provides the 'best' name in the form of each picture's dominant name. If the goal therefore became to use the specified name instead of the best name, one might argue that this experiment is not strictly relevant to core theoretical claims of single-factor accounts. In fact, we fully agree that our corrective familiarization changed the goal of the Phase 3 naming task, but maintain that the criticism equally applies to the 'dominant name familiarization' studies that Nozari & Hepner's (2019) flexible criterion proposal explicitly embraced as evidence of competitive selection (as well as other paradigms with clearly arbitrary goals, e.g., picture-word interference). From that perspective, this study can be seen as identifying appropriate limits for the scope of the proposal. Because speakers maintain stable preferences even for non-dominant names (Balatsou et al., under review), requiring participants to

change to a dominant name will nonetheless impose a cost (particularly for low-agreement items), meaning that naming following corrective familiarization is already a 'choose the specified word' task instead of a 'choose the best word' task. And we have previously identified similar within-item corrective familiarization-based name change costs in a behavioral task where we used all dominant names and merely tracked these costs rather than manipulating them (Oppenheim, 2014). The difference is only that, without orthogonally manipulating name change, one cannot measure the name change cost for high-agreement items and may therefore mistake exogeneous competition for endogenous conflict. More generally, a major theme in this article is that seemingly innocuous, pragmatically motivated practices, like pre-experiment corrective familiarization can deeply affect the way a participant approaches a laboratory task, rendering the resulting data less relevant to the theories that researchers wish to build or test.

4.5 Conclusion

So, how do speakers select an appropriate word for production? The answer may depend on the precise nature of the task. Speakers can accurately, and often quickly, adjust their observable behavior. The inherent needs of communication clearly require broad adjustments like applying pre-lexical semantic control to craft an appropriate message, incorporating language and word class constraints to select syntagmatically appropriate words, and imposing speed/accuracy trade-offs to maintain fluency (whether these affect sensitivity to coactivation or merely evidence accumulation). However, when tasks require participants to impose ad hoc constraints, like using a specified word instead of a synonym or naming a picture instead of a superimposed distractor, they may do so by engaging adjunct mechanisms that are less critical to production in the wild. Our evidence indicates that speakers employ distinct mechanisms to address such ad hoc constraints via both early and late processes. We suggest that key evidence for the competition-based account of lexical selection instead reflects the contribution of these adjunct mechanisms. Remaining questions include what if any role

such mechanisms play in typical communicative language production and how precisely they might integrate with the core system.

CHAPTER 5- Robust effects of picture name agreement for stable word preferences

This chapter investigates the electrophysiological and behavioral effects of picture name agreement in consistent responses across speakers. In a repeated naming task, I report robust effects of name agreement for both naming latencies and ERPs in respect to individuals' idiolects, which include both the dominant (e.g., "couch") and secondary (e.g., "sofa") names. These modulations indicate that picture name agreement is a valid measure of lexical co-activation but do not suggest response competition. The results are also discussed in respect to theories of word production.

5.1. Introduction

While models of word production generally agree that a series of necessary processes have to take place before speaking (i.e., conceptualization, formulation and articulation) (e.g., Levelt, et al., 1999; Dell, 1986), their major difference lies in the nature of the mechanism that determines how one of the appropriate co-activated candidate words will be eventually selected. Competitive models of lexical selection suggest that the activation of other candidates will slow down the selection of the target word, which will eventually be chosen after it passes a *relative* threshold (Levelt et al. 1999; Roelofs, 1992), while non-competitive accounts instead argue that the target word will be selected when it reaches an *absolute* threshold, irrespective the relative activation of the other words in the mental lexicon (Mahon et al., 2007). A second point of debate lies in the dynamics of these encoding processes: one view is that the processing levels are discrete and serial and the activation from one level to the other is only feed-forward (Levelt et al., 1999), while an alternative, cascading interactivity hypothesis states that activation cascades within the system, also in the form of feedback processing (Dell, 1986).

While historically the computational principles of seriality and competition have dominated interpretations in the literature, nowadays it is inarguably accepted that at least some cascading activation or interactivity are inherent to the production system (see Dell, Nozari, & Oppenheim, 2014, for a review of the evidence in favor of interactivity), while the competition debate is also subject to re-examinations (e.g., Nozari & Hepner, 2019). To date the major evidence in favor of the lexical competition hypothesis derives from the behavioral effects in picture word interference tasks, in which participants are slower to produce the target word for the picture (e.g., “cat”) in the context of semantically related visual or auditory distractors (e.g., “dog”) compared to those that have no semantic relationship with the target (e.g., “pencil”) (e.g., La Heij, 1988; Schriefers et al., 1990). By increasing the levels of co-activation in this simultaneous processing, it is assumed that the levels of lexical competition in production increase in a similar fashion, as indicated by naming latency analyses

(Belke et al., 2005; Bloem & La Heij, 2003; Bloem et al, 2004). This interpretation, grounded on the notion of a time-dependent selection mechanism, and not on the cost of increased perceptual or semantically-driven lexical co-activation which is externally-induced, has been severely challenged by the finding that as distractor words become more semantically close to the target (e.g., "zebra" to "horse") production latencies become faster (Mahon et al., 2007), meaning that as "competition" becomes stronger, the production speed is instead facilitated.

Driven by this observation and by the fact that competitive effects in other naming tasks have been successfully modeled with non-competitive criteria (e.g., Howard et al.2006 vs Oppenheim et al., 2010, for lexical competition in cumulative semantic interference), it is worth considering whether competition during selection, as the one reported in picture-word interference, is inherent to simple word selection or merely reflects task-oriented production (Nozari & Hepner, 2019). We have recently argued that it is also crucial to distinguish between laboratory-reported effects and the extent to which they account for the processes that are inherent to communicative word production and re-evaluate the competition hypothesis through empirical findings in simple production tasks, instead of manipulating competition levels in psychologically challenging experimental designs (Oppenheim & Balatsou, 2019). One such approach is collecting picture naming norms within and across languages, which involves simple picture naming (e.g., Bates et al. 2003; Szekely et al., 2004). In norms, researchers note the primary names given for each picture by the participants and then analyze the variables associated with those names in relation to theoretical predictions about word production processes. One of the most widely examined variables is picture name agreement, the empirically derived measure of the proportion of speakers who produce the same name for a picture. Name agreement variations produce robust effects in retrieval success and speed (Alario et al., 2004) cross linguistically (Bates et al., 2003; Snodgrass and Yuditsky, 1996; Ellis and Morrison, 1998; Szekely et al., 2004; Cuetos et al., 1999; Bonin et al., 2002; Dell'Acqua et al., 2000; Dimitropoulou et al., 2009; Nishimoto

et al., 2012; Bakhtiar et al., 2013) and are observed independently of other linguistic variables, like lexical frequency (Alario et al., 2004).

Apart from being perhaps the most widely used norming measure, name agreement variations in simple naming have been directly interpreted as evidence of endogenous lexical competition during word production (e.g., LaGrone & Spieler, 2006; Shao et al., 2014; Bose & Schafer, 2017). Because pictures with high agreement (e.g., "dog") are generally faster and easier to name compared to those with low name agreement (e.g., "couch"), it is assumed that this lower consensus induces a challenge for the individual speaker during lexical selection, specifically in the level of semantic-to-lexical processing in cases of pictures with multiple alternative words (Alario et al., 2004; see Figure 1.): the robust naming latency effects for pictures with multiple appropriate names has been suggested to index the ongoing conflict between the co-activated lexical representations, which eventually increases the relative threshold of activation, delaying lexical selection and eventually retrieval speed (Levelt et al., 1999). This is the most commonly adopted interpretation of the strong behavioral effects of name agreement, however it is mostly a priori grounded on the assumption of lexical competition in which word selection and corresponding production latency is guided by a stochastic function: the probability of selecting a particular word is determined by the ratio of the activation of that word to that of co-activated alternatives (e.g., Luce, 1959) and when that relative ratio is decreased, as in pictures with multiple or strong competitor words, selection time for the target word increases in the same fashion (Levelt et al., 1999; Roelofs, 1992; 2003; Roelofs & Piai, 2015).

These competition-based interpretations of name agreement behavioral effects have also been used as basis of evaluation of the robust neuroimaging and electrophysiological findings in the literature. Pictures with lower consensus induce greater left inferior frontal gyrus (LIFG) activity before naming, leading towards the interpretation that Broca's area, as a language-sensitive region, mediates the selection among competing lexical representations during production (Kan & Thompson-Schill, 2004; Thompson-Schill, et al., 1997). In ERP research name agreement effects are less

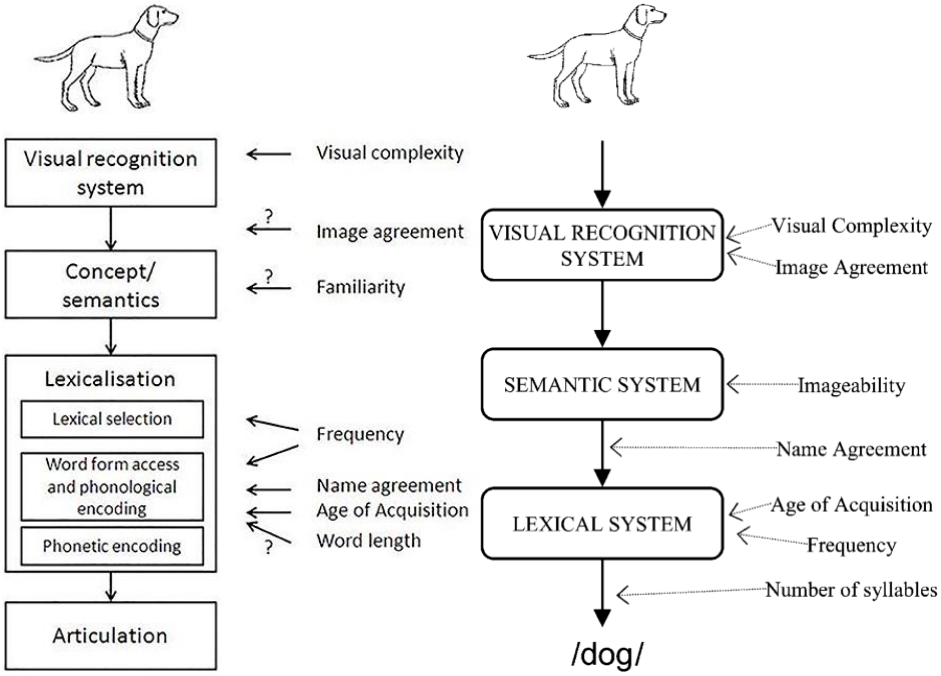

Figure 1). Picture naming models reprinted from a) Valente, A., Bürki, A., & Laganaro, M. (2014). ERP correlates of word production predictors in picture naming: a trial by trial multiple regression analysis from stimulus onset to response. *Frontiers in neuroscience*, 8, 390. and b) Alario, F. X., Ferrand, L., Laganaro, M., New, B., Frauenfelder, U. H., & Segui, J. (2004). Predictors of picture naming speed. *Behavior Research Methods, Instruments, & Computers*, 36(1), 140-155., indicating the most common psycholinguistic factors which affect each specific encoding substage of word production.

consistent: pictures with lower name agreement produce effects in both early time windows and modulations at the P1 and N2 components (see Chapter 4; Cheng et al., 2010; Shao et al., 2014) as well as later differences arising approximately after 400 ms post picture onset, mostly associated with later processing stages, like phonological encoding (Valente et al., 2014; see Figure 1.). Most interpretations are based on Indefrey and Levelt's (2004) assumed timeline of word production processes according to which lexical selection takes place between 175 ms and 250 ms post-stimulus-onset, lexical phonological code retrieval between 250 ms and 330 ms post-stimulus-onset and syllabification between 330 ms and 455 ms post-stimulus-onset. Early ERP effects, those reported in the P1 range, have been interpreted as indexes of visual-to-conceptual ambiguity before any lexical-level processing has taken place (Cheng et al., 2010), which in the competition framework reflects increased shared co-activation between conceptually related representations leading to the automatic co-activation of semantically similar lexical entries (e.g., Belke, 2013). The N2 range effects, in

particular, have been specifically tied to the time window of lexical selection according to Indefrey and Levelt's (2004) timeline and in similar fashion to the fMRI findings, researchers have attributed these differences to a mechanism responsible for conflict resolution. Pictures with lower name agreement (e.g., "couch") are assumed to require the recruitment of response inhibition, a mechanism that suppresses the activated competitor words (e.g., "sofa", "settee"), in order to resolve the ongoing lexical competition and eventually allow the selection the target word (Shao et al., 2014). Although competitive word production models do not explicitly incorporate a response inhibition mechanism (e.g., Levelt et al., 1999; Piai et al., 2014; Roelofs, 1992, 2003, 2008), Shao et al. (2014) have argued that inhibition could complement a stochastic lexical selection process in order to resolve the small differences between the co-activated competitor words and eventually achieve target word selection instead of resulting in an omission.

However, as the empirical finding that increased lexical co-activation facilitates production in picture-word interference tasks (Mahon et al., 2007), the recent finding that increased secondary name agreement (i.e., the measure for the strongest alternative name for each picture) also facilitates production speed (Oppenheim, 2017; in prep.) challenges the core assumption of the competition hypothesis: when the core feature of competition, that is the impact of the strength of lexical co-activation, is independently found to make retrieval speed faster in both more complex and simple naming tasks, then it is reasonable to assume that either word production is not strictly competitive as non-competitive/chronometrically agnostic accounts generally argue (e.g., Mahon et al., 2007; Oppenheim et al., 2010; Dell, 1986), that picture name agreement measures do not directly index variations in levels of lexical competition as generally assumed, or both. Complementary to the latter, comes the recent empirical observation that population-level name agreement, apart from directly reflecting the options that speakers consider during lexical selection, also indexes individual speakers' stable word preferences, which include non-dominant responses (Balatsou, Fischer Baum, & Oppenheim; in revision). For instance, in repeated naming, I found that speakers select the same name

(e.g., “couch”) in two consecutive naming sessions much more frequently than it is predicted by a strictly stochastic selection and despite the equally high probability that the competitor word (e.g., “sofa”) has in population-level norms. Importantly, this tendency to reproduce specific words also holds for non-dominant responses: apart from reproducing strong alternative names (e.g., “sofa”), speakers are also likely to reproduce alternative names that they themselves have previously chosen (e.g., “settee”). We suggested that this is compatible with *idiolects* and demonstrated that name agreement also reflects the heterogeneity in individual speakers’ word preferences.

Because speakers’ idiolects have not been previously considered nor extensively studied in experimental work, it is, thus, worth re-evaluating how this heterogeneity in responses could have modulated some of the previously reported name agreement effects and their competitive interpretations. For instance, the use of a familiarization, which aims is to reduce variance and at the same time increase the power of the analyses, has been previously found to strongly modulate the behavioral effects in picture-word interference studies (Collina et al.,2013; Gauvin et al., 2018). In Chapter 4, I have shown that its use in studies that investigate the effects of picture name agreement as a psychologically meaningful predictor of individuals’ states is problematic: when speakers have preferences for alternative names (e.g., “sofa”), the instruction to use only the norm-assessed dominant names (e.g., “couch”) creates an additional task demand to switch to dispreferred labels, which severely distorts the name agreement effects in simple selection. In contrast, when subsequent production was consistent with speakers’ idiolects, the cost of repetition priming created by overt-naming familiarization (as in Shao et al., 2014) is also another point of consideration, because as also demonstrated in Chapter 4, repeating a previously selected picture name has robust priming effects in both naming latencies and in ERP amplitudes.

Such methodological inconsistencies and conceptual misuses can, thus, explain the lack of consistency in the electrophysiological effects reported in the literature and the subsequent differences in the interpretations of name agreement effects. A point of consideration is whether picture name

agreement variations index the effects of endogenous competition during lexical selection (e.g., Alario et al., 2004; Cheng et al., 2010; Shao et al., 2014) or the dynamics of the production system related to other necessary subprocesses, such as lexical co-activation and/or phonological encoding (e.g., Valente et al., 2014), which, can also be explained without competition. In particular cases where name agreement effects have been directly linked to competitive lexical selection, either as indexing "lemma-level" competition (Cheng et al., 2010) or response inhibition (Shao et al., 2014), investigating these variations in respect to individual speakers' idiolects provides a better measure to assess inherent effects of co-activation in normal production. Concurrently, by accepting as idiolect-appropriate the non-dominant names that speakers consistently produce (i.e., including secondary responses from norms), it is possible to incorporate the measures of *relative activation* of all lexical representations in the population (population-level name agreement for the dominant and secondary names) into speakers' own consistent word choices which have reached the *absolute activation* for selection in their mental lexicons (individual-level name selections for the dominant and secondary names). Using this approach, I believe that it is possible to report name agreement behavioral and electrophysiological effects in a simple and straightforward production task and be able to directly associate them with different encoding processes during word production, without necessarily basing the interpretations of the electrophysiological effects on Levelt and Indefrey's (2004) timeline, which is mostly unsupported by ERP data.

In the present study I, thus, aim to: (1) establish whether and when electrophysiological name agreement effects emerge in simple picture naming, without restricting them to particular components or time windows, (2) observe how these effects differ with repetition, in order to be able to estimate individuals' stable word preferences and at the same time be consistent with previous practices that have included repeated naming (3) evaluate whether the previously reported electrophysiological and behavioral differences between low and high name agreement pictures still emerge in a reduced competition scenario, like that of name consistency. By examining name agreement effects for picture

names which have not previously reached the selection threshold, the accessibility of the "competitor" word is automatically reduced. If the previously reported robust name agreement effects during formulation dissipate, then it is reasonable to assume that the previously reported variations are related to lexical-level response competition. If, however, name agreement effects previously associated with lexical selection, still persist for name consistency, then they likely reflect the cost linked to other encoding processes, like increased target and non-target semantic-to-lexical activation (e.g., Alario et al., 2014) or increased activation of phonological nodes (e.g., Valente et al., 2014).

5.2. Methods

5.2.1 Participants

Twenty-three Bangor University students (16 females; Mean age: 20.6 years, *SD*= 3.3; eighteen right-handed) participated in a two-session picture naming study. All participants were native English speakers and had normal or corrected-to-normal vision, no neurological impairment and no self-reported symptoms of any language disorders. The study was approved by Bangor University Ethics Committee, and participants gave informed consent and received course credit or cash compensation.

5.2.2 Materials

Pictures for the naming task were the 525 black-and-white line drawings of common objects from the International Picture Naming Project (Bates et al., 2003). Standard methods were adapted from Bates et al. (2003). The stimuli were grouped into 5 blocks of 105 pictures each, including one filler at the beginning of each block, followed by 104 experimental items (520 experimental items in total).

5.2.3 Apparatus

Stimuli were presented via E-Prime (v. 1.2.1.847) on a 17" CRT in a soundproof testing booth at Bangor University's Language and Electrophysiology Team (BULET) laboratory. Responses and naming latencies were recorded via a small diaphragm condenser microphone that was positioned approximately 15cm from the participants' mouth, feeding into both a digital recorder and a delayed-threshold voicekey.

5.2.4 Design and Procedure

The experimental design consisted of twenty-three unique sequences of approximately counterbalanced stimulus orders across Sessions and participants. Session 2 of the study took place one to-two weeks apart from Session 1. The experimenter in the testing booth manually recorded the participants' naming accuracy (i.e., whether participants selected the dominant or secondary name for each picture or gave any other response). Each trial began with a small black fixation cross at the center of the screen which appeared for 150 ms. Next, a picture (317 x 317 pixels) appeared at the center of the screen for 3000 ms or until the voicekey was triggered by the participant's voice. Participants were asked to name each picture as quickly and accurately as they could, while refraining from making any other noises. The interstimulus interval time was set at 500 ms. Short self-paced rests followed each 105-trial block. One to two weeks later, each participant returned to repeat the full procedure.

5.2.5 Electrophysiological recording and pre-processing

5.2.5.1 Recording

Electrophysiological data were recorded in reference to Cz at a rate of 1 kHz from 64 Ag/AgC1 electrodes placed according to the 10-20 convention (Jasper, 1958). Impedances were kept below 5 kΩ. Ocular artefacts were monitored using vertical electro-oculogram channels (VEOG) set above and below the left eye. All pre-processing steps and analyses were performed using EEGLAB Toolbox

(version 14.1.2b; Delorme & Makeig, 2004) in MATLAB (v. R2019a, Mathworks Inc.) and Neuroscan (Scan 4.5, Compumedics).

5.2.5.2 Preprocessing

EEG data were filtered bandpass using zero-phrase shift digital filtering (0.1 Hz, 24 dB/oct- 20 Hz, 48 dB/oct) in Neuroscan. Ocular artefacts were mathematically corrected using independent component analysis (ICA) in EEGLAB. The independent components (ICs) were inspected by plotting component activations as well as component spectra and maps to see which ICs contributed the most at 5 Hz and 20 Hz frequencies. ICs containing ocular artefacts were removed. Prior to accepting ICA correction, I plotted the EEG data before and after ICA correction to make sure that rejecting ICs did not impact the data in an adverse way. In total, a mean number of 1.89 ICs (*SD* = .67; min = 1, max = 3) was rejected per participant. The eye channels were then removed from the signal and a visual inspection on the channels was carried out, in order to select the electrodes for interpolation. In total a mean number of 1.09 channels (*SD* = 1.54; min = 0, max = 5) was interpolated per participant. Individual averages were then digitally re-referenced to the global average reference. Continuous recordings were cut into epochs starting 200 ms before picture onset and ending 600 ms after for stimulus-locked ERP analyses. Baseline correction was performed in reference to pre-stimulus activity (-200 to 0 ms) for stimulus-locked analyses. Further visual inspection of individual epochs was carried out and epochs exceeding -100 to 100 μV were automatically rejected from each dataset in a total mean number of 3.47 per dataset (*SD* = 4.93, min = 0, max = 24). Epochs containing contaminated signal were also manually dismissed from each data set (Mean = 20.52, *SD* = 7.43, min = 6, max = 36) and an average number of 3.21 channels (*SD* = 1.86, min = 0, max = 8) were re-interpolated, in order to achieve the best maximum quality of the data.

5.2.6 Analytical Approach

The main analytical approach used followed that of Chapter 3, with the exception that I only considered consistent responses for the behavioral and ERP analysis. Responses were initially transcribed on-line, later confirmed via audio recordings and were also manually checked offline for accuracy. Oppenheim's (in prep.) recent norms from the same population provided dominant and secondary names for each picture. Following the Oppenheim (in prep.) norms, responses that deviated from an expected name only in plurality or the addition of an article (e.g., "toe"/"toes", "boat"/"a boat") were accepted as tokens of that name; possible abbreviated forms (e.g., plane and aeroplane), however, were considered as distinct responses. In cases where a participant produced two or more codable responses in a single trial (e.g., "dog. . . cat"), I analyzed the first. Because a main purpose of the current study was to assess the electrophysiological effects of picture Name Agreement in repeated picture naming, while trying to control for the covariate of voluntary name switching between Session 1 & 2, I only accepted consistent responses across the two Sessions, i.e., trials in which the same (either dominant or secondary) name was selected in both naming instances by each participant.

5.2.6.1 Behavioral

Naming latencies were analyzed with confirmatory mixed effects regression, via the lmer function in the lme4 v1.12 library (Bates et al., 2016) in R (v5.5.1, R Development Core Team, 2016). Naming latencies were inverse transformed, in order to reduce the variability of the data (Whelan, 2008), the fixed effects were centered and contrast-coded, while the model included a maximal random effects structures (Barr et al., 2013) for participants and items, omitting correlations between random effects to facilitate convergence. *P*-value estimations used the Satterthwaite approximation method.

5.2.6.2 Electrophysiological

The aim of the mass multivariate analysis was to determine the time course and nature of picture Name Agreement and experimental Session effects on ERPs. I used picture Name Agreement[6] as a continuous predictor and Session as a categorical predictor for the analyses.

These analyses were conducted using the LIMO toolbox (Pernet et al., 2011) in MATLAB. In LIMO, the data were analyzed using a hierarchical general linear model (GLM), following a two-step procedure. At a first level of analysis, the data of each participant (individual trials) were analyzed independently to estimate the parameters of the GLM based on electrophysiological activity at each time point and electrode. The first-level analysis delivered beta coefficients (i.e., the strength of the effect of the dependent variable) for each experimental condition and each individual dataset. The second level analysis used these parameters to test for statistical significance across participants, by performing robust statistics. Note that the GLM approach implemented in LIMO can be seen as a mixed effects model, but this model differs from traditional mixed effects models used with behavioral psycholinguistic data, because LIMO does not allow to use many predictors at a time (see 2.2.4 for a detailed description of the LIMO analysis).

I initially investigated the main effects of picture Name Agreement (continuous predictor) and Session (categorical predictor) on the signal and a potential interaction between the two. I assessed the main effect for the continuous variable using a one sample t-test and a paired t-test for the categorical variable. The interaction between Name Agreement and Session was investigated using a repeated measures ANOVA, with picture Name Agreement and Session as repeated measures for the analysis. The analysis was performed using 1000 bootstraps, to control for multiple comparisons. Corrections for multiple comparisons were applied using spatio-temporal clustering (Maris & Oostenveld, 2007, see also Pernet et al., 2011). In the present analysis, I applied a spatiotemporal clustering with alpha set to 0.05 and neighboring distance at .44 (approx. 5 electrodes per cluster).

[6] I had also intended to assess the ERP correlates of secondary Name Agreement in this study, but a coding error early in the analysis pipeline prevents the inclusion of that analysis at present. In a more exploratory approach, I also additionally assessed the ERP correlates of Lexical Frequency, though, and have included them in the Appendix D.

5.3 Results

Within the 23920 experimental trials recorded, I first excluded 550 trials (2.2%) with naming latencies shorter than 500ms; without repetition within a Session, such short naming latencies typically indicate voice key errors, due to audible hesitations or other task-irrelevant sounds. To specifically focus on items for which each participant appeared to have a strongly preferred name, I also excluded from each participant's dataset any item for which they gave different names across Sessions (1786 trials, 7.4%) or trials with omissions (992 trials, 4.1%). Finally, for the ERP analyses I also excluded an additional 983 (4%) trials due to contaminated EEG signal. Thus, the current ERP analysis was restricted to 19609 (81%) of the originally recorded trials, 9704 in Session 1 and 9907 in Session 2, and the current behavioral analysis was restricted to 20592 (86%) of the originally recorded trials, 10186 in Session 1 and 10406 in the Session 2.

5.3.1 Naming latencies

Participants' responses and mean naming latencies across the two sessions are summarized in Table 1. Response patterns across the two naming Sessions for the dominant and secondary names and corresponding mean naming latencies are summarized in Table 2.

Table 1. Response frequencies and mean naming latencies (in ms) for each Session.

	Responses				**Naming Latencies**			
	Session 1		**Session 2**		**Session 1**		**Session 2**	
	Mean	**N**	**Mean**	**N**	**Mean**	**SD**	**Mean**	**SD**
Dominant	.75	9280	.79	9531	1050	382	1022	354
Secondary	.09	1191	.07	1140	1154	472	1128	459
Other	.07	931	.09	855	-	-	-	-
Omissions	.04	558	.03	434	-	-	-	-
Total		11960		11960	1039	463	1021	424

By-item response frequencies correlated well between Session 1 and 2 within this experiment, for both dominant (r= .91, $p < .001$) and secondary (r= .83, $p < .001$) Name Agreement separately. In

reference to Oppenheim's (in prep.) observed frequencies for the dominant (p = .78, $Mean_{RT}$ = 978, SD = 217) and secondary (p = .10,$Mean_{RT}$ =1125, SD = 399) name, by-item response frequencies corresponded well to these estimates for both the dominant (r= .93, p < .001) and secondary name (r= .83, p < .001) in Session 1 and for the dominant (r= .92, p < .001) and secondary name (r= .83, p < .001) in Session 2. These frequencies also corresponded well between the current experiment and the frequencies observed in Chapter 3 for both the dominant (r= .88, p < .001) and secondary name (r= .79, p < .001) in Session 1 and the dominant (r= .89, p < .001) and secondary name (r= .78, p < .001) in Session 2.

Table 2. Response patterns across the two Sessions for the dominant and secondary names and mean naming latencies (ms).

	Responses		**Naming Latencies**	
	Mean	**N**	**Mean**	**SD**
Dominant Name in both Sessions	.71	17054	1024	351
Secondary Name in both Sessions	.05	1224	1156	445
Switched from Dominant to Secondary Name	.03	780	1102	427
Switched from Secondary to Dominant Name	.03	918	1102	420
Other Responses patterns	.18	3944	1021	424
Total		23920	1030	444

As described in the Methods section, I used maximal linear mixed effects regression, to predict participants' naming latencies as a function of (1) Session (an ordinal measure from 1:2, centered), (2) picture Name Agreement from Oppenheim's (in prep) recent Bangor norming study (a continuous measure from 0:1, centered). Collinearity was not an issue in the current study, because the experimental conditions were not correlated. I also report two restricted models, with Session 1 and Session 2 naming latencies respectively (see Table 3).

Table 3. Summary of LMM analyses of inverse-transformed naming latencies.

Both Sessions

	β	*SE*	*t*	*p*
Intercept	10.012	0.253	39.47	<.001***

Session	-.263	.119	-2.218	<.001***
Name Agreement	-5.113	.297	-17.206	<.001***
Session*Name Agreement	.354	.164	2.154	.04*

Session 1 only

	β	*SE*	*t*	*p*
Intercept	-9.885	.240	-41.12	<.001**
Name Agreement	-5.325	.325	-16.34	<.001***

Session 2 only

	β	*SE*	*t*	*p*
Intercept	-10.139	.278	-36.41	<.001**
Name Agreement	-4.908	.239	-16.67	<.001***

Mixed effects analyses revealed significant effects of Session and picture Name Agreement on naming latency data. Speakers were generally faster to name pictures with higher Name Agreement ($\beta_{NameAgreement}$ = -5.11, *SE* = .29, $p < .001$), while naming latencies in total decreased significantly in Session 2 ($\beta_{Session}$ = -.26, *SE* = .11, $p < .001$). A significant interaction between Session and Name Agreement emerged ($\beta_{Session*NameAgreement}$ = .35, *SE* = .16, $p < .04$), suggesting that the Name Agreement effect in naming latencies was much stronger in Session 1. Restricted models further demonstrated the robust effects of picture Name Agreement independently during the first ($\beta_{NameAgreement}$ = -5.32, *SE* = .32, $p < .001$) and the second naming Session ($\beta_{NameAgreement}$ = -4.90, *SE* = .22, $p < .001$).

5.3.2 ERPs

Because the aim of the current study is to assess the inherent effects of picture Name Agreement without any a priori restriction over time or spatial localization of these effects, the ERP analyses followed an exploratory approach to observe effects across the whole scalp throughout the response period. Stimulus-locked (from picture onset to 600 ms) single trial ERPs were analyzed across all space and time dimensions using a hierarchical general linear model in LIMO (Pernet et al., 2011). After estimating beta coefficients for each participant (1st level analysis), I used these parameters to identify effects that were consistent across participants (2nd level analysis).

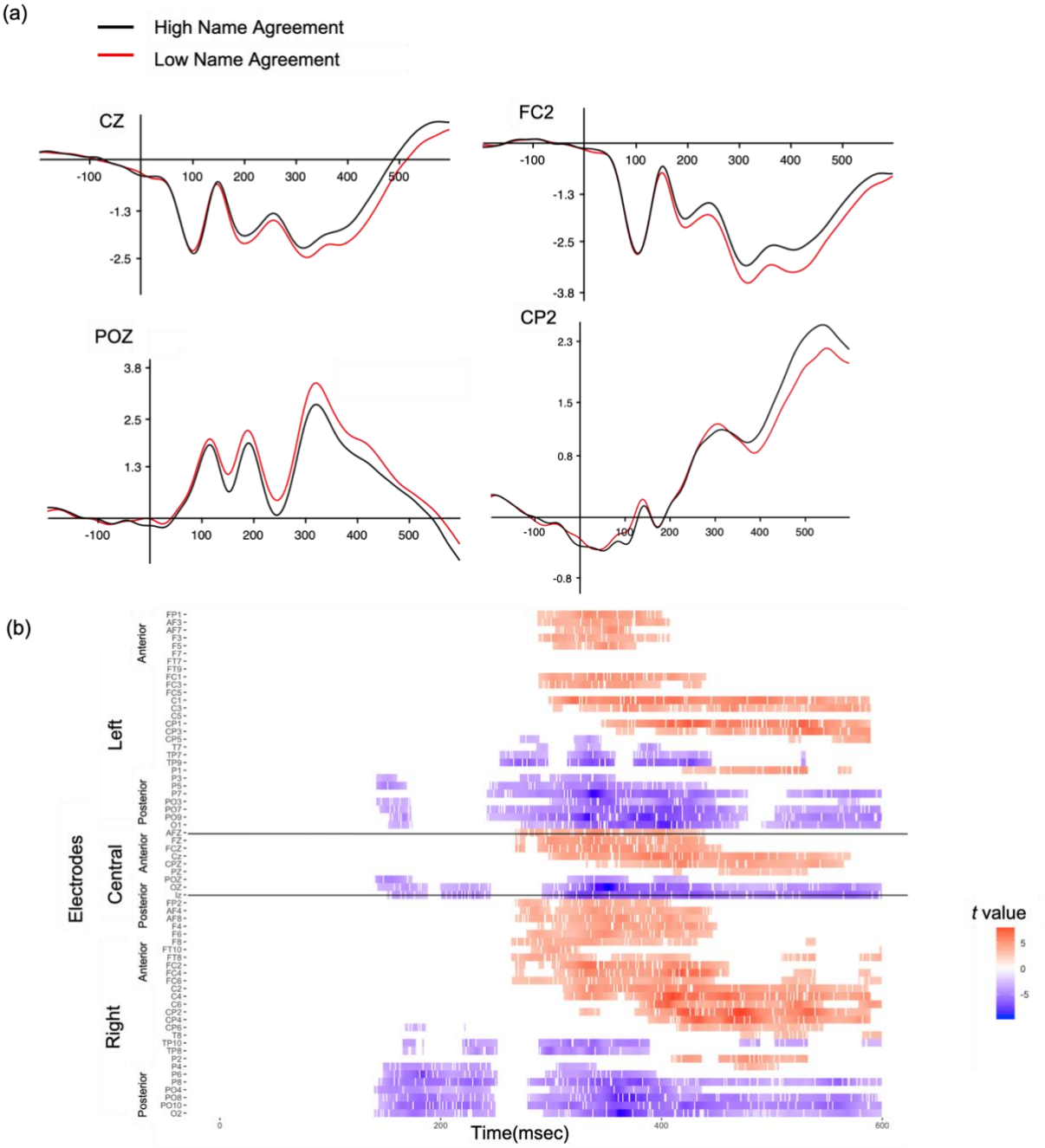

Figure 2.(a) Grand-average ERP waveforms elicited by high and low Name Agreement pictures across the two Sessions. ERPs were computed for a visual illustration, based on a mean split on the dataset, resulting in 269 items being classed as low name agreement ($Mean_{NA}$ <.86) and 256 items being classed as high Name Agreement ($Mean_{NA}$ >.86). (*b*) Results of the robust one sample t-test (representation of significant *t*-values) for the main effect of Name Agreement across the two Sessions after correction for multiple comparisons at all electrodes and timepoints, with the colors representing electrodes with higher (red) and lower (blue) amplitudes for trials with Name Agreement; each line is an electrode.

Session

After correcting for multiple comparisons, no main effect of Session on ERP amplitudes exceeded the significance threshold for any electrode at any timepoint (all $p > .05$).

Name Agreement

As Figure 2b demonstrates, ERP amplitudes differed significantly ($p < .05$) as a function of picture Name Agreement in several time windows. The earliest differences are observed after 150 ms post picture onset at mostly posterior areas of the scalp, in which decreasing Name Agreement resulted in increased electrophysiological negativity. The second robust difference emerged at 280 ms post picture onset and lasted throughout the entire temporal window of analysis up until 600 ms, in which I observed increased positivity in posterior areas for trials with higher Name Agreement and increased negativity in anterior areas for trials with lower Name Agreement. These differences between trials with high and low Name Agreement produced robust effects in separate sub-windows, with decreasing Name Agreement resulting in lower ERP amplitudes in anterior areas of the scalp bilaterally from 290 until 440 ms and increasing Name Agreement producing more positive electrophysiological activity in posterior areas of the scalp from 290 ms until 600 ms post-picture onset.

*Session * Name Agreement Interaction*

After examining the main effects separately, I performed a repeated measures ANOVA in LIMO (using Session and Name Agreement as repeated measures), in order to investigate for a potential interaction. However, once I corrected for multiple comparisons, no significant interaction between Session and Name Agreement exceeded the significance threshold on ERPs ($p > .05$).

Post-hoc t tests on Name Agreement

Driven by the main effects of Name Agreement and by the significant interaction between Session and Name Agreement in naming latencies, I computed post-hoc comparisons between high

and low Name Agreement in each of the two Sessions. One sample t-tests of Name Agreement failed to reach statistical significance in the first Session ($p > .05$) but came as significant in the second naming Session ($p < .05$).

5.4. Discussion

Naming pictures with multiple appropriate names is associated with robust behavioral effects, widely reported in the literature and some ERP modulations which are less consistent. The majority of these name agreement effects have been interpreted as evidence in favor of the lexical competition hypothesis and primarily assume to reflect the cost that arises during selection: naming a picture of a couch takes longer time than naming a picture of a dog and modulates ERP peaks because, during formulation of the linguistic concepts, the selection of "couch" is proportionally delayed by the activation of its alternative names (e.g., "sofa", "settee") (e.g., Levelt et al., 1999; Roelofs, 1992; 2003). Such picture name agreement variations are thought to index the struggle of the production system to systematically select the most appropriate word according to the populations' tendencies amongst different equally suitable alternatives. This default conceptual use of picture name agreement is challenged by the recent finding that speakers develop stable word preferences in naming (Balatsou, Fischer Baum & Oppenheim; in revision), even for non-dominant responses, suggesting that low name agreement does not necessarily imply that individual speakers have *actively competing* alternative words in their production systems, or at least not to the extent previously assumed. This demonstration consistent with *idiolects*, which is disproportionally greater for low name agreement pictures, challenges the default use of a population-level measure as directly indexing *relative* activation thresholds for each speaker, and therefore, the basis of interpretation of name agreement modulations. In the present study, I aimed to assess for the first time, the electrophysiological effects of picture name agreement for individuals' most accessible word forms, and the extent to which they can account for endogenous lexical competition within the production system by evaluating the behavioral and

electrophysiological effects of name agreement in response consistency. In a repeated naming task with minimal endogenous or exogenous conflict, I hypothesized that if the previously reported effects of name agreement (e.g., Cheng et al., 2010; Shao et al., 2014) dissipated, then they were most likely previously indexing the endogenous lexical competition the response competition created by task-demands. If, though, picture name agreement still produced robust behavioral and electrophysiological modulations, then I predicted that these effects are likely associated with increased activation in the system during other necessary sub-stages, such as lexical activation and/or phonological encoding (e.g., Valente et al., 2014).

The findings of the current study demonstrate that name agreement produces robust behavioral and electrophysiological effects, even after controlling for the heterogeneity in speakers' idiolects. A regression analysis of the naming latencies replicated the widely reported effect of picture name agreement on retrieval speed (e.g., Alario et al., 2004; Bates et al., 2003). The longer naming latencies for pictures with lower name agreement in the first naming session, thus, reflect the endogenous conflict of the production system in simple naming, as in norms (Bates et al., 2003; Szekely et al., 2004). Such conflict has been previously interpreted as evidence for endogenous lexical competition driven by the notion of stochastic word selection (e.g., Alario et al., 2004), however, in the case of idiolects, there is no theoretical basis to assume that lexical competition is the driving force of this effect, since Session 2 retrievals suggested that none of the alternative word choices has reached the *relative* selection threshold inside speakers' mental lexicons, per Levelt et al.'s (1999) Luce choice-inspired account of selection probability.

The behavioral name agreement effect was also robust during the second naming session: speakers were faster to name a picture of a dog compared to a picture of a couch, despite the reduced conflict of response consistency after reproducing "dog" and "couch". This finding replicates the effect observed in the first session, while it is surprisingly compatible with the previously reported name agreement modulations on naming latencies in studies that have used repeated naming in the process

of familiarization (e.g., Alario et al., 2004; Shao et al., 2014), despite they could have possibly misidentified some of speakers' word preferences. The significant interaction between Name Agreement and experimental Session on naming latencies suggested that, even though speakers were slower to name pictures with multiple appropriate names in both sessions, this effect was slightly weaker in Session 2. This replicates the widely reported repetition priming in picture naming (e.g., Cave, 1997), which can eventually be the basis for formulation of stable word preferences. For instance, in the error-based learning model of cumulative semantic interference, each word retrieval results in the strengthening of the connection between the semantic-to-lexical mappings, making the target words more accessible for future selections (Oppenheim et al., 2010; 2007). If we assume that such competitive learning can also emerge as long-lasting, then it is possible that speakers' vocabularies are formed through the strengthening of the associations between the intent and the linguistic output. However, because these effects are reported in a two-session naming task, I can only remain speculative as to how speakers' response profiles would change with additional naming and whether this tendency to repeat ones' previous selections persists or not. In any case, the name agreement effects reported in the current study indicate that, at least some kind of conflict is also endogenous in speakers' stable word preferences, even with reduced lexical co-activation.

Concerning the electrophysiological analysis, the results indicate that name agreement variations modulated ERP amplitudes in several time windows related to both early and later processes of word production. These results are compatible with both the suggestion that name agreement affects the substage of semantic-to-lexical mappings (Alario et al., 2004) and suggestion that effects arise during phonological encoding (Valente et al., 2014). However, in the current study I also observed modulations in earlier time windows, before lexical processing. Even though an exploratory mass-univariate analysis cannot directly be related to traditional "component" analyses, some of the earliest differences observed (150ms to 240ms post picture onset) in parietal areas of the scalp are compatible with the differences in the P1 component that Cheng et al. (2010) previously reported. The differences

in this early time window most likely indicate differences in the earliest perceptual levels between pictures that elicit multiple alternative words instead of pictures corresponding to one lexical candidate: for instance, “dog” may require less visual or perceptual processing than “can opener”, which has increased lexical and visual ambiguity and, in an interactivity-based interpretation, could pass on increased activation to the earliest perceptual levels of processing. In the next time window (280 ms-600 ms post picture onset), the observed effects can be associated with both the N200 component previously reported (see Chapter 5; Shao et al., 2014) and other language-sensitive peaks, like the N400 (see Chapter 5; cf. Valente et al., 2014), which has been previously linked to increased semantic-to-lexical processing during production (e.g., Costa et al., 2009). Finally, the later effects are partly compatible with Valente et al.’s (2014) observations, who were also not temporally or spatially restricted and reported name agreement differences arising at a continuous time window of around 380ms post picture onset until before articulation. This time window of name agreement differences is partly compatible with Indefrey and Levelt’s (2004) timeline, which suggests that this is when phonological encoding takes place, however the continuity, overlap and duration of the ERP effect suggests that activation cascades throughout the response period. Increased activation for lower name agreement pictures seems to cascade from the earliest stages of semantic-to-lexical processing, continuing during phonological encoding, a finding that serves as an additional empirical observation to question the seriality and discreteness of the timeline (also see Strijkers et al., 2010; Munding et al., 2016, Nozari & Pinet 2020).

Even though the interaction between Name Agreement and experimental Session did not emerge as significant in ERPs, post-hoc comparisons revealed that the electrophysiological differences in name agreement variations were not independently significant in the first naming Session, but emerged as significant in Session 2. Unlike the behavioral effects of name agreement which were significant in both naming Sessions, the robust ERP effect emerging during the second session could be explained as improved efficiency in word learning, indexed by enhanced neural synchronization:

the same kind of increased brain-to-stimulus synchronization underlying associative learning tasks and speech processing (e.g., Gotts, Chow & Martin, 2012; Assaneo et al. 2019), which increases the differences in the semantic-to-lexical mappings between the condition that is harder (i.e., low name agreement pictures) and the condition that is easier to learn (i.e., high name agreement pictures). Nonetheless, the electrophysiological differences elicited by name agreement variations indicates that repeated naming robustly modulates ERPs, but the absence of a significant Name Agreement x Session interaction cannot explicitly indicate whether these differences observed after subsequent naming are inherent to the production system (e.g., Nozari & Hepner, 2019; Oppenheim & Balatsou, 2019), or reflect the dynamics of a system that is continuously learning by tuning its semantic-to-lexical mappings (Oppenheim et al., 2010).

By investigating reduced competition scenarios (i.e., speakers' idiolects) in repeated naming, I examined whether the previously reported modulations of Name Agreement, which were interpreted as indexes of lexical competition, still emerged. Using a more theoretically constraining task than previous ERP name agreement studies (e.g., Cheng et al., 2010; Shao et al, 2014), and a robust exploratory analytical method, I identified differences in electrodes and time windows which would have otherwise been attributed to the lexical selection by competition hypothesis. The benefit that comes with controlling for speakers' lexical consistency, though, reduces the basic interpretive framework for attributing these effects to lexical competition: robust modulations were reported contra to the stochastic selection that Levelt et al. (1999) proposed and with the absence of a strong "competing" word in speakers' mental lexicons. These robust ERP effects of name agreement, that remained after eliminating *relative* activation thresholds, demonstrated the differences in processing for pictures that activate multiple words, which cascades within the production system from the earliest towards the later stages of production and not differences in selection of words, as previously assumed.

5.5. Conclusion

In this simple naming task, I have, thus, identified several behavioral and electrophysiological effects of high and low name agreement pictures, even after taking steps to minimize within-speaker lexical "competition". Previous research has predominantly interpreted similar robust effects as directly indexing lexical-level competition, that is, the cost associated with the need to suppress alternative lexical representations which race for selection (e.g. Shao et al., 2014). This interpretation is mostly based on the default assumption that "couch" and "sofa" are actively competing for selection at any given time (e.g., Levelt et al., 1999; Roelofs, 1992). Here, I showed that these "couch"-only and "sofa"-only speakers, which I identified in Chapter 3, still appear recruit more cognitive recourses in naming pictures with lower name agreement, while I surprisingly replicated some of the robust effects that previous literature reported with only "couch-sofa" speakers in mind. By challenging this core assumption of selection probability, I further showed that several robust effects associated with the encoding processes during production emerge in more naturalistic designs, like simple picture naming. I, therefore, suggest that such effects are much more indicative of the nature of the encoding processes and the challenges associated with selecting a word with multiple co-activated lexical alternatives. More generally, a major theme in the current research is that identifying and respecting individual differences in relation to population-derived predictors in cognitive psychology provides a more valid basis to associate the empirical findings with the hypotheses that researchers aim to examine.

CHAPTER 6-General Discussion

6.1 Overview of the thesis

So, how do speakers select one appropriate word for production? It is reasonable to assume that the process of finding the appropriate word for a concept is straightforward when there is only a single candidate linguistic unit in one's mental lexicon. However, that is not how language production normally works: Speakers' vocabularies consist of tens of thousands of words in their native language, with multiple units usually corresponding to the same, or approximately the same, meaning (Levelt, 1989). The question of how speakers select the appropriate word out of multiple similarly suitable alternatives has dominated the field of psycholinguistics for decades, while the nature of the encoding processes involved in speech production are hotly debated to this day. There is controversy as to whether the mechanism guiding lexical selection is chronometrically dependent (e.g., Levelt et al., 1999) or independent (e.g., Mahon et al., 2007) upon the availability of responses which are determined during the previous step, namely lexical co-activation. By measuring the effect of the availability of alternative candidate words, which in simple naming tasks is indexed by picture name agreement, empirical observations are usually interpreted in favor of lexical competition. But, until now, the field has lacked a critical evaluation of picture name agreement as a psychologically meaningful predictor of individuals' internal states, necessary for its subsequent evaluation with respect to the nature of lexical selection. By observing how variations in picture name agreement affect individuals' behavioral and electrophysiological profiles in simple naming tasks, this thesis aimed to elucidate the nature of the lexical selection mechanism and, at the same time, provide a new empirical basis for future research to use picture name agreement as a measure of lexical co-activation.

Chapter 1 introduced the relevant literature and critically evaluated the empirical findings in favor of the lexical selection by competition hypothesis and the alternative account that selection is a non-competitive process. I identified conceptual misuses and overinterpretations of the relevant effects in favor of competition and targeted new research questions to demonstrate the genuine effects of name agreement and their relation to competition. In Chapter 2, I detailed the methodological and technical

approach used in the current work and justified its relevance for answering my research questions. Chapter 3 investigated within-speaker stability of name choices across multiple sessions, showing that picture name agreement is a valid measure of lexical co-activation within individual speakers. Importantly, it further demonstrated that it also indexes heterogeneity between individuals' stable word preferences, that is, their idiolects. In Chapter 4, I reported how we tested the hypothesis that directing production against speakers' idiolects would largely increase the existing competition in the system. Instead, I found that the inherent conflict reflected in name agreement effects is naturally distinct from the exogenous competition in task-oriented word production. In Chapter 5, I showed that consistent responses also produce robust electrophysiological and behavioral effects and demonstrated that in this minimal conflict scenario, picture name agreement is a valid measure of lexical co-activation for dominant and secondary responses, but not of response competition. The key contributions of the current work are that (1) it provides the first empirical evaluation of picture name agreement showing that it is a reliable measure of co-activation in word production, but not necessarily of competition, (2) it assesses for the first time the heterogeneity in individuals speakers' lexical preferences, which we termed as idiolects, and (3) it demonstrates that the selection mechanism in simple word production is qualitatively different from the mechanism involved in production with response competition. By examining this measure in simple designs, I aimed to mirror how language production normally works and wished to provide reproducible effects, which can be extended in new and useful directions in the field.

6.2 Empirical evaluation of picture name agreement

Picture name agreement is often the measure of interest in norming studies (Bates et al., 2003; Szekely et al., 2004) and a widely used predictor in cognitive science research. Its robust effects are most often associated with the process of lexical selection according to the competition hypothesis (Alario et al., 2004; Kan & Thompson-Schill, 2004; Valente et al., 2014; Bose & Schafer, 2017).

Perhaps because this default explanation of name agreement effects seemed so cognitively plausible, it has generally escaped critical empirical assessment. A main goal of the current work was to provide a body of evidence to assess and evaluate the theoretical basis of these interpretations for the first time.

6.2.1 Luce Choice-inspired selection is partially verified

Levelt et al.'s (1999) production theory was undeniably influential in the field of language production. The core computational principles of their model have guided most interpretations on the robust name agreement effects researchers have observed in production tasks. One of the most critical components of their theory, the Luce Choice-inspired (Luce, 1959) ratio is the computationally explicit formulation of the lexical competition hypothesis, which determines both the probability that the target word will be selected and the time it will take to be selected according to the *relative* activation of other words in the system. The notion that selection probability and latency is guided by the stochastic function of the Luce choice rule has been the basis of explanations of name agreement effects, but to my knowledge, has never been explicitly tested before.

In Chapter 3, I showed that the Luce Choice ratio guides the probability of selection in individual speakers. By examining name-selection consistency, I observed that population-level norms, which index the distributions of possible responses across speakers, predict within-speaker variability in word selections for both stronger and less modal responses (i.e., secondary names). The demonstration that the Luce Choice-inspired ratio guides individuals' probability of re-selecting either "couch" or "sofa" provides strong empirical justification for the use of picture name agreement variations as cognitively meaningful predictors in word production studies. The demonstration that within-speakers' selection probability is guided by lexical co-activation across speakers makes it, therefore, possible to partially justify a *relative* activation threshold in selection. However, this is a between-trial empirical observation of response consistency and not a within-trial estimate of the co-activation of those options, which researchers generally assume affects the production of the target

word in both simple naming tasks (e.g., Alario et al., 2004) and in picture-word interference studies (e.g., Bloem & La Heij, 2003; Roelofs, 2003; Vigliocco et al., 2004). Although the basis for interpreting the competition effects of picture name agreement is the increased naming latencies for items with multiple candidate words, in the current research I have not been able to evaluate naming latency effects for voluntary stochastic selections, since speakers are less likely to switch between names for subsequent productions. The major contribution of these findings is, therefore, in relation to response consistency and not latency which, to date, remains the strongest principle of competition (Levelt et al., 1999).

6.2.2 Picture name agreement also reflects variation in idiolects

The demonstration that variability in name agreement also reflects the heterogeneity between speakers' independent word preferences, which I termed as *idiolects*, at the same time questions some of the normative uses of this measure (Alario et al., 2004; Shao et al., 2014; Valente et al., 2014) and the strictly competition-based interpretations of its effects (Alario et al., 2004; Cheng et al., 2010; Shao et al., 2014; Bose & Schafer, 2017). In Chapter 3, we showed that speakers develop stable preferences in repeated naming for both dominant and secondary responses and, in Chapter 5, I reported some of the widely observed effects of name agreement in naming latencies and ERPs emerging for consistent responses. In the strictly non-competitive framework, according to which a word is chosen after it passes an *absolute* threshold (e.g., Dell, 1986), idiolects should reflect the minimal conflict scenario within the system, reflecting the strengthened one-to-one mappings between concepts and intended words. Similarly, within the competitive framework, and given that picture name agreement is tied to selection difficulty, idiolects should reflect a significantly reduced competition scenario in comparison to stochastic selections. More specifically, the relative threshold of activation of the competitor words should be minimal, if not non-existent, since they do not reach the point of selection as predicted by a stochastic axiom. Despite these predictions, I found that the main effects in naming latencies and ERPs

persist even for individuals' consistent responses: speakers are significantly slower in re-naming pictures with lower name agreement (e.g., "couch" or "sofa") than pictures with higher name agreement (e.g., "dog"), while the nature of the electrophysiological effects indicates increased co-activation in multiple processing stages for pictures with numerous names.

The observation of such strong name agreement effects for repeated name selections is surprisingly compatible with the previously reported effects in the name agreement literature in studies in which individuals' word preferences were not explicitly taken into consideration in the experimental design (Alario et al., 2004; Cheng et al., 2010; Shao et al., 2014; Valente et al., 2014). This is significantly more paradoxical, when considering the effects that involuntary name switches can produce, as shown in Chapter 4: by respecting and contradicting speakers' idiolects, we demonstrated the difference between the inherent conflict of the system during typical lexical selection (name agreement in line with speakers' idiolects) and the conflict resulting from task-induced competition (name agreement contra to speakers' idiolects). The most probable reason why name agreement effects are still evident in the most minimal conflict scenario, is that it they indeed measure within-subjects lexical co-activation: even though "sofa" will not be reaching a selection threshold within the "couch" speaker, it still appears to be an active option, which makes production more effortful. It is, yet, still unknown whether the majority of the effects previously reported are in line with individuals' word preferences (e.g., Alario et al., 2004; Valente et al., 2014; Shao et al., 2014), so the nature of those effects in reference to name consistency or voluntary switching remains a point of speculation.

Nonetheless, this heterogeneity in individuals' responses provides the first empirical observation that there are individual differences in name selections, a finding useful to evaluate in relation to the nature of the encoding processes in other naming tasks, as well. By accepting words that deviate from the dominant group-level norms as appropriate responses, it is possible to re-examine the previously reported effects of within-trial co-activation (e.g., Vigliocco et al., 2004) or within-block co-activation (e.g., Howard et al., 2006) in respect to speakers' own absolute activation thresholds. If,

for instance, semantically similar entries in respect to speakers' idiolects significantly delay production speed in interference tasks, then it is reasonable to associate these effects with conflict during lexical selection. An additional benefit of keeping track of such preferences, is that researchers can include response consistency as a predictor both in studies which directly or indirectly manipulate the levels of lexical co-activation.

6.2.3 Name agreement is sensitive to name change

Findings from the current study replicated the widely documented effect of name agreement on naming latencies (Snodgrass & Yuditsky, 1996; Ellis & Morrison, 1998; Bates et al., 2003; Szekely et al., 2004; Alario et al., 2004; Shao et al., 2014; Bose & Schafer, 2017) and partially extended this effect in situations of both low and high-conflict production demands. In Phase 1 of Chapter 4, I reported significant differences in naming latencies for low (e.g., "couch") and high (e.g., "dog") name agreement pictures. In Phase 2 of Chapter 4, I further showed that name agreement effects still emerge for the dominant responses, when they are following speakers' preferences in repeated naming: retrieving "couch" twice is slower than retrieving "dog" twice despite previous findings that repetition priming is stronger for lower name agreement pictures (e.g. Park & Gabrieli, 1995). The naming latency effect in response consistency was extended in Chapter 5, by demonstrating that naming low name agreement pictures significantly delays production both for dominant and secondary responses together: selecting either "couch" or "sofa" twice is still slower than producing "dog" twice. Production delay for secondary naming preferences broadens the interpretation of picture name agreement as both a measure of retrieval difficulty and as an index of lexical co-activation, because it demonstrates that the alternative word, irrespective of whether it is the modal response, has not reached the criterion for selection within-speakers, but still modulates semantic-to-lexical processing.

However, this modulation is not as systematic as the competition hypothesis would predict (Levelt et al., 1999; Roelofs 1992; 2003; 2018), because, when within-speaker lexical co-activation

was artificially exceeding the criterion for selection, we observed that competition did not increase in the same fashion. In Phase 2 of Chapter 4, the differences between low and high name agreement effects largely dissipated when the task coerced alternative names for both conditions (i.e., switching responses between "couch and "sofa" became surprisingly similar to switching between "dog" and the less modal "mutt"). By inducing exogenous activation in the production system and creating an artificial environment to test the axiom of selection, we have shown that name agreement effects are sensitive to name change.

However, we were also able to provide some indirect measures of "competition" in the narrow sense of Levelt et al. (1999) and in the adjustable framework of Nozari and Hepner (2019). It is worth highlighting that the latency effects in non-voluntary name switches are partially compatible with the core principle of a stochastic selection, in which the activation of the competitor words delays target word retrieval (Levelt et al., 1999; Belke et al., 2005; Bloem & La Heij, 2003; Bloem et al., 2004). Instead, though, of the traditional competitive explanation of this effect, I suggest that delayed production in this high conflict scenario most closely resembles the exogenously-driven competition observed in picture word interference studies (e.g., Schriefers et al., 1990), and appears to be an actual "infrequent derailment" of producing speech. One possible explanation for this is characterization is that name switches in the current research, and elsewhere in the literature, resulted from explicit task manipulations and not from speakers' own voluntary need to use alternative labels. Nevertheless, the finding that picture name agreement produced robust effects even in such high conflict production situations provides new evidence that it is related to the automatic, top-down process of lexical co-activation, which increases the likelihood that its effects are replicable.

6.2.4 New electrophysiological modulations of picture name agreement

Because the ERP name agreement literature is limited and the effects reported are inconsistent, a main goal of the current work was to be able to evaluate electrophysiological effects in relation to

the encoding processes. Previous literature has identified name agreement modulations in the P100 range (Cheng et al., 2010), the N200 range (Cheng et al., 2010; Shao et al., 2014), and in a later processing window starting at ~380 ms post stimulus onset and 100 ms before articulation (Valente et al., 2014). These effects have been associated with lexical selection and phonological encoding, respectively, mostly based in Indefrey and Levelt's (2004) widely questioned timeline. In the current experimental work, we found name agreement differences in the N200 (Chapter 4), observed a novel N400 modulation (Chapter 4), and identified additional differences arising 290–600 ms after stimulus onset (Chapter 5). These observations served as additional evidence that picture name variability indeed indexes lexical co-activation, while the temporal and spatial distribution of these effects, are compatible with cascading interactivity in the production system.

N200

The N200 effect was explicitly observed in the current work when we examined picture name agreement as a categorical variable in simple naming which omitted the previously used practice of familiarization (Chapter 4). Previous observations of name agreement effects in the N200 time window were interpreted as indexing response inhibition following familiarization (Shao et al., 2014) and "lemma"-level competitive selection in overt naming without familiarization (Cheng et al., 2010). Naming a picture with multiple words was assumed to require additional resources at the level of lexical selection or required the suppression of the active competitors in order to achieve target word retrieval. Both interpretations, though, were based on the competition hypothesis and its respective timeline. These two effects were also spatiotemporally district, which makes it probable that they also related to different cognitive processes.

These competition interpretations would predict the opposite behavior of an inhibitory component, compared to the effects reported here. That is, the need to recruit additional inhibition or increase the existing competition within the system, when the task imposes a response as correct-only

and thus the levels of co-activation are artificially increased. Contra to this hypothesis, the N200 was found to decrease with increasing "competition", which made us attribute it to differences in the earliest points of lexical access, driven by similar findings in the literature which associated increased N200 magnitude with increased ordinal position in cumulative semantic interference tasks (as in Costa et al., 2009; see also Strijkers & Costa, 2011).

In simple naming, effort in lexical access should disproportionally increase by increased semantic-to-lexical activation, reflected in pictures with multiple words. This effort appears to later be resolved with repetition priming resulting from subsequent naming, even in high conflict production demand, such as violating one's preferences. This interpretation was broadened with the electrophysiological analysis of consistent responses in Chapter 5. Even though the analytical approach of ERPs followed in Chapter 5 cannot generally reveal distinct component modulations, I found increased bilateral negativity associated with lower name agreement in several time windows with distinct topographies, which can relate to both the N200 effect observed in Chapter 4, the anterior N200 that Shao et al. (2014) reported and the posterior negativity that Cheng et al. (2010) interpreted as an N200. The N200 in the current studies is, thus, associated with picture name agreement variations in the earliest processing levels, but does not appear to reflect response competition or selective inhibition.

N400

An additional contribution of the current work is that it identified novel N400 modulations for lower name agreement pictures in simple and coerced naming (Chapter 4) and electrophysiological effects compatible with an N400, when speakers produced their consistent responses (Chapter 5). I did not explicitly predict such effect, largely because it has not been previously reported in the name agreement literature. Another reason is that naming pictures with multiple alternative names does trivially relate to semantic manipulation, which is usually driving the N400 (Kutas & Federmeier,

2011). N400 modulations have been found in other language production tasks, such as primed picture naming (Ganushchak & Schiller, 2008; Koester & Schiller, 2008), picture word interference studies (Blackford et al., 2012; Piai et al., 2012; Shitova et al., 2017), and lexical decision tasks (Rabovsky, Sommer & Abdel Rahman, 2012). While some previous interpretations have linked N400 effects to lexical competition (e.g., Piai et al., 2012; Rabovsky et al., 2012; Shitova et al., 2017), in speech production it is generally associated with semantic-to-lexical mappings during the formulation of words. In the current research, I have interpreted the N400 effect as evidence for cascading interactivity within the production system (Dell, 1986) and, at the same time, an empirical demonstration that lower picture name agreement is indeed more resource-intensive for the production system. This is without necessarily indexing resources that are revealing active race between words.

If we follow Indefrey and Levelt's (2004) temporal stages of word production, the N400 effect falls into the syllabification window, which they place between 330 ms and 455 ms post-stimulus-onset. In a strictly serial and discrete model of language production (Levelt et al, 1999), it is difficult to explain why the N400-an established meaning-related index-would reflect such lower-level processing, or why semantic effects would follow lexical selection (175 ms and 250 ms post stimulus onset according to the timeline). Compatible with previous literature in both language production and comprehension (see Kutas & Federmeier, 2011, for an extensive discussion), I suggested that the N400 in the current research is indeed indexing semantic-to-lexical processing difficulty for pictures with multiple words. At the same time, this interpretation of the N400 is compatible with accounts that question the reliability of the proposed timeline (see Strijkers & Costa, 2011; Nozari & Hepner, 2019, Munding et al., 2016 for similar discussions).

The fact that the observed N400 effect persisted in both high and low production demands and was not reduced by repetition priming suggests that it is sensitive to this inherent conflict. The enhanced N400 for lower name agreement pictures could be explained by their increased semantic co-activation and increased semantic density compared to higher name agreement pictures (Rabovsky et

al., 2012). For example, “couch”, “sofa” and “settee” taken together should elicit more semantic processing than “dog”, because “couch”, “sofa” and “settee” have highly intercorrelated, yet partially distinct, semantic features. This interpretation, though, makes us speculate why such N400 modulations have not been previously reported in the ERP name agreement literature (Cheng et al., 2010; Shao et al., 2014; Valente et al., 2014). Although Valente et al.’s (2014) exploratory analysis did not report a distinct N400 either, differences associated with picture name agreement were reported in the N400 window, which resemble the findings of Chapter 5. In previous studies that used a predictive component-based approach, it is possible that the N400 was not a component of interest or not particularly evoked by stimulus-specific properties.. Another point of consideration is that pre-experimental familiarization in the current research was largely omitted, and the experimental manipulations were reduced, possibly allowing for previously masked differences to be enhanced. Nevertheless, I showed that even in low conflict situations, such as repeating one’s previously used names, robust effects in the N400 are reported for low name agreement pictures. The N400 in the current research, therefore, serves as additional evidence that name agreement is a measure of co-activation in the production system, yet not as concrete evidence to associate this increased co-activation with increased lexical competition.

Later effects

The current research also identified ERP name agreement effects in a later window coinciding with response initiation. In Chapter 5, I observed a continuous name agreement modulation for consistent responses towards later stages of processing, compatible with some previously reported effects (Cheng et al., 2010; Valente et al., 2014). Such effects are thought to be associated with post-lexical phonological processing differences for pictures with multiple alternative words (Valente et al., 2014) and are compatible with theories which argue that conflict arises in post-lexical, response buffer stages of processing (Mahon et al., 2007). In the context of the Response Exclusion Hypothesis, the

robustness of name agreement effects towards the endpoint of production is compatible with the idea that some lexical information, that must be cleared before the articulation of the target, word remains in the output buffer, and this why pictures with multiple co-activated words demand more resources in that time-window (Mahon et al., 2007; 2012). This interpretation, of course, is opposed to traditional competitive explanations of name agreement effects, which assume that conflict arises during lexical selection (e.g., Shao et al., 2014; Bose & Schafer, 2017). In addition, even if I consider scaling the timeline of Indefrey and Levelt (2004) to longer naming latencies, these later effects still continue into the window of phonological encoding, which is past formulation, in a strictly discrete model. In any case, the differences in such later processing stages demonstrates that the increased challenge in producing a word with multiple alternatives persists, even when high conflict is reduced in the system.

6.3 A stance on lexical competition

Apart from providing the first empirical evaluation of picture name agreement in the literature, the major theoretical motivation of the current work was to provide additional insight into the nature of the lexical selection mechanism by observing how variations in name agreement affects naming latencies and ERP amplitudes in simple naming. The behavioral and electrophysiological name agreement effects across the three experimental chapters indicate that the production system indeed puts more effort into processing words with strong alternatives. However, I argue that this effort does not provide strong support for the systematic competition that Levelt et al. (1999) have described.

6.3.1 Evidence in favor of lexical co-activation but against systematic competition

In the current work, I have demonstrated that name agreement variations are associated with production difficulty. For example, naming a picture of a couch, that can also be called a “sofa” or a “settee” (low name agreement), is more neurally effortful and more time consuming than naming a picture of a dog (high name agreement), regardless of speaker-specific preferences. We have also

empirically supported the assumption of a stochastic response selection. In other words, “sofa” and “settee” are both active within individuals (Chapter 3) and there is sufficient behavioral and electrophysiological evidence to suggest that they are also active within “couch” individuals (Chapters 3 & 5). Following the general trend in the literature, some would immediately argue that these observations are sufficient evidence in favor of the lexical selection-by-competition hypothesis (Levelt et al., 1999; Roelofs, 1992; 2003), or merely proof that selection is minimally competitive (Nozari & Hepner, 2019).

But is this kind of conflict reflecting the systematic racing of words for selection? When we introduced an actual competitor word into this race, not only the inherent lexical conflict in the system largely dissipated, but the entire selection process turned to task-dependent production, as seen in naming latencies and ERP modulations of picture name agreement (Chapter 4). This distinction between simple and targeted naming could, therefore, mirror the distinction between typical and competitive lexical selection, which seems to concern two distinct processes. The former is task-independent and most closely related to the actual mechanism that governs communicative speech, while the latter can emerge under extreme laboratory manipulations, such as picture naming with a distractor word or directing ones’ speech (cf. Oppenheim & Balatsou, 2019). However, this kind of competition is qualitatively different from the within-trial systematic lexical competition that Levelt et al.’s (1999) theory proposed.

Of course, these findings and their interpretations do not aim to dismiss the possibility and plausibility of a competitive mechanism, nor do they disregard the contribution of Levelt et al.’s (1999) theory in understanding how word production works. It is one thing to demonstrate that conflict exists during lexical processing and another to show that it reflects the active race between words to reach the selection threshold. Findings from the current work have undoubtedly provided sufficient evidence for the former, while future focused research can inform the latter. One useful way to revisit the competition hypothesis, as I partially aimed to do, is to extensively investigate empirical findings and

build theories which integrate both the effects of lexical co-activation and effects of subsequent lexical selection, as Nozari & Hepner (2019) recently did. In the current research, I have provided substantial evidence to suggest that lower name agreement reflects increased lexical co-activation, which produces robust effects during production. However, these effects are qualitatively and quantitatively different from those of response competition. I, therefore, suggest that in order to contribute to the conclusion of the debate, future research should be carefully looking for competition where it actually exists or in the form it exists in the production system.

6.3.2 Idiolects in speakers' vocabularies and competitive lexical learning

A final point of discussion driven by the present findings is whether an element of competitive learning might exist within the production system. We have introduced speakers' idiolects (Chapter 3), findings which suggest that these word preferences are also subject to more effortful retrieval in simple production (Chapter 5), and demonstrated that keeping track of individuals' tendencies is a necessary procedure in understanding how typical lexical selection works (Chapter 4). The observation and evaluation of idiolects is in line with the notion that speakers become better at retrieving words they have recently produced, which is compatible with theories of error-based learning in word production (Oppenheim et al., 2010). In empirical research, incremental vocabulary re-optimization has been used to explain findings in other word production tasks, as the effects of cumulative semantic interference in typical and atypical adults (Oppenheim et al., 2007; 2010, Oppenheim, Tainturier & Barr, 2016; Irons, Oppenheim & Fischer-Baum, 2017) and in novel word production (Oppenheim, 2018). This theory can be easily extended to account for typical word selection, as well. If we imagine that each time an individual retrieves "couch" from their mental lexicon, the semantic-to-lexical mappings for the word "couch" are strengthened and the mappings for the unselected activated words (i.e., "sofa" and "settee") are weakened, then this continuous relearning and unlearning of words can be the basis for the formulation of idiosyncrasies. Of course, this type of lexical learning also takes

into consideration the dispersity of available names given by other speakers in the same linguistic community, otherwise I would only observe “couch” speakers, “sofa” speakers, and “settee” speakers who would not be able to communicate their intents with one another. Perhaps this is resolved by the stochastic axiom of lexical selection, which keeps the “sofa” and “settee” options active within “couch” individuals and, thus, the variability of the linguistic population within the speaker. Although investigating lexical learning in the production system was beyond the scope of the current work, the observation of these idiosyncrasies is compatible with the dynamics of communicative everyday speech, because in the majority of cases speakers continuously reuse, and thus re-optimize, words within their vocabularies.

6.4 Limitations and future directions

Although the degree of consistency of the behavioral and electrophysiological data across the three empirical chapters, as well as their compatibility with the effects in the literature is encouraging, I would like to acknowledge a number of limitations and propose new research directions.

A general limitation of the current research is that it was largely based on observations of between-item comparisons. Although between-item comparisons were the most optimal way to assess the effects of stimulus name agreement, it is also possible that increased variability in uncontrolled variables, such as lexical frequency, image agreement, and concept familiarity, could potentially obscure some observations. We aim to assess those effects and their interactions with picture name agreement in future research. In an attempt to evaluate the effects of picture name agreement in simple designs by speakers, I also used repeated naming in native speakers of British English, which by definition cannot directly generalize to the wider population. It is, thus, useful for future studies to replicate these effects in other populations of the same language and, ultimately, cross-linguistically. Some more specified areas for future research driven by the limitations and the scope of the current work are discussed below.

6.4.1 Secondary name agreement as a window into lexical co-activation

Driven by the finding that speakers can develop stable word preferences even for non-dominant names (Chapter 3), I suggest that secondary responses in the population should be further investigated in both norming studies and in word production tasks that include name agreement as a measure of selection difficulty. For instance, it has been shown behaviorally that both dominant and secondary name agreement affect production latencies in the same fashion. Despite the widely replicated dominant RT effect (i.e., "dog" is faster to name than "couch"), it was additionally found that low name agreement pictures with strong alternative names are, at the same, time faster to name than low name agreement pictures with weak alternative names. For example, "couch" is faster to retrieve than "stroller", even though "sofa" is a stronger response than "pram" in the population (Oppenheim, 2017). This finding poses a major threat to competition-based interpretations of name agreement, as it contradicts the assumption that increased co-activation should increase competition (Levelt et al., 1999). I suggest that future research should investigate individual and population-level secondary name agreement, both behaviorally and with the use of more time-sensitive techniques, like ERPs, to understand where and how it affects target word retrieval. If stronger name agreement produces qualitatively similar effects to dominant name agreement, that is, the previously reported N200 (as reported in Chapter 2; Cheng et al., 2010; Shao et al., 2014) or N400 modulations (as reported in Chapter 4), then it is reasonable to assume that the target and competitor word have a common source of lexical conflict. By, therefore, measuring secondary naming preferences and separately evaluating them in relation to the production of the target word, future research can gain invaluable insight into the nature of co-activation and selection.

6.4.2 How is stochastic selection different from consistency?

In the current work, I have been able to partially support the response pattern predicted by a Luce choice-inspired lexical selection by observing participants' name switches in repeated naming (e.g., speakers initially selecting "couch" and then choosing "sofa" or vice versa; Chapter 3). The overall naming latency and ERP analyses across the empirical chapters were, however, performed in consistent responses (Chapter 5 & Chapter 4, Phase 1) or non-voluntary switching (Chapter 4; Phase 3). This is because voluntary name switches had a relatively low distribution within the data (~15% of the experimental trials in both Chapters 3 & 5), reducing the power of analysis critically, especially in the case of ERPs, which are more sensitive to power loss (e.g., Boudewyn et al., 2018). I have, thus, not been able to report name agreement behavioral and ERP effects in voluntary "couch-sofa" name switches in repeated naming and evaluate how these differ from "couch-couch" or "sofa-sofa" responses.

Because the chronometric and electrophysiological effects of stochastic responses are the centerpiece of the competition hypothesis, I suggest that future research should evaluate these effects in great detail, The naming latency and N200/N400 name agreement modulations reported here were observed for those trials in which the same response has reached the selection threshold twice, while, surprisingly their effects are still compatible with previous observations that did not control this name consistency (e.g., Alario et al., 2004; Shao et al., 2014). One area for future research would be to investigate whether "couch-sofa" would modulate ERPs and behavior in the same fashion that "couch-couch" or "sofa-sofa" did and evaluate whether the cumulative semantic-to-lexical activation of words in former scenario, increases the effort in production (Levelt et al., 1999; Roelofs 2002; 2003). If, on the other hand, name consistency and stochastic selection ultimately produce similar effects, then this could be in line with system flexibility that Nozari and Hepner (2019) proposed.

6.4.3 Less is more

A final scope for future research is to explore the temporal and spatial markers of word production processes in simple language production tasks, which I consider as a necessary approach to increase the replicability of effects and consequently the number of published studies in the field. This can ultimately elevate language production research to the same standards that we observe in other research fields, such as language comprehension. In the present study, I have aimed to understand how single word selection works by observing effects in simple tasks with minimal experimental manipulations, which is not the default practice in ERP and behavioral production studies. Nonetheless, robust effects were still observed, by additional securing that extreme experimental manipulations, such as the process of familiarization, were eliminated. Of course, re-engineering a system, by observing deviants of typical behavior such as speech errors or evaluating how it operates under extreme situations, as in "infrequent derailments", is also useful in understanding the features and mechanisms that constitute it. But, since language production research ultimately aims to understand how humans communicate linguistic knowledge throughout their lives, future studies should incorporate experimental designs which observe effects in the simplest cases of language retrieval and develop theories that are both cognitively plausible and empirically verified by those examples.

6.5 Conclusion

Speakers generally select the appropriate word to denote their intents, according to both their idiosyncratic preferences and influenced by the words other speakers in their linguistic communities use. Here, I have shown that this linguistic variability in names indexes the co-activation of those entries in individuals' systems, evident in speakers' increased behavioral and neural effort, but not necessarily the systematic struggle to select the best word amongst numerous equally suitable alternatives.

CHAPTER 7-Appendices

Appendix A- Chapter 4

Table A1. Stimuli used in this study, and their dominant and secondary name frequencies of use in our norms (Oppenheim, in prep) and Phase 1 free naming (*i.e.*, before corrective feedback).

	Dominant Name			Secondary Name		
IPNP Object ID	**Name**	**Freq. in norms**	**Freq. in Phase 1**	**Name**	**Freq. in norms**	**Freq. in Phase 1**
obj002acorn	acorn	0.79	0.82	nut	0.13	0.06
obj004alligator	crocodile	0.86	0.18	alligator	0.12	0.82
obj006ant	ant	0.84	0.82	bug	0.07	0.06
obj007antlers	antlers	0.74	0.94	horns	0.20	0.06
obj017baby	baby	0.98	0.88	child	0.02	0.06
obj021badge	badge	0.46	0.06	medal	0.26	0.88
obj031basket	basket	0.98	0.82	hamper	0.02	0.06
obj037bed	bed	1.00	1.00	-	-	-
obj038bee	bee	0.51	0.82	wasp	0.22	0.06
obj041belt	belt	0.90	0.71	collar	0.05	0.12
obj045bird	bird	0.93	0.88	sparrow	0.04	-
obj047wood	wood	0.45	0.59	plank	0.33	0.35
obj049bomb	bomb	0.97	1.00	dynamite	0.01	-
obj051book	book	1.00	0.94	-	-	-
obj056box	box	0.97	0.94	cardboard box	0.02	0.06
obj059bra	bra	0.97	0.94	underwear	0.01	-
obj060bread	bread	0.94	1.00	toast	0.02	-
obj063broom	broom	0.50	0.65	brush	0.38	0.29
obj067butterfly	butterfly	0.95	1.00	moth	0.04	-
obj070cage	cage	0.95	0.82	cell	0.01	0.06
obj074can	can	0.63	0.53	tin	0.32	0.35
obj076cane	walking stick	0.43	0.47	cane	0.42	0.35
obj078canoe	boat	0.46	0.71	canoe	0.46	0.29
obj080hat	hat	0.51	0.24	cap	0.41	0.65
obj081car	car	0.99	0.82	-	-	-
obj082carousel	carousel	0.43	0.65	merry go round	0.43	0.06
obj083carrot	carrot	0.99	0.88	turnip	0.01	-
obj084tape	tape	0.58	0.47	cassette	0.28	0.35
obj095church	church	0.96	1.00	house	0.02	-
obj097city	city	0.48	0.41	town	0.33	0.41
obj099clock	clock	0.96	0.94	time	0.03	0.06
obj103coat	coat	0.64	0.76	jacket	0.28	-
obj105pillar	pillar	0.49	0.53	column	0.21	0.29
obj106comb	comb	0.96	0.76	brush	0.03	0.12
obj107cookie	biscuit	0.58	0.53	oreo	0.26	0.18
obj108cork	cork	0.62	0.35	can	0.05	0.06
obj111cow	cow	0.97	0.71	goat	0.01	0.12

obj114crackers	crackers	0.62	0.71	biscuits	0.25	0.12
obj115crib	cot	0.49	0.71	bed	0.37	-
obj123desert	cactus	0.56	0.71	desert	0.41	0.24
obj128dog	dog	1.00	1.00	-	-	-
obj133dragon	dragon	0.99	0.94	dinosaur	0.01	-
obj137drill	drill	0.59	0.65	screwdriver	0.13	0.12
obj138drum	drum	1.00	1.00			
obj141eagle	eagle	0.49	0.35	bird	0.31	0.35
obj153finger	finger	0.98	1.00	fingernail	0.01	-
obj156firetruck	fire engine	0.51	0.53	firetruck	0.34	0.24
obj157fish	fish	1.00	1.00			
obj162floor	floor	0.57	0.47	tiles	0.17	0.24
obj163flower	flower	0.95	0.88	sunflower	0.01	0.06
obj167football	rugby ball	0.45	0.53	football	0.25	0.24
obj171frog	frog	0.95	0.94	toad	0.05	0.06
obj173trash	rubbish	0.65	0.76	junk	0.10	0.06
obj177ghost	ghost	1.00	0.94			
obj179girl	girl	0.84	0.65	child	0.08	0.06
obj180glass	glass	0.67	0.53	cup	0.31	0.41
obj182globe	globe	0.91	0.94	world	0.03	0.06
obj184goat	goat	0.98	0.94	animal	0.01	-
obj189gun	gun	0.93	0.82	revolver	0.06	-
obj205highchair	highchair	0.61	0.65	chair	0.15	0.35
obj211horse	horse	0.98	1.00	pony	0.02	-
obj222puzzle	jigsaw	0.59	0.47	puzzle	0.36	0.47
obj226king	king	0.98	0.94	crown	0.01	-
obj228knife	knife	0.97	0.88	butterknife	0.02	0.12
obj232ladle	ladle	0.61	0.65	spoon	0.31	0.18
obj239leopard	leopard	0.55	0.35	cheetah	0.27	0.53
obj241lettuce	lettuce	0.56	0.76	cabbage	0.35	0.18
obj258mask	mask	1.00	0.82			
obj265priest	monk	0.52	0.24	priest	0.27	0.53
obj272mountain	mountain	0.96	0.88	volcano	0.02	0.06
obj277nail	nail	0.54	0.76	screw	0.36	0.24
obj278neck	neck	0.67	0.53	chin	0.25	0.41
obj279necklace	necklace	0.96	0.65	necklace of pearls	0.01	-
obj280needle	needle	0.58	0.41	sewing needle	0.06	0.59
obj281nest	nest	0.47	0.47	eggs	0.45	0.29
obj291package	parcel	0.62	0.47	package	0.24	0.41
obj292bucket	bucket	0.98	0.94	pail	0.01	0.06
obj294paint	paint	0.43	0.71	palette	0.17	0.18
obj304peach	peach	0.59	0.53	plum	0.12	0.12
obj321pinecone	pine cone	0.53	0.65	acorn	0.21	0.12
obj323pirate	pirate	0.95	0.94	captain	0.02	-

obj334porcupine	porcupine	0.53	0.53	animal	0.12	-
obj335pot	pot	0.46	0.41	pan	0.22	0.18
obj338priest	priest	0.59	0.71	vicar	0.29	0.12
obj341pyramid	pyramid	0.99	0.94	temple	0.01	-
obj342queen	queen	0.95	0.88	princess	0.01	0.06
obj343rabbit	rabbit	0.96	0.94	bunny	0.02	0.06
obj354gun	gun	0.73	0.82	rifle	0.20	0.06
obj355ring	ring	1.00	1.00	-	-	-
obj364rooster	chicken	0.64	0.65	cockerel	0.13	-
obj367rug	rug	0.58	0.82	carpet	0.23	0.12
obj375sandwich	sandwich	0.97	0.94	bread	0.03	-
obj391boat	boat	0.50	0.76	ship	0.36	0.24
obj393shoe	shoe	0.98	1.00	-	-	-
obj394shoulder	shoulder	0.68	0.88	arm	0.28	0.12
obj395shovel	spade	0.51	0.59	shovel	0.42	0.41
obj399skeleton	skeleton	1.00	0.82	-	-	-
obj403sled	sledge	0.54	0.35	sled	0.20	0.29
obj405slingshot	slingshot	0.58	0.71	catapult	0.14	0.06
obj407smoke	smoke	0.57	0.35	chimney	0.37	0.47
obj409snake	snake	1.00	0.82	-	-	-
obj412couch	sofa	0.75	0.47	couch	0.12	0.35
obj414spaghetti	spaghetti	0.92	0.76	pasta	0.04	0.18
obj415spatula	spatula	0.41	0.18	shovel	0.22	0.53
obj416spider	spider	0.99	0.94	spider web	0.01	-
obj418spoon	spoon	1.00	0.94	-	-	-
obj424stocking	tights	0.64	0.53	stocking	0.23	-
obj426stove	oven	0.43	0.12	cooker	0.34	0.71
obj428stroller	pram	0.62	0.71	pushchair	0.20	0.06
obj433sweater	jumper	0.87	0.12	sweater	0.05	0.76
obj435sword	sword	0.97	0.88	sabre	0.02	0.06
obj437table	table	0.99	1.00	desk	0.01	-
obj442tear	tear	0.64	0.53	crying	0.28	0.35
obj443teepee	tent	0.60	0.65	teepee	0.30	0.18
obj449tent	tent	1.00	0.88	-	-	-
obj452thumb	thumb	0.97	0.82	fingernail	0.01	0.06
obj458toilet	toilet	0.99	0.88	-	-	-
obj459tomato	tomato	0.96	0.82	apple	0.03	-
obj467train	train	0.96	0.94	lorry	0.02	-
obj469tree	tree	1.00	0.94	-	-	-
obj470tripod	tripod	0.53	0.71	stand	0.15	-
obj471trophy	trophy	0.58	0.53	cup	0.32	0.35
obj472truck	truck	0.59	0.41	lorry	0.37	0.53
obj473trumpet	trumpet	0.94	0.82	trombone	0.02	0.06
obj474chest	chest	0.50	0.41	box	0.31	0.35

obj476turtle	turtle	0.54	0.53	tortoise	0.46	0.41
obj477tweezers	tweezers	0.87	0.82	clippers	0.03	-
obj482vacuum	hoover	0.75	0.59	vacuum	0.18	0.29
obj484vest	waistcoat	0.63	0.65	vest	0.19	0.29
obj485violin	violin	0.91	1.00	guitar	0.06	-
obj486volcano	volcano	0.97	0.94	volcano eruption	0.02	0.06
obj496watch	watch	0.99	0.94	-	-	-
obj504wheelbarrow	wheelbarrow	0.95	1.00	wagon	0.01	-
obj508wig	wig	0.76	0.82	hair	0.22	0.18
obj512wing	wing	0.93	0.94	feathers	0.02	0.06
obj517wrench	spanner	0.53	0.71	wrench	0.19	0.06

Appendix B-Chapter 4

Results from the linear mixed effects regressions of ex-Gaussian components, estimated for each subject in each condition, including by-subject random intercepts. For the purpose of these analyses, the Name Agreement predictor was coded as a centered *binomial* contrast, but to match the main RT analyses the levels of the Name Agreement are set at their condition means {high=.96, low=.56}.

Table B1. LMM analyses for the ex-Gaussian μ (mu) component, from the naming latency data.

Both Phases

	β	*SE*	*t*	*p*
Intercept	801.73	17.21	46.58	<.001
Phase	46.51	17.24	2.70	.008
Name Agreement	-101.28	43.72	-2.32	.021
Name Change	38.98	17.24	2.26	.024
Phase*Name Agreement	368.56	87.43	4.22	<.001
Phase*Name Change	129.64	34.47	3.76	<.001
Name Agreement*Name Change	69.30	87.43	0.79	.43
Phase*Name Agreement* Name Change	88.31	174.86	0.51	.61

Phase 1 only

	β	*SE*	*t*	*p*
Intercept	778.48	23.76	32.77	<.001
Name Agreement	-285.56	52.83	-5.41	<.001
Name Change	-25.84	20.83	-1.24	.22
Name Agreement* Name Change	25.15	105.66	0.24	.81

Phase 3 only

	β	*SE*	*t*	*p*
Intercept	824.98	19.97	41.32	<.001
Name Agreement	83.00	54.30	1.53	.13
Name Change	103.79	21.41	4.85	<.001
Name Agreement* Name Change	113.45	108.59	1.04	.30

Table B2. LMM analyses for the ex-Gaussian σ (sigma) component, from the naming latency data.

Both Phases

	β	*SE*	*t*	*p*
Intercept	124.02	7.37	16.83	<.001
Phase	-11.36	9.71	-1.17	.24
Name Agreement	-113.07	24.62	-4.59	<.001
Name Change	5.16	9.71	0.53	.60
Phase*Name Agreement	155.11	49.24	3.15	.002
Phase*Name Change	60.51	19.41	3.12	.002
Name Agreement*Name Change	111.95	49.24	2.27	.025
Phase*Name Agreement* Name Change	109.48	98.48	1.11	.27

Phase 1 only

	β	*SE*	*t*	*p*
Intercept	129.70	10.00	12.97	<.001
Name Agreement	-190.62	34.09	-5.59	<.001
Name Change	-25.10	13.44	-1.87	.068
Name Agreement* Name Change	57.21	68.17	0.84	.41

Phase 3 only

	β	*SE*	*t*	*p*
Intercept	118.34	9.94	11.91	<.001

Name Agreement	-35.51	29.88	-1.19	.23
Name Change	35.41	11.78	3.01	.004
Name Agreement* Name Change	166.69	59.77	2.79	.008

Table B3. LMM analyses for the ex-Gaussian τ (tau) component, from the naming latency data.

Both Phases

	β	*SE*	*t*	*p*
Intercept	223.35	11.19	19.96	<.001
Phase	-9.94	14.25	-0.70	.49
Name Agreement	-206.59	36.14	-5.72	<.001
Name Change	32.60	14.25	2.29	.024
Phase*Name Agreement	9.72	72.27	0.13	.89
Phase*Name Change	24.05	28.49	0.84	.40
Name Agreement*Name Change	46.78	72.27	0.65	.52
Phase*Name Agreement* Name Change	-13.97	144.54	-0.10	.92

Phase 1 only

	β	*SE*	*t*	*p*
Intercept	228.32	12.75	17.90	<.001
Name Agreement	-211.45	51.80	-4.08	<.001
Name Change	20.57	20.42	1.01	.32
Name Agreement* Name Change	53.77	103.60	0.52	.61

Phase 3 only

	β	*SE*	*t*	*p*
Intercept	218.38	14.59	14.97	<.001
Name Agreement	-201.73	48.35	-4.17	<.001
Name Change	44.63	19.06	2.34	.023
Name Agreement* Name Change	39.80	96.70	0.41	.68

Appendix C-Chapter 4

Results from the linear mixed effects regressions of ERP amplitudes, estimated for each subject in each condition, including by-subject random slopes and intercepts, and by-electrode random intercepts. The Name Agreement predictor was coded as a centered binomial contrast {High, Low}, as was the Name Change predictor {No Change, Change}.

Table C1. LMM analyses for the N200 window.

Both Phases

	β	*SE*	*t*	*p*
(Intercept)	-0.65	0.64	-1.01	.33
Phase	1.07	0.32	3.36	.004
NameAgreement	0.48	0.18	2.65	.018
Phase x NameAgreement	-0.20	0.34	-0.59	.56

Phase 1 only

	β	*SE*	*t*	*p*
(Intercept)	-1.36	0.67	-2.02	.058
NameAgreement	0.61	0.23	2.68	.016

Phase 3 only

	β	*SE*	*t*	*p*
(Intercept)	-0.29	0.65	-0.44	.66
NameChange	-0.44	0.22	-2.02	.061
NameAgreement	0.41	0.25	1.65	.12
NameChange x NameAgreement	-0.30	0.54	-0.56	.58

Table C2. LMM analyses for the N400 window.

Both Phases

	β	*SE*	*t*	*p*
(Intercept)	-0.79	0.55	-1.42	.17
Phase	0.66	0.33	2.02	.06
NameAgreement	0.67	0.18	3.67	.002
Phase x NameAgreement	-0.39	0.29	-1.38	.17

Phase 1 only

	β	*SE*	*t*	*p*
(Intercept)	-1.23	0.58	-2.11	.049
NameAgreement	0.94	0.20	4.63	<.001

Phase 3 only

	β	*SE*	*t*	*p*
(Intercept)	-0.56	0.56	-1.00	.33
NameChange	-0.02	0.27	-0.08	.93
NameAgreement	0.54	0.24	2.31	.035
NameChange x NameAgreement	0.54	0.39	1.37	.19

Table C3. LMM analyses for the late anterior positive modulation.

Phase 3 only

	β	*SE*	*t*	*p*
(Intercept)	0.78	0.46	1.71	0.1
NameChange	1.01	0.41	2.44	0.027

Appendix D-Chapter 5

Including Lexical Frequency as a predictor of naming latency analysis and ERP analysis

1. Naming Latencies

Lexical frequency values were obtained from the SUBTLEX-UK database (Van Heuven, Mandera, Keuleers, & Brysbaert, 2014). The mixed effects model predicted naming latencies as a function of (1) Session (an ordinal measure from 1:2, centered), (2) dominant name agreement from Oppenheim's (in prep) recent Bangor norming study (a continuous measure from 0:1, centered), (3) secondary name agreement from Oppenheim's (in prep.) recent Bangor norming study (a continuous measure from 0:.48, centered) and (4) lexical frequency from SUBLTEX, UK (Zipf[7] values from 0:6.56, centered).

Table 2. Summary of LMM analyses of inverse-transformed naming latencies.

Both Sessions

	β	*SE*	*t*	*p*
Intercept	-9.996	.253	-39.470	<.001***
Session	-.278	.118	-2.347	<.001***
Dominant Name Agreement	-7.827	.418	-18.691	<.001***
Secondary Name Agreement	-5.884	.634	-9.271	<.001***
Lexical Frequency	-.239	.047	-5.104	<.001***
Session*Dominant Name Agreement	1.298	.323	4.017	.010
Session*Secondary Name Agreement	1.904	.483	3.942	<.001***
Session*Lexical Frequency	.138	.029	4.620	<.001***
Dominant Name Agreement*Lexical Frequency	-1.165	.242	-4.810	<.001***
Secondary Name Agreement*Lexical Frequency	-.607	.422	-1.439	<.001***
Session*Dominant Name Agreement*Lexical Frequency	.482	.209	2.303	.142
Session*Secondary Name Agreement*Lexical Frequency	.566	.319	1.773	.074

Session 1 only

	β	*SE*	*t*	*p*
Intercept	-9.871	.240	-41.113	<.001**
Dominant Name Agreement	-8.674	.478	-18.126	<.001***
Secondary Name Agreement	-7.066	.717	-9.848	.009**
Lexical Frequency	-0.309	.054	-5.674	<.001***
Dominant Name Agreement*Lexical Frequency	-1.477	.285	-.5.175	<.001
Secondary Name Agreement*Lexical Frequency	-.991	.482	-.2.073	<.001

Session 2 only

[7] The Zipf scale (Van Heuven, Mandera, Keuleers, & Brysbaert, 2014) is a standardized measure of lexical frequency. A Zipf value of 1 corresponds to words with frequencies of 1 per 100 million words, a Zipf value of 2 corresponds to words with frequencies of 1 per 10 million words, and so on. For example, the word "dog" which is considered a highly frequent word in British English has a Zipf value of 5.17, while a less frequent word, such as "eskimo" has a Zipf value of 2.84.

	β	*SE*	*t*	*p*
Intercept	-10.11	.278	-36.311	<.001**
Dominant Name Agreement	-7.438	.431	-17.233	<.001***
Secondary Name Agreement	-5.258	.642	-8.183	.001**
Lexical Frequency	1.164	.043	-3.741	<.001**
Dominant Name Agreement*Lexical Frequency	-.979	.244	-4.009	<.001**
Secondary Name Agreement*Lexical Frequency	-.397	.426	-.938	.024*

2. ERPS

In LIMO analysis, I used Lexical Frequency as a continuous predictor and Session as a categorical predictor for the analyses. The interaction between Lexical Frequency and Session were investigated using two repeated measures ANOVAs, with Frequency and Session as repeated measures for each analysis.

Lexical Frequency

The main effect of Lexical Frequency was also examined using a one-sample t test. However, the main effect of Lexical Frequency did not reach statistical significance ($p > .01$), when I corrected for multiple comparisons.

*Session * Lexical Frequency Interaction*

An interaction between Lexical Frequency and Session, also failed to provide a significant result ($p > .01$), once corrected for multiple comparisons ($p > .01$).

Post-hoc t tests on Lexical Frequency

In an exploratory analysis and even though the main effect of Lexical Frequency did not reach statistical significance, the one sample t-tests in the two sessions revealed that in the first Session the effect of Lexical Frequency on ERPs was not statistically significant, but in the second Session the effect reached statistical significance (see Figure 1) ($p < .01$). Lexical frequency modulated ERP amplitudes as early as 100 ms post picture onset at parietal and occipital electrodes (P1), continuing at 200ms with the same topographic distribution (P2) and further modulating the ERP signal from 250-420 onwards and 500-600 ms post-stimulus onset. Notably, the effects of dominant name agreement with lexical frequency have a slight temporal and spatial overlap, however, their polarity (see *t* value in Figures 2b Chapter 5, 1b & 2b here) is reverse, with effects of Lexical Frequency arising at frontal and fronto-central areas of the scalp in the right hemisphere and lexical frequency affecting parietal and occipital areas mostly in the left scalp.

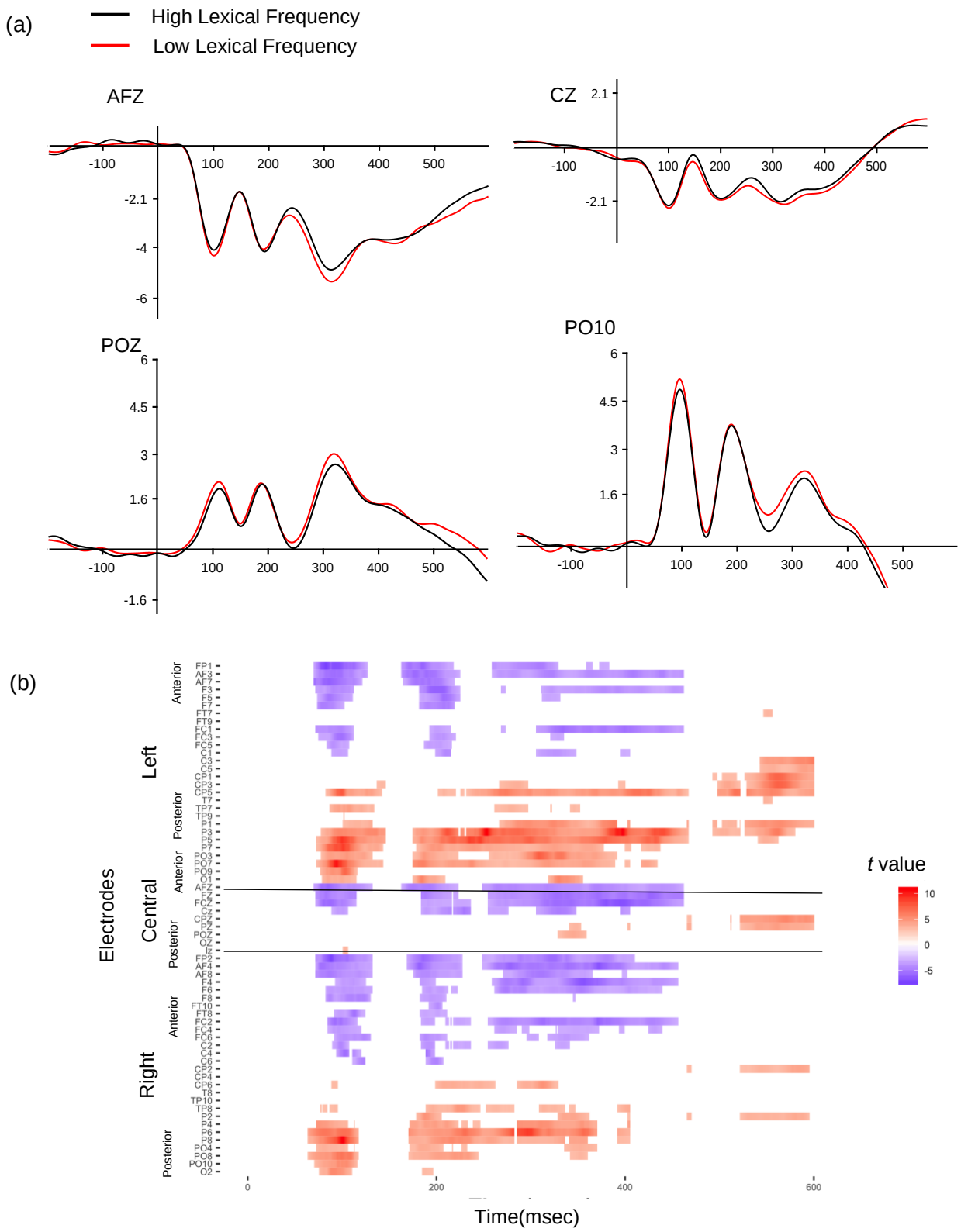

Figure1.(a) Grand-average ERP waveforms elicited high and low frequency pictures in the second naming session. ERPs were computed for a visual illustration, based on a mean split on the dataset, resulting in 245 items of low lexical frequency (MeanFreq < 4.11) and 280 items of high lexical frequency (MeanFreq > 4.11). (*b*) Results of the robust one sample t-test (representation of significant t-values) for lexical frequency after correction for multiple comparisons at all electrodes and timepoints; with the colors representing electrodes with higher (red) and lower (blue) amplitudes for trials with lexical frequency; each line is an electrode.

References

Abdel Rahman, R., & Melinger, A. (2007). When bees hamper the production of honey: lexical interference from associates in speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(3), 604. https://doi.org/10.1037/0278-7393.33.3.604

Abdel Rahman, R., & Melinger, A. (2009). Dismissing lexical competition does not make speaking any easier: A rejoinder to Mahon and Caramazza (2009). *Language and Cognitive Processes*, *24*(5), 749-760. https://doi.org/10.1080/01690960802648491

Abdel Rahman, R., & Sommer, W. (2008). Seeing what we know and understand: How knowledge shapes perception. *Psychonomic bulletin & review*, *15*(6), 1055-1063. https://doi.org/10.3758/pbr.15.6.1055

Abdel Rahman, R., Van Turennout, M., & Levelt, W. J. (2003). Phonological encoding is not contingent on semantic feature retrieval: An electrophysiological study on object naming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(5), 850. https://doi.org/10.1037/0278-7393.29.5.850

Abel, S., Dressel, K., Bitzer, R., Kümmerer, D., Mader, I., Weiller, C., & Huber, W. (2009). The separation of processing stages in a lexical interference fMRI-paradigm. *Neuroimage*, *44*(3),11131124. https://doi.org/10.1016/j.neuroimage.2008.10.018

Anders, R., Riès, S., Van Maanen, L., & Alario, F. X. (2015). Evidence accumulation as a model for lexical selection. *Cognitive Psychology*, *82*, 57-73. https://doi.org/10.1016/j.cogpsych.2015.07.002

Alario, F. X., Segui, J., & Ferrand, L. (2000). Semantic and associative priming in picture naming. *The Quarterly Journal of Experimental Psychology: Section A*, *53*(3), 741-764. ttps://doi.org/10.1080/713755907

Alario, F. X., Ferrand, L., Laganaro, M., New, B., Frauenfelder, U. H., & Segui, J. (2004). Predictors of picture naming speed. *Behavior Research Methods, Instruments, & Computers*, *36*(1), 140-155. https://doi.org/10.3758/bf03195559

Alonso-Prieto, E., Pancaroglu, R., Dalrymple, K. A., Handy, T., Barton, J. J., & Oruc, I. (2015). Temporal dynamics of the face familiarity effect: bootstrap analysis of single-subject event-related potential data. *Cognitive neuropsychology*, *32*(5), 266-282. https://doi.org/10.1080/02643294.2015.1053852

Aristei, S., Zwitserlood, P., & Abdel Rahman, R. (2012). Picture-induced semantic interference reflects lexical competition during object naming. *Frontiers in psychology*, *3*, 28. https://doi.org/10.3389/fpsyg.2012.00028

Assaneo, M. F., Ripollés, P., Orpella, J., Lin, W. M., de Diego-Balaguer, R., & Poeppel, D. (2019). Spontaneous synchronization to speech reveals neural mechanisms facilitating language learning. *Nature neuroscience*, *22*(4), 627-632. https://doi.org/10.1038/s41593-019-0353-z

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of memory and language*, *59*(4), 390-412. https://doi.org/10.1016/j.jml.2007.12.005

Bakhtiar, M., Nilipour, R., & Weekes, B. S. (2013). Predictors of timed picture naming in Persian. *Behavior Research Methods*, *45*(3), 834-841. https://doi.org/10.3758/s13428-012-0298-6

Balaguer, R. D. D., Sebastián-Gallés, N., Díaz, B., & Rodríguez-Fornells, A. (2005). Morphological processing in early bilinguals: An ERP study of regular and irregular verb processing. *Cognitive Brain Research*, *25*(1), 312-327. https://doi.org/10.1016/j.cogbrainres.2005.06.003

Balatsou, E., Fischer-Baum, S., & Oppenheim, G. M. (in revision). *The psychological validity of picture name agreement*. *Cognition.*

Barsalou, L. W. (1983). Ad hoc categories. *Memory & cognition*, *11*(3), 211-227. https://doi.org/10.3758/bf03196968

Bates, E., D'Amico, S., Jacobsen, T., Székely, A., Andonova, E., Devescovi, A., ... & Wicha, N. (2003). Timed picture naming in seven languages. *Psychonomic bulletin & review*, *10*(2), 344-380. https://doi.org/10.3758/bf03196494

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2016). lme4: Linear mixed- effects models using Eigen and S4. 2014. R package version 1.1-7. *Google Scholar*. https://doi.org/10.18637/jss.v067.i01

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015a). Fitting linear mixed-effects models using lme4. Journal of Statistical Software, 67(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Barkley, C., Kluender, R., & Kutas, M. (2015). Referential processing in the human brain: An Event-Related Potential (ERP) study. *Brain research*, *1629*, 143-159. https://doi.org/10.1016/j.brainres.2015.09.017

Barber, H., Vergara, M., & Carreiras, M. (2004). Syllable-frequency effects in visual word recognition: evidence from ERPs. *Neuroreport*, *15*(3), 545-548. https://doi.org/10.1097/00001756-200403010-00032

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of memory and language*, *68*(3), 255-278. https://doi.org/10.1016/j.jml.2012.11.001

Barrett, S. E., & Rugg, M. D. (1990). Event-related potentials and the semantic matching of pictures. *Brain and cognition*, *14*(2), 201-212. https://doi.org/10.1016/0278-2626(90)90029-n

Barry, C., Morrison, C. M., & Ellis, A. W. (1997). Naming the Snodgrass and Vanderwart pictures: Effects of age of acquisition, frequency, and name agreement. *The Quarterly Journal of*

Experimental Psychology: Section A, *50*(3), 560-585. https://doi.org/10.1080/027249897392026

Becker, S., Moscovitch, M., Behrmann, M., & Joordens, S. (1997). Long-term semantic priming: A computational account and empirical evidence. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*(5), 1059. https://doi.org/10.1037/0278-7393.23.5.1059

Belke, E. (2008). Effects of working memory load on lexical-semantic encoding in language production. *Psychonomic bulletin & review*, *15*(2), 357-363. https://doi.org/10.3758/pbr.15.2.357

Belke, E. (2013). Long-lasting inhibitory semantic context effects on object naming are necessarily conceptually mediated: Implications for models of lexical-semantic encoding. *Journal of Memory and Language*, *69*(3), 228-256. https://doi.org/10.1016/j.jml.2013.05.008

Belke, E. (2017). The role of task-specific response strategies in blocked-cyclic naming. *Frontiers in Psychology*, *7*, 1955. https://doi.org/10.3389/fpsyg.2016.01955

Belke, A., & Dreger, C. (2013). Current Account Imbalances in the Euro A rea: Does Catching up Explain the Development?. *Review of International Economics*, *21*(1), 6-17. https://doi.org/10.1111/roie.12016

Belke, A., Gros, D., & Osowski, T. (2017). The effectiveness of the Fed's quantitative easing policy: New evidence based on international interest rate differentials. *Journal of International Money and Finance*, *73*, 33534. https://doi.org/10.1016/j.jimonfin.2017.02.011

Belke, E., Meyer, A. S., & Damian, M. F. (2005). Refractory effects in picture naming as assessed in a semantic blocking paradigm. Quarterly Journal of Experimental Psychology, 58, 667–692. https://doi.org/10.1080/02724980443000142

Beres (2017). Time is of the essence: A review of electroencephalography (EEG) and event-related brain potentials (ERPs) in language research. *Applied psychophysiology and biofeedback*, 42(4), 247-255). https://doi.org/10.1007/s10484-017-9371-3

Best, W., Herbert, R., Hickin, J., Osborne, F., & Howard, D. (2002). Phonological and orthographic facilitation of word-retrieval in aphasia: Immediate and delayed effects. *Aphasiology*, *16*(1-2), 151-168. https://doi.org/10.1080/02687040143000483

Bierwisch, M., & Schreuder, R. (1992). From concepts to lexical items. *Cognition*, *42*(1-3), 23-60. https://doi.org/10.1016/0010-0277(92)90039-k

Blackford, T., Holcomb, P. J., Grainger, J., & Kuperberg, G. R. (2012). A funny thing happened on the way to articulation: N400 attenuation despite behavioral interference in picture naming. *Cognition*, *123*(1), 84-99. https://doi.org/10.1016/j.cognition.2011.12.007

Bloem, I., & La Heij, W. (2003). Semantic facilitation and semantic interference in word translation: Implications for models of lexical access in language production. *Journal of Memory and language*, *48*(3), 468-488. https://doi.org/10.1037/e501882009-509

Bloem, I., van den Boogaard, S., & La Heij, W. (2004). Semantic facilitation and semantic interference in language production: Further evidence for the conceptual selection model of lexical access. *Journal of Memory and Language*, *51*(2), 307-323. https://doi.org/10.1016/j.jml.2004.05.001

Bodner, G. E., & Masson, M. E. (1997). Masked repetition priming of words and nonwords: Evidence for a nonlexical basis for priming. *Journal of Memory and Language*, *37*(2), 268-293. https://doi.org/10.1006/jmla.1996.2507

Boisgontier, M. P., & Cheval, B. (2016). The anova to mixed model transition. *Neuroscience & Biobehavioral Reviews*, *68*, 1004 1005. https://doi.org/10.1016/j.neubiorev.2016.05.034

Bonin, P., Chalard, M., Méot, A., & Fayol, M. (2002). The determinants of spoken and written picture naming latencies. *British Journal of Psychology*, *93*(1), 89-114. https://doi.org/10.1348/000712602162463

Bookheimer, S. Y., Zeffiro, T. A., Blaxton, T., Gaillard, W., & Theodore, W. (1995). Regional cerebral blood flow during object naming and word reading. *Human Brain Mapping*, *3*(2), 93-106. https://doi.org/10.1002/hbm.460030206

Bormann, T., Kulke, F., Wallesch, C. W., & Blanken, G. (2008). Omissions and semantic errors in aphasic naming: Is there a link?. *Brain and language*, *104*(1), 24-32. https://doi.org/10.1016/j.bandl.2007.02.004

Bose, A., & Schafer, G. (2017). Name agreement in aphasia. *Aphasiology*, *31*(10), 1143-1165. https://doi.org/10.1080/02687038.2016.1254148

Boudewyn, M. A., Luck, S. J., Farrens, J. L., & Kappenman, E. S. (2018). How many trials does it take to get a significant ERP effect? It depends. *Psychophysiology*, *55*(6), e13049. https://doi.org/10.1111/psyp.13049

Bridwell, D. A., Cavanagh, J. F., Collins, A. G., Nunez, M. D., Srinivasan, R., Stober, S., & Calhoun, V. D. (2018). Moving beyond ERP components: a selective review of approaches to integrate EEG and behavior. *Frontiers in human neuroscience*, *12*, 106. https://doi.org/10.3389/fnhum.2018.00106

Britt, A. E., Ferrara, C., & Mirman, D. (2016). Distinct effects of lexical and semantic competition during picture naming in younger adults, older adults, and people with aphasia. *Frontiers in Psychology*, *7*, 813. https://doi.org/10.3389/fpsyg.2016.00813

Bruin, K. J., Wijers, A. A., & Van Staveren, A. S. J. (2001). Response priming in a go/nogo task: do we have to explain the go/nogo N2 effect in terms of response activation instead of inhibition?. *Clinical Neurophysiology*, *112*(9), 1660-1671. https://doi.org/10.1016/s1388-2457(01)00601-0

Bürki, A., Elbuy, S., Madec, S., & Vasishth, S. (2020). What did we learn from forty years of research on semantic interference? A Bayesian meta-analysis. *Journal of Memory and Language*, *114*(April), 104125. https://doi.org/10.1016/j.jml.2020.104125

Bürki, A., Elbuy, S., Madec, S., & Vasishth, S. (2020). Picture naming in the context of semantically related distractors: A Bayesian meta-analysis 40 years after Lupker 1979s first reaction time study. https://doi.org/10.1016/j.jml.2020.104125

Bürki, A., Frossard, J., & Renaud, O. (2018). Accounting for stimulus and participant effects in event-related potential analyses to increase the replicability of studies. *Journal of neuroscience methods*, *309*, 218227. https://doi.org/10.1016/j.jneumeth.2018.09.016

Butterfield, G. B., & Butterfield, E. C. (1977). Lexical codability and age. *Journal of Verbal Learning and Verbal Behavior*, *16*(1), 113-118. https://doi.org/10.1016/s0022-5371(77)80013-3

Cameron-Jones, C. M., & Wilshire, C. E. (2007). Lexical competition effects in two cases of non-fluent aphasia. *Brain and Language*, *103*(1-2), 136-137. https://doi.org/10.1016/j.bandl.2007.07.083

Caramazza, A. (1997). How many levels of processing are there in lexical access?. *Cognitive neuropsychology*, *14*(1), 177-208. https://doi.org/10.1080/026432997381664

Caramazza, A., Hillis, A. E., Rapp, B. C., & Romani, C. (1990). The multiple semantics hypothesis: Multiple confusions?. *Cognitive neuropsychology*, *7*(3), 161-189. https://doi.org/10.1080/02643299008253441

Carriero, L., Zalla, T., Budai, R., & Battaglini, P. P. (2007). Inhibition of wrong responses and conflict resolution: an electroencephalogram study. *Neuroreport*, *18*(8), 793-796. https://doi.org/10.1097/wnr.0b013e3280c1e330

Carroll, J. B., & White, M. N. (1973). Word frequency and age of acquisition as determiners of picture-naming latency. *The Quarterly Journal of Experimental Psychology*, *25*(1), 85-95. https://doi.org/10.1080/14640747308400325

Cattell, J. M. (1886). The time it takes to see and name objects. *Mind*, *11*(41), 63-65. https://doi.org/10.1093/mind/os-xi.41.63

Cave, C. B. (1997). Very long-lasting priming in picture naming. *Psychological science*, *8*(4), 322-325. https://doi.org/10.1111/j.1467-9280.1997.tb00446.x

Chen, S., & Bates, E. (1998). The dissociation between nouns and verbs in Broca's and Wernicke's aphasia: Findings from Chinese. *Aphasiology*, *12*(1), 5-36. https://doi.org/10.1080/02687039808249441

Cheng, X., Schafer, G., & Akyürek, E. G. (2010). Name agreement in picture naming: an ERP study. *International Journal of Psychophysiology*, *76*(3), 130-141. https://doi.org/10.1016/j.ijpsycho.2010.03.003

Clark, H. H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of verbal learning and verbal behavior*, *12*(4), 335-359. https://doi.org/10.1016/s0022-5371(73)80014-3

Collina, S., Tabossi, P., & De Simone, F. (2013). Word Production and the Picture-Word Interference Paradigm: The Role of Learning. *Journal of Psycholinguistic Research*, *42*(5), 461–473. https://doi.org/10.1007/s10936-012-9229-z

Costa, A., Alario, F. X., & Caramazza, A. (2005). On the categorical nature of the semantic interference effect in the picture-word interference paradigm. *Psychonomic Bulletin & Review*, *12*(1), 125-131. https://doi.org/10.3758/bf03196357

Costa, A., Caramazza, A., & Sebastian-Galles, N. (2000). The cognate facilitation effect: implications for models of lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(5), 1283. https://doi.org/10.1037/0278-7393.26.5.1283

Costa, A., Strijkers, K., Martin, C., & Thierry, G. (2009). The time course of word retrieval revealed by event-related brain potentials during overt speech. *Proceedings of the National Academy of Sciences*, *106*(50), 21442-21446. https://doi.org/10.1093/cercor/bhp153

Crowther, J. E., & Martin, R. C. (2014). Lexical selection in the semantically blocked cyclic naming task: the role of cognitive control and learning. *Frontiers in human neuroscience*, *8*, 9. https://doi.org/10.3389/fnhum.2014.00009

Cuetos, F., Ellis, A. W., & Alvarez, B. (1999). Naming times for the Snodgrass and Vanderwart pictures in Spanish. *Behavior Research Methods*, *31*(4), 650-658. https://doi.org/10.3758/bf03200741

Cutting, J. C., & Ferreira, V. S. (1999). Semantic and phonological information flow in the production lexicon. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(2), 318. https://doi.org/10.1037/0278-7393.25.2.318

Cycowicz, Y. M., Friedman, D., Rothstein, M., & Snodgrass, J. G. (1997). Picture naming by young children: Norms for name agreement, familiarity, and visual complexity. *Journal of experimental child psychology*, *65*(2), 171-237. https://doi.org/10.1006/jecp.1996.2356

Damasio, H., Grabowski, T. J., Tranel, D., Ponto, L. L., Hichwa, R. D., & Damasio, A. R. (2001). Neural correlates of naming actions and of naming spatial relations. *Neuroimage*, *13*(6), 1053-1064. https://doi.org/10.1006/nimg.2001.0775

Damian, M. F., & Als, L. C. (2005). Long-lasting semantic context effects in the spoken production of object names. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(6), 1372. https://doi.org/10.1037/0278-7393.31.6.1372

Damian, M. F., & Bowers, J. S. (2003). Locus of semantic interference in picture-word interference tasks. *Psychonomic Bulletin & Review*, *10*(1), 111-117. https://doi.org/10.3758/bf03196474

D'Amico, S., Devescovi, A., & Bates, E. (2001). Picture naming and lexical access in Italian children and adults. *Journal of Cognition and Development*, *2*(1), 71-105. https://doi.org/10.1207/s15327647jcd0201_4

Davidoff, J., & Masterson, J. (1996). The development of picture naming: Differences between verbs and nouns. *Journal of Neurolinguistics*, *9*(2), 69-83. https://doi.org/10.1016/0911-6044(96)00004-8

Deacon, D., & Shelley-Tremblay, J. (2000). How automatically is meaning accessed: a review of the effects of attention on semantic processing. *Frontiers in Bioscience*, *5*(Part E), 82-94. https://doi.org/10.2741/a569

Dell, G. S. (1984). Representation of serial order in speech: evidence from the repeated phoneme effect in speech errors. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *10*(2), 222. https://doi.org/10.1037/0278-7393.10.2.222

Dell, G.S. (1986). A spreading-activation theory of retrieval in sentence production. Psychological Review, 93, 283–321. https://doi.org/10.1037/0033-295x.93.3.283

Dell, G. S. (1988). The retrieval of phonological forms in production: Tests of predictions from a connectionist model. *Journal of memory and language*, 27(2), 124-142. https://doi.org/10.1016/0749-596x(88)90070-8

Dell, G. S., Nozari, N., & Oppenheim, G. M. (2014). Word production: Behavioral and computational considerations. *The Oxford handbook of language production*, 88-104. https://doi.org/10.1093/oxfordhb/9780199735471.013.014

Dell, G. S., Schwartz, M. F., Martin, N., Saffran, E. M., & Gagnon, D. A. (1997). Lexical access in aphasic and nonaphasic speakers. *Psychological review*, *104*(4), 801. https://doi.org/10.1037/0033-295x.104.4.801

Dell, G. S., & O'Seaghdha, P. G. (1991). Mediated and convergent lexical priming in language production: A comment on Levelt et al(1991). https://doi.org/10.1037/0033-295x.98.4.604

Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of neuroscience methods*, *134*(1), 9-21. https://doi.org/10.1016/j.jneumeth.2003.10.009

Dell'Acqua, R., Job, R., Peressotti, F., & Pascali, A. (2007). The picture-word interference effect is not a Stroop effect. *Psychonomic bulletin & review*, *14*(4), 717-722. https://doi.org/10.3758/bf03196827

Dell'acqua, R., Lotto, L., & Job, R. (2000). Naming times and standardized norms for the Italian PD/DPSS set of 266 pictures: Direct comparisons with American, English, French, and Spanish published databases. *Behavior Research Methods, Instruments, & Computers*, *32*(4), 588-615. https://doi.org/10.3758/bf03200832

Dell'Acqua, R., Sessa, P., Peressotti, F., Mulatti, C., Navarrete, E., & Grainger, J. (2010). ERP evidence for ultra-fast semantic processing in the picture–word interference paradigm. *Frontiers in psychology*, *1*,177. https://doi.org/10.3389/fpsyg.2010.00177

De Zubicaray, G. I., Hansen, S., & McMahon, K. L. (2013). Differential processing of thematic and categorical conceptual relations in spoken word production. *Journal of Experimental Psychology: General*, *142*(1), 131. https://doi.org/10.1037/a0028717

De Zubicaray, G. I., & McMahon, K. L. (2009). Auditory context effects in picture naming investigated with event-related fMRI. *Cognitive, Affective, & Behavioral Neuroscience*, *9*(3), 260-269.https://doi.org/10.3758/cabn.9.3.260

De Zubicaray, G., McMahon, K., Eastburn, M., & Pringle, A. (2006). Top-down influences on lexical selection during spoken word production: A 4T fMRI investigation of refractory effects in picture naming. *Human brain mapping*, *27*(11), 864-873. https://doi.org/10.1002/hbm.20227

Dhooge, E., De Baene, W., & Hartsuiker, R. J. (2013). A late locus of the distractor frequency effect in picture–word interference: Evidence from event-related potentials. *Brain and language*, *124*(3), 232-237. https://doi.org/10.1016/j.bandl.2012.12.005

Di Flumeri, G., Aricó, P., Borghini, G., Colosimo, A., & Babiloni, F. (2016, August). A new regression-based method for the eye blinks artifacts correction in the EEG signal, without using any EOG channel. In *2016 38th Annual International Conference of the IEEE Engineering in*

Medicine and Biology Society (EMBC) (pp. 3187-3190). IEEE. https://doi.org/10.1109/embc.2016.7591406

Dimitropoulou, M., Duñabeitia, J. A., Blitsas, P., & Carreiras, M. (2009). A standardized set of 260 pictures for Modern Greek: Norms for name agreement, age of acquisition, and visual complexity. *Behavior Research Methods*, *41*(2), 584-589. https://doi.org/10.3758/brm.41.2.584

Dhooge, E., De Baene, W., & Hartsuiker, R. J. (2013). A late locus of the distractor frequency effect in picture–word interference: Evidence from event-related potentials. *Brain and language*, *124*(3), 232-237. https://doi.org/10.1016/j.bandl.2012.12.005

Dhooge, E., & Hartsuiker, R. J. (2010). The distractor frequency effect in picture–word interference: Evidence for response exclusion. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(4), 878. https://doi.org/10.1037/a0019128

Dhooge, E., & Hartsuiker, R. J. (2011). The distractor frequency effect in a delayed picture-word interference task: Further evidence for a late locus of distractor exclusion. *Psychonomic Bulletin & Review*, *18*(1), 116-122. https://doi.org/10.3758/s13423-010-0026-0

Dockrell, J. E., Messer, D., & George, R. (2001). Patterns of naming objects and actions in children with word finding difficulties. *Language and Cognitive processes*, *16*(2-3), 261-286. https://doi.org/10.1080/01690960042000030

Donchin, E., Ritter, W., & McCallum, W. C. (1978). Cognitive psychophysiology: The endogenous components of the ERP. *Event-related brain potentials in man*, *349*, 411. https://doi.org/10.1016/b978-0-12-155150-6.50019-5

Donkers, F. C., & Van Boxtel, G. J. (2004). The N2 in go/no-go tasks reflects conflict monitoring not response inhibition. *Brain and cognition*, *56*(2), 165-176. https://doi.org/10.1016/j.bandc.2004.04.005

Dong, G., Yang, L., Hu, Y., & Jiang, Y. (2009). Is N2 associated with successful suppression of behavior responses in impulse control processes?. *Neuroreport*, *20*(6), 537-542. https://doi.org/10.1097/wnr.0b013e3283271e9b

Druks, J., & Froud, K. (2002). The syntax of single words: Evidence from a patient with a selective function word reading deficit. *Cognitive Neuropsychology*, *19*(3), 207-244. https://doi.org/10.1080/02643290143000141

Druks, J., & Shallice, T. (2000). Selective preservation of naming from description and the "restricted preverbal message". *Brain and Language*, *72*(2), 100-128. https://doi.org/10.1006/brln.1999.2165

Dylman, A. S., & Barry, C. (2018). When having two names facilitates lexical selection: Similar results in the picture-word task from translation distractors in bilinguals and synonym distractors in monolinguals. *Cognition*, *171*, 151-171. https://doi.org/10.1016/j.cognition.2017.09.014

Eimer, M. (1993). Effects of attention and stimulus probability on ERPs in a Go/Nogo task. *Biological psychology*, *35*(2), 123-138. https://doi.org/10.1016/0301-0511(93)90009-w

Ellis, A. W., & Morrison, C. M. (1998). Real age-of-acquisition effects in lexical retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*(2), 515. https://doi.org/10.1037/0278-7393.24.2.515

Ellis, A. W., Young, A. W., Flude, B. M., & Hay, D. C. (1987). Repetition priming of face recognition. *The Quarterly Journal of Experimental Psychology Section A*, *39*(2), 193-210. ttps://doi.org/10.1080/14640748708401784

Falkenstein, M., Hoormann, J., & Hohnsbein, J. (1999). ERP components in Go/Nogo tasks and their relation to inhibition. *Acta psychologica*, *101*(2-3), 267-291. https://doi.org/10.1016/s0001-6918(99)00008-6

Fieder, N., Nickels, L., Biedermann, B., & Best, W. (2014). From “some butter” to “a butter”: An investigation of mass and count representation and processing. *Cognitive neuropsychology*, *31*(4), 313-349. https://doi.org/10.1080/02643294.2014.903914

Fisher, S. E. (2019). Human genetics: the evolving story of FOXP2. *Current Biology*, *29*(2), R65-R67. https://doi.org/10.1016/j.cub.2018.11.047

Finkbeiner, M., & Caramazza, A. (2006). Now you see it, now you don't: On turning semantic interference into facilitation in a Stroop-like task. *Cortex*, *42*(6), 790-796. https://doi.org/10.1016/s0010-9452(08)70419-2

Fodor, J. A. (1983). *The modularity of mind*. MIT press. https://doi.org/10.7551/mitpress/4737.001.0001

Forster, K. I., & Davis, C. (1984). Repetition priming and frequency attenuation in lexical access. *Journal of experimental psychology: Learning, Memory, and Cognition*, *10*(4), 680. https://doi.org/10.1037/0278-7393.10.4.680

Fraisse, P. (1968). Motor and verbal reaction times to words and drawings. *Psychonomic Science*, *12*(6), 235-236. https://doi.org/10.3758/bf03331287

Fraisse, P. (1969). Why is naming longer than reading?. *Acta Psychologica*, *30*, 96-103. https://doi.org/10.1016/0001-6918(69)90043-2

Francis, W. S. (2014). Repetition priming in picture naming: Sustained learning through the speeding of multiple processes. *Psychonomic bulletin & review*, *21*(5), 1301-1308. https://doi.org/10.3758/s13423-014-0610-9

Francis, W. S., & Sáenz, S. P. (2007). Repetition priming endurance in picture naming and translation: Contributions of component processes. *Memory & Cognition*, *35*(3), 481-493. https://doi.org/10.3758/bf03193288

Friederici, A. D., Hahne, A., & Mecklinger, A. (1996). Temporal structure of syntactic parsing: early and late event-related brain potential effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(5), 1219. https://doi.org/10.1037/0278-7393.22.5.1219

Friedrich, C. K., Eulitz, C., & Lahiri, A. (2006). Not every pseudoword disrupts word recognition: an ERP study. *Behavioral and Brain Functions*, *2*(1), 36. https://doi.org/10.1186/1744-9081-2-36

Ganushchak, L. Y., & Schiller, N. O. (2008). Motivation and semantic context affect brain error-monitoring activity: an event-related brain potentials study. *Neuroimage*, *39*(1), 395-405. https://doi.org/10.1016/j.neuroimage.2007.09.001

Garrod, S., & Anderson, A. (1987). Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. *Cognition*, *27*(2), 181-218. https://doi.org/10.1016/0010-0277(87)90018-7

Gaspar, C. M., Rousselet, G. A., & Pernet, C. R. (2011). Reliability of ERP and single-trial analyses. *Neuroimage*, *58*(2), 620-629. https://doi.org/10.1016/j.neuroimage.2011.06.052

Gauvin, H. S., Jonen, M. K., Choi, J., McMahon, K., & de Zubicaray, G. I. (2018). No lexical competition without priming: Evidence from the picture–word interference paradigm. *Quarterly Journal of Experimental Psychology*, *71*(12), 2562–2570. https://doi.org/10.1177/1747021817747266

Gehring, W. J., Goss, B., Coles, M. G., Meyer, D. E., & Donchin, E. (2018). The error-related negativity. *Perspectives on Psychological Science*, *13*(2), 200-204. https://doi.org/10.1177/1745691617715310

Gilhooly, K. J., & Gilhooly, M. L. (1979). Age-of-acquisition effects in lexical and episodic memory tasks. *Memory & Cognition*, *7*(3), 214-223. https://doi.org/10.3758/bf03197541 Actions

Glaser, W. R. (1992). Picture naming. *Cognition*, *42*(1-3), 61105.

https://doi.org/10.1016/0010-0277(92)90040-o

Glaser, W. R., & Düngelhoff, F. J. (1984). The time course of picture-word interference. *Journal of Experimental Psychology: Human Perception and Performance*, *10*(5), 640. https://doi.org/10.1037/0096-1523.10.5.640

Goodglass, H. (1993). *Understanding aphasia*. Academic Press. https://doi.org/10.2307/416147

Gotts, S. J., Chow, C. C., & Martin, A. (2012). Repetition priming and repetition suppression: A case for enhanced efficiency through neural synchronization. *Cognitive neuroscience*, *3*(3-4), 227-237. https://doi.org/10.1080/17588928.2012.670617

Gratton, G., Coles, M. G., & Donchin, E. (1983). A new method for off-line removal of ocular artifact. *Electroencephalography and clinical neurophysiology*, *55*(4), 468-484. https://doi.org/10.1016/0013-4694(83)90135-9

Griffin, Z. M., & Ferreira, V. S. (2006). Properties of spoken language production. In *Handbook of psycholinguistics* (pp. 21-59). Academic Press. https://doi.org/10.1016/b978-012369374-7/50003-1

Gupta, P., & Cohen, N. J. (2002). Theoretical and computational analysis of skill learning, repetition priming, and procedural memory. *Psychological review*, *109*(2), 401. https://doi.org/10.1037/0033-295x.109.2.401

Harley, T. A. (1993). Phonological activation of semantic competitors during lexical access in speech production. *Language and Cognitive Processes*, *8*(3), 291-309. https://doi.org/10.1080/01690969308406957

Harley, T. A., & Grant, F. (2004). The role of functional and perceptual attributes: evidence from picture naming in dementia. *Brain and Language*, *91*(2), 223-234. https://doi.org/10.1016/j.bandl.2004.02.008

Harley, T. A., & MacAndrew, S. B. (1995). Interactive models of lexicalization: Some constraints from speech error, picture naming, and neuropsychological data. *Connectionist models of memory and language*, 311-331. https://doi.org/10.1080/09602019508520174

Harvey, D. Y., & Schnur, T. T. (2016). Different loci of semantic interference in picture naming vs. word-picture matching tasks. *Frontiers in psychology*, *7*, 710. https://doi.org/10.3389/fpsyg.2016.00710

Hirschfeld, G., Jansma, B., Bölte, J., & Zwitserlood, P. (2008). Interference and facilitation in overt speech production investigated with event-related potentials. *Neuroreport*, *19*(12), 1227-1230. https://doi.org/10.1097/wnr.0b013e328309ecd1

Heil, M., Osman, A., Wiegelmann, J., Rolke, B., & Hennighausen, E. (2000). N200 in the Eriksen-task: Inhibitory executive process?. *Journal of Psychophysiology*, *14*(4), 218. https://doi.org/10.1027//0269-8803.14.4.218

Hernandez, A. E., Dapretto, M., Mazziotta, J., & Bookheimer, S. (2001). Language switching and language representation in Spanish–English bilinguals: An fMRI study. *NeuroImage*, *14*(2), 510-520. https://doi.org/10.1006/nimg.2001.0810

Hoffmann, S., & Falkenstein, M. (2008). The correction of eye blink artefacts in the EEG: a comparison of two prominent methods. *PLoS One*, *3*(8), e3004. ttps://doi.org/10.1371/journal.pone.0003004

Holcomb, P. J., Grainger, J., & O'rourke, T. (2002). An electrophysiological study of the effects of orthographic neighborhood size on printed word perception. *Journal of Cognitive Neuroscience*, *14*(6), 938-950. https://doi.org/10.1162/089892902760191153

Holcomb, P. J., Kounios, J., Anderson, J. E., & West, W. C. (1999). Dual-coding, context-availability, and concreteness effects in sentence comprehension: an electrophysiological

investigation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(3), 721. https://doi.org/10.1037/0278-7393.25.3.721

Howard, D., Nickels, L., Coltheart, M., & Cole-Virtue, J. (2006). Cumulative semantic inhibition in picture naming: Experimental and computational studies. *Cognition*, *100*(3), 464–482. https://doi.org/10.1016/j.cognition.2005.02.006

Hubbard, R., & Lindsay, R. M. (2008). Why P values are not a useful measure of evidence in statistical significance testing. *Theory & Psychology*, *18*(1), 69-88. https://doi.org/10.1177/0959354307086923

Humphreys, G. W., Lloyd-Jones, T. J., & Fias, W. (1995). Semantic interference effects on naming using a postcue procedure: Tapping the links between semantics and phonology with pictures and words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(4), 961. https://doi.org/10.1037/0278-7393.21.4.961

Humphreys, G. W., Riddoch, M. J., & Quinlan, P. T. (1988). Cascade processes in picture identification. *Cognitive neuropsychology*, *5*(1), 67-104. https://doi.org/10.1080/02643298808252927

Indefrey, P. (2011). The spatial and temporal signatures of word production components: a critical update. *Frontiers in psychology*, *2*, 255. https://doi.org/10.3389/fpsyg.2011.00255

Indefrey, P., & Levelt, W. J. M. (2004). The spatial and temporal signatures of word production components. *Cognition*, *92*(1–2), 101–144. https://doi.org/10.1016/j.cognition.2002.06.001

Irons, S., Oppenheim, G., & Fischer-Baum, S. (2017). Breaking the Dark Side: A computational neuropsychological approach. Abstract from 55th Annual Meeting of the Academy of Aphasia, Baltimore, United States. https://doi.org/10.3389/conf.fnhum.2017.223.00037

Jackson, G. M., Swainson, R., Mullin, A., Cunnington, R., & Jackson, S. R. (2004). ERP correlates of a receptive language-switching task. *The Quarterly Journal of Experimental Psychology*

Section A, *57*(2), 223-240. https://doi.org/10.1016/s1053-8119(01)91665-9

Janssen, N., Hernández-Cabrera, J. A., van der Meij, M., & Barber, H. A. (2015). Tracking the time course of competition during word production: Evidence for a post-retrieval mechanism of conflict resolution. *Cerebral Cortex*, *25*(9), 2960-2969. https://doi.org/10.1093/cercor/bhu092

Janssen, N., Schirm, W., Mahon, B. Z., & Caramazza, A. (2008). The semantic interference effect in the picture–word interference paradigm: Evidence for the response selection hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 249-256. https://doi.org/10.1037/e512682013-476

Jasper, H. H. (1958). The ten-twenty electrode system of the International Federation. *Electroencephalogr. Clin. Neurophysiol.*, *10*, 370-375. https://doi.org/10.1080/00029238.1961.11080571

Jensen, A. R. (2006). *Clocking the mind: Mental chronometry and individual differences*. Elsevier. https://doi.org/10.1111/j.1744-6570.2008.00111_7.x

Johnson, C. J., Paivio, A., & Clark, J. M. (1996). Cognitive components of picture naming. *Psychological bulletin*, *120*(1), 113. https://doi.org/10.1037/0033-2909.120.1.113

Jung, T. P., Humphries, C., Lee, T. W., Makeig, S., McKeown, M. J., Iragui, V., & Sejnowski, T. J. (1998). Extended ICA removes artifacts from electroencephalographic recordings. In *Advances in neural information processing systems* (pp. 894-900). https://doi.org/10.1111/1469-8986.3720163

Jung, T. P., Makeig, S., Westerfield, M., Townsend, J., Courchesne, E., & Sejnowski, T. J. (2000). Removal of eye activity artifacts from visual event-related potentials in normal and clinical subjects. *Clinical Neurophysiology*, *111*(10), 1745-1758. https://doi.org/10.1016/s1388-2457(00)00386-2 Actions

Kan, I. P., Kable, J. W., Van Scoyoc, A., Chatterjee, A., & Thompson-Schill, S. L. (2006). Fractionating the left frontal response to tools: Dissociable effects of motor experience and lexical competition. *Journal of cognitive neuroscience*, *18*(2), 267-277. https://doi.org/10.1162/jocn.2006.18.2.267

Kan, I. P., & Thompson-Schill, S. L. (2004). Effect of name agreement on prefrontal activity during overt and covert picture naming. *Cognitive, Affective, & Behavioral Neuroscience*, *4*(1), 43-57. https://doi.org/10.3758/cabn.4.1.43

Koester, D., & Schiller, N. O. (2008). Morphological priming in overt language production: Electrophysiological evidence from Dutch. *Neuroimage*, *42*(4), 1622-1630. https://doi.org/10.1016/j.neuroimage.2008.06.043

Kohn, S. E., & Goodglass, H. (1985). Picture-naming in aphasia. *Brain and language*, *24*(2), 266-283. https://doi.org/10.1016/0093-934x(85)90135-x

Kraljic, T., Samuel, A. G., & Brennan, S. E. (2008). First impressions and last resorts: How listeners adjust to speaker variability. *Psychological science*, *19*(4), 332-338. https://doi.org/10.1037/e527342012-194

Kremin, H., Perrier, D., De Wilde, M., Dordain, M., Le Bayon, A., Gatignol, P., ... & Arabia, C. (2001). Factors predicting success in picture naming in Alzheimer's disease and primary progressive aphasia. *Brain and cognition*, *46*(1-2), 180-183. https://doi.org/10.1006/brcg.2000.1270

Kremin, H., Akhutina, T., Basso, A., Davidoff, J., De Wilde, M., Kitzing, P., ... & Weniger, D. (2003). A cross-linguistic data bank for oral picture naming in Dutch, English, German, French, Italian, Russian, Spanish, and Swedish (PEDOI). *Brain and Cognition*, *53*(2), 243-246. https://doi.org/10.1016/s0278-2626(03)00119-2

Kristjánsson, Á., & Campana, G. (2010). Where perception meets memory: A review of repetition priming in visual search tasks. *Attention, Perception, & Psychophysics*, *72*(1), 5-18. https://doi.org/10.3758/app.72.1.5

Ku, Y., Ohara, S., Wang, L., Lenz, F. A., Hsiao, S. S., Bodner, M., ... & Zhou, Y. D. (2007). Prefrontal cortex and somatosensory cortex in tactile crossmodal association: an independent component analysis of ERP recordings. *PloS one*, *2*(8), e771. https://doi.org/10.1371/journal.pone.0000771

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*(4427), 203-205. https://doi.org/10.1126/science.7350657

Lachman, R. (1973). Uncertainty effects on time to access the internal lexicon. *Journal of Experimental Psychology*, *99*(2), 199. https://doi.org/10.1037/h0034633

Lachman, R., & Lachman, J. L. (1980). Picture naming: Retrieval and activation of long-term memory. In *New Directions in Memory and Aging (PLE: Memory): Proceedings of the George A. Talland Memorial Conference* (p. 313). https://doi.org/10.4324/9781315774886

Lachman, R., Lachman, J. L., Thronesbery, C., & Sala, L. S. (1980). Object salience and code separation in picture naming. *Bulletin of the Psychonomic Society*, *16*(3), 187-190. ttps://doi.org/10.3758/bf03329517

Lachman, R., Shaffer, J. P., & Hennrikus, D. (1974). Language and cognition: Effects of stimulus codability, name-word frequency, and age of acquisition on lexical reaction time. *Journal of verbal learning and verbal behavior*, *13*(6), 613-625. https://doi.org/10.1016/s0022-5371(74)80049-6

LaGrone, S., & Spieler, D. H. (2006). Lexical competition and phonological encoding in young and older speakers. *Psychology and Aging*, *21*(4), 804. https://doi.org/10.1037/0882-7974.21.4.804

La Heij, W. (1988). Components of Stroop-like interference in picture naming. *Memory & Cognition*, *16*(5), 400-410. https://doi.org/10.3758/bf03214220

Laiacona, M., Luzzatti, C., Zonca, G., Guarnaschelli, C., & Capitani, E. (2001). Lexical and semantic factors influencing picture naming in aphasia. *Brain and Cognition*, *46*(1-2), 184-187. https://doi.org/10.1016/s0278-2626(01)80061-0

Laganaro, M., & Perret, C. (2011). Comparing electrophysiological correlates of word production in immediate and delayed naming through the analysis of word age of acquisition effects. *Brain Topography*, *24*(1), 19-29. https://doi.org/10.1007/s10548-010-0162-x

Laganaro, M., Python, G., & Toepel, U. (2013). Dynamics of phonological–phonetic encoding in word production: Evidence from diverging ERPs between stroke patients and controls. *Brain and Language*, *126*(2), 123-132. https://doi.org/10.1016/j.bandl.2013.03.004

Laganaro, M., Valente, A., & Perret, C. (2012). Time course of word production in fast and slow speakers: a high density ERP topographic study. *NeuroImage*, *59*(4), 3881-3888. https://doi.org/10.1016/j.neuroimage.2011.10.082

Laiacona, M., & Capitani, E. (2001). A case of prevailing deficit of nonliving categories or a case of prevailing sparing of living categories?. *Cognitive Neuropsychology*, *18*(1), 39-70. https://doi.org/10.1080/02643290042000035

Lei, X., & Liao, K. (2017). Understanding the influences of EEG reference: a large-scale brain network perspective. *Frontiers in neuroscience*, *11*, 205. https://doi.org/10.3389/fnins.2017.00205

Leske, S., & Dalal, S. S. (2019). Reducing power line noise in EEG and MEG data via spectrum interpolation. *NeuroImage*, *189*, 763-776. https://doi.org/10.1016/j.neuroimage.2019.01.026

Levelt, W. J. (1983). Monitoring and self-repair in speech. *Cognition*, *14*(1), 41-104. https://doi.org/10.1016/0010-0277(83)90026-4

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge: MIT Press.

Levelt, W. J., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and brain sciences*, *22*(1), 1-38. https://doi.org/10.1017/s0140525x99001776

Levelt, W. J., Schriefers, H., Vorberg, D., Meyer, A. S., Pechmann, T., & Havinga, J. (1991). The time course of lexical access in speech production: A study of picture naming. *Psychological review*, *98*(1), 122. https://doi.org/10.1037/0033-295x.98.1.122

Levinson, S. C. (1997). *From outer to inner space: linguistic categories and non-linguistic thinking* (pp. 13-45). Cambridge University Press. https://doi.org/10.1017/cbo9781139086677.002

Liljeström, M., Hultén, A., Parkkonen, L., & Salmelin, R. (2009). Comparing MEG and fMRI views to naming actions and objects. *Human brain mapping*, *30*(6), 1845-1856. https://doi.org/10.1016/s1053-8119(09)71399-0

Llorens, A., Trébuchon, A., Riès, S., Liégeois-Chauvel, C., & Alario, F. X. (2014). How familiarization and repetition modulate the picture naming network. *Brain and language*, *133*, 47-58. https://doi.org/10.1016/j.bandl.2014.03.010

Lorsbach, T. C., & Morris, A. K. (1991). Direct and indirect testing of picture memory in second and sixth grade children. *Contemporary Educational Psychology*, *16*(1), 18-27. https://doi.org/10.1016/0361-476x(91)90003-4

Luce, R. D. (1959). Individual Choice Behavior: A theoretical analysis, New York, NY: John Willey and Sons.

Luck, S. J. (2005). Ten simple rules for designing ERP experiments. *Event-related potentials: A methods handbook*, *262083337*.

Luck, S. J. (2014). *An introduction to the event-related potential technique*. MIT press.

Luck, S. J., Hillyard, S. A., Mouloua, M., Woldorff, M. G., Clark, V. P., & Hawkins, H. L. (1994). Effects of spatial cuing on luminance detectability: psychophysical and electrophysiological

evidence for early selection. *Journal of experimental psychology: human perception and performance*, *20*(4), 887. https://doi.org/10.1037/0096-1523.20.4.887

Makeig, S., Bell, A. J., Jung, T. P., & Sejnowski, T. J. (1996). Independent component analysis of electroencephalographic data. In *Advances in neural information processing systems* (pp. 145-151). https://doi.org/10.1007/978-1-4615-5351-9_17

Makeig, S., Westerfield, M., Jung, T. P., Covington, J., Townsend, J., Sejnowski, T. J., & Courchesne, E. (1999). Functionally independent components of the late positive event-related potential during visual spatial attention. *Journal of Neuroscience*, *19*(7), 2665-2680. https://doi.org/10.1523/jneurosci.19-07-02665.1999

Mandler, G., Goodman, G. O., & Wilkes-Gibbs, D. L. (1982). The word-frequency paradox in recognition. *Memory & Cognition*, *10*(1), 33-42. https://doi.org/10.3758/bf03197623

Mahon, B. Z., Costa, A., Peterson, R., Vargas, K. A., & Caramazza, A. (2007). Lexical selection is not by competition: a reinterpretation of semantic interference and facilitation effects in the picture-word interference paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(3), 503. https://doi.org/10.1037/0278-7393.33.3.503

Martin, N. (1996). Phonological facilitation of semantic errors in normal and aphasic speakers. *Language and Cognitive Processes*, *11*(3), 257-282. https://doi.org/10.1080/016909696387187

McCulloch, C. E., Searle, S. R., & Neuhaus, J. M. (2008). Generalized, linear, and mixed models. Hoboken. *NJ: Wiley. QA*, *279*, M38.

Mehl, M. R., Vazire, S., Ramírez-Esparza, N., Slatcher, R. B., & Pennebaker, J. W. (2007). Are women really more talkative than men?. *Science*, *317*(5834), 82-82. https://doi.org/10.1126/science.1139940

Meier, E. L., Kapse, K. J., & Kiran, S. (2016). The relationship between frontotemporal effective connectivity during picture naming, behavior, and preserved cortical tissue in chronic aphasia. *Frontiers in human neuroscience*, *10*, 109. https://doi.org/10.3389/fnhum.2016.00109

Meyer, A. S. (1992). Investigation of phonological encoding through speech error analyses: Achievements, limitations, and alternatives. *Cognition*, *42*, 181-211. https://doi.org/10.1016/0010-0277(92)90043-h

Mitchell, D. B. (1989). How many memory systems? Evidence from aging. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(1), 31. https://doi.org/10.1037/0278-7393.15.1.31

Mitchell, D. B., & Brown, A. S. (1988). Persistent repetition priming in picture naming and its dissociation from recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(2), 213. https://doi.org/10.1037/0278-7393.14.2.213

Mitchell, R. C., Carson, R. T., & Carson, R. T. (1989). *Using surveys to value public goods: the contingent valuation method*. Resources for the Future.

Morsella, E., & Miozzo, M. (2002). Evidence for a cascade model of lexical access in speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(3), 555. https://doi.org/10.1037/0278-7393.28.3.555

Munding, D., Dubarry, A. S., & Alario, F. X. (2016). On the cortical dynamics of word production: A review of the MEG evidence. *Language, Cognition and Neuroscience*, *31*(4), 441-462. https://doi.org/10.1080/23273798.2015.1071857

Murtha, S., Chertkow, H., Beauregard, M., & Evans, A. (1999). The neural substrate of picture naming. *Journal of cognitive neuroscience*, *11*(4), 399-423. https://doi.org/10.1162/089892999563508

Nation, K., Marshall, C. M., & Snowling, M. J. (2001). Phonological and semantic contributions to children's picture naming skill: Evidence from children with developmental reading disorders. *Language and Cognitive Processes*, *16*(2-3), 241-259. https://doi.org/10.1080/01690960042000003

Navarrete, E., Del Prato, P., Peressotti, F., & Mahon, B. Z. (2014). Lexical selection is not by competition: Evidence from the blocked naming paradigm. *Journal of Memory and Language*, *76*, 253-272. https://doi.org/10.1016/j.jml.2014.05.003

Niedermeyer, E., & da Silva, F. L. (Eds.). (2005). *Electroencephalography: basic principles, clinical applications, and related fields*. Lippincott Williams & Wilkins.

Nieuwenhuis, S., Yeung, N., Van Den Wildenberg, W., & Ridderinkhof, K. R. (2003). Electrophysiological correlates of anterior cingulate function in a go/no-go task: effects of response conflict and trial type frequency. *Cognitive, affective, & behavioral neuroscience*, *3*(1), 17-26. https://doi.org/10.3758/cabn.3.1.17

Nilipour, R., Bakhtiar, M., Momenian, M., & Weekes, B. S. (2017). Object and action picture naming in brain-damaged Persian speakers with aphasia. *Aphasiology*, *31*(4), 388-405. https://doi.org/10.1080/02687038.2016.1234583

Nilipour, R., Pourshahbaz, A., & Momenian, M. (2017). The effect of psycholinguistic factor on picture naming in Persian speaking children. *International Journal of Children and Adolescents*, *3*(4), 13-21.

Nishimoto, T., Ueda, T., Miyawaki, K., Une, Y., & Takahashi, M. (2012). The role of imagery-related properties in picture naming: A newly standardized set of 360 pictures for Japanese. *Behavior research methods*, *44*(4), 934-945. https://doi.org/10.3758/s13428-011-0176-7

Novick, J. M., Kan, I. P., Trueswell, J. C., & Thompson-Schill, S. L. (2009). A case for conflict across multiple domains: Memory and language impairments following damage to ventrolateral

prefrontal cortex. *Cognitive neuropsychology*, *26*(6), 527-567. https://doi.org/10.1080/02643290903519367

Nozari, N., Dell, G. S., & Schwartz, M. F. (2011). Is comprehension necessary for error detection? A conflict-based account of monitoring in speech production. *Cognitive psychology*, *63*(1), 1-33. https://doi.org/10.1016/j.cogpsych.2011.05.001

Nozari, N., Freund, M., Breining, B., Rapp, B., & Gordon, B. (2016). Cognitive control during selection and repair in word production. *Language, cognition and neuroscience*, *31*(7), 886-903. https://doi.org/10.1080/23273798.2016.1157194

Nozari, N., & Hepner, C. R. (2019). To select or to wait? The importance of criterion setting in debates of competitive lexical selection. *Cognitive neuropsychology*, *36*(5-6), 193-207. https://doi.org/10.1080/02643294.2018.1476335

Nozari, N., & Pinet, S. (2020). A critical review of the behavioral, neuroimaging, and electrophysiological studies of co-activation of representations during word production. *Journal of Neurolinguistics*, *53*, 100875. https://doi.org/10.1016/j.jneuroling.2019.100875

Oldfield, R. C., & Wingfield, A. (1965). Response latencies in naming objects. *Quarterly Journal of Experimental Psychology*, *17*(4), 273-281. https://doi.org/10.1080/17470216508416445

Osterhout, L., & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of memory and language*, *31*(6), 785-806. https://doi.org/10.1016/0749-596x(92)90039-z

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251). https://doi.org/10.1126/science.aac4716

Oppenheim, G. M. (2017, September). *Strong competitors facilitate target name retrieval in simple picture naming*. Poster session presented at Architectures and Mechanisms of Language Processing 2017, Lancaster, UK.

Oppenheim, G. M. (2018). The paca that roared: Immediate cumulative semantic interference among newly acquired words. *Cognition*, *177*, 21-29. https://doi.org/10.1016/j.cognition.2018.02.014

Oppenheim, G. M. (in preparation). Competition in normal production.

Oppenheim, G. M., & Balatsou, E. (2019). Lexical competition on demand. *Cognitive neuropsychology*, *36*(5-6), 216-219. https://doi.org/10.1080/02643294.2019.1580189

Oppenheim, G. M., Barr, P., & Tainturier, M. J. (2016). Preserved cumulative semantic interference despite explicit memory impairment. In *Proceedings of the 2016 international meeting of the psychonomic society*. https://doi.org/10.3389/conf.fpsyg.2015.65.00002

Oppenheim, G. M., Dell, G. S., & Schwartz, M. F. (2007). Cumulative semantic interference as learning. *Brain and Language*, *103*(1), 175-176. https://doi.org/10.1016/j.bandl.2007.07.102

Oppenheim GM, Tainturier M and Barr P (2015). Preserved cumulative semantic interference despite amnesia. *Front. Psychol. Conference Abstract: Academy of Aphasia 53rd Annual Meeting.* doi: 10.3389/conf.fpsyg.2015.65.00002

Paivio, A., Clark, J. M., Digdon, N., & Bons, T. (1989). Referential processing: Reciprocity and correlates of naming and imaging. *Memory & Cognition*, *17*(2), 163-174. https://doi.org/10.3758/bf03197066

Park, S. M., & Gabrieli, J. D. (1995). Perceptual and nonperceptual components of implicit memory for pictures. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(6), 1583. https://doi.org/10.1037/0278-7393.21.6.1583

Pashler, H., & Harris, C. R. (2012). Is the replicability crisis overblown? Three arguments examined. *Perspectives on Psychological Science*, *7*(6), 531-536. https://doi.org/10.1177/1745691612463401

Pernet, C. R., Chauveau, N., Gaspar, C., & Rousselet, G. A. (2011). LIMO EEG: a toolbox for hierarchical LInear MOdeling of ElectroEncephaloGraphic data. *Computational intelligence and neuroscience*, *2011*. https://doi.org/10.1155/2011/831409

Peterson, R. R., & Savoy, P. (1998). Lexical selection and phonological encoding during language production: Evidence for cascaded processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*(3), 539. https://doi.org/10.1037/0278-7393.24.3.539

Piai, V., & Roelofs, A. (2013). Working memory capacity and dual-task interference in picture naming. *Acta psychologica*, *142*(3), 332-342. https://doi.org/10.1016/j.actpsy.2013.01.006

Piai, V., Roelofs, A., & van der Meij, R. (2012). Event-related potentials and oscillatory brain responses associated with semantic and Stroop-like interference effects in overt naming. *Brain Research*, 1450, 87– https://doi.org/10.1016/j.brainres.2012.02.050

Princeton University "About WordNet." WordNet. Princeton University. 2010.

Prior, A., MacWhinney, B., & Kroll, J. F. (2007). Translation norms for English and Spanish: The role of lexical variables, word class, and L2 proficiency in negotiating translation ambiguity. *Behavior Research Methods*, *39*(4), 1029-1038. https://doi.org/10.3758/bf03193001

Python, G., Fargier, R., & Laganaro, M. (2018). ERP evidence of distinct processes underlying semantic facilitation and interference in word production. *Cortex*, 99, 1-12. https://doi.org/10.1016/j.cortex.2017.09.008

RStudio Team (2020). RStudio: Integrated Development for R. RStudio, PBC, Boston, MA URL http://www.rstudio.com/.

Rabovsky, M., Sommer, W., & Abdel Rahman, R. (2012). Implicit word learning benefits from semantic richness: Electrophysiological and behavioral evidence. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(4), 1076. https://doi.org/10.1037/a0025646

Rapp, B., & Goldrick, M. (2000). Discreteness and interactivity in spoken word production. *Psychological review*, *107*(3), 460. https://doi.org/10.1037/0033-295x.107.3.460

Rodd, J. M., Cutrin, B. L., Kirsch, H., Millar, A., & Davis, M. H. (2013). Long-term priming of the meanings of ambiguous words. *Journal of Memory and Language*, *68*(2), 180-198. https://doi.org/10.1016/j.jml.2012.08.002

Rodríguez-Ferreiro, J., Menéndez, M., Ribacoba, R., & Cuetos, F. (2009). Action naming is impaired in Parkinson disease patients. *Neuropsychologia*, 47(14), 3271-3274. https://doi.org/10.1016/j.neuropsychologia.2009.07.007

Rodriguez-Fornells, A., De Diego Balaguer, R., & Münte, T. F. (2006). Executive control in bilingual language processing. *Language Learning*, *56*, 133-190. https://doi.org/10.1111/j.1467-9922.2006.00359.x

Rodriguez-Fornells, A., Lugt, A. V. D., Rotte, M., Britti, B., Heinze, H. J., & Münte, T. F. (2005). Second language interferes with word production in fluent bilinguals: brain potential and functional imaging evidence. *Journal of cognitive neuroscience*, *17*(3), 422-433. https://doi.org/10.1162/0898929053279559

Rodriguez-Fornells, A., Rotte, M., Heinze, H. J., Nösselt, T., & Münte, T. F. (2002). Brain potential and functional MRI evidence for how to handle two languages with one brain. *Nature*, *415*(6875), 1026-1029. https://doi.org/10.1038/4151026a

Roelofs, A. (1992). A spreading-activation theory of lemma retrieval in speaking. *Cognition*, *42*(1-3), 107-142. https://doi.org/10.1016/0010-0277(92)90041-f

Roelofs, A. (1993). Testing a non-decompositional theory of lemma retrieval in speaking: Retrieval of verbs. *Cognition*, *47*(1), 59-87. https://doi.org/10.1016/0010-0277(93)90062-z

Roelofs, A. (1997). The WEAVER model of word-form encoding in speech production. *Cognition*, *64*(3), 249-284. ttps://doi.org/10.1016/s0010-0277(97)00027-9

Roelofs, A. (2001). Set size and repetition matter: Comment on Caramazza and Costa (2000). *Cognition*, *80*(3), 283-290. https://doi.org/10.1016/s0010-0277(01)00134-2

Roelofs, A. (2003). Goal-referenced selection of verbal action: modeling attentional control in the Stroop task. *Psychological review*, *110*(1), 88. https://doi.org/10.1037/0033-295x.110.1.88

Roelofs, A. (2004). Comprehension-based versus production-internal feedback in planning spoken words: a Rejoinder to Rapp and Goldrick (2004). ttps://doi.org/10.1037/0033-295x.111.2.579

Roelofs, A. (2018). A unified computational account of cumulative semantic, semantic blocking, and semantic distractor effects in picture naming. *Cognition*, *172*, 59-72. https://doi.org/10.1016/j.cognition.2017.12.007

Roelofs, A. P. A., & Piai, V. (2015). Aspects of competition in word production: Reply to Mahon and Navarrete. https://doi.org/10.1016/j.cortex.2014.10.016

Roelofs, A., & Piai, V. (2017). Distributional analysis of semantic interference in picture naming. *Quarterly Journal of Experimental Psychology*, *70*(4), 782-792. https://doi.org/10.1080/17470218.2016.1165264

Rose, S. B., & Abdel Rahman, R. (2017). Semantic similarity promotes interference in the continuous naming paradigm: behavioral and electrophysiological evidence. *Language, Cognition and Neuroscience*, *32*(1), 55-68. https://doi.org/10.1080/23273798.2016.1212081

Rose, S. B., Aristei, S., Melinger, A., & Abdel Rahman, R. (2019). The closer they are, the more they interfere: Semantic similarity of word distractors increases competition in language production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45*(4), 753. https://doi.org/10.1037/xlm0000592

Rousselet, G. A., Pernet, C. R., Caldara, R., & Schyns, P. G. (2011). Visual object categorization in the brain: what can we really learn from ERP peaks?. *Frontiers in Human Neuroscience*, *5*, 156. https://doi.org/10.3389/fnhum.2011.00156

Saccuman, M. C., Cappa, S. F., Bates, E. A., Arevalo, A., Della Rosa, P., Danna, M., & Perani, D. (2006). The impact of semantic reference on word class: An fMRI study of action and object naming. *Neuroimage*, *32*(4), 1865-1878. https://doi.org/10.1016/j.neuroimage.2006.04.179

Sanfeliu, M. C., & Fernandez, A. (1996). A set of 254 Snodgrass-Vanderwart pictures standardized for Spanish: Norms for name agreement, image agreement, familiarity, and visual complexity. *Behavior Research Methods, Instruments, & Computers*, *28*(4), 537-555. https://doi.org/10.3758/bf03200541

Sanoudaki, E., & Thierry, G. (2015). Language non-selective syntactic activation in early bilinguals: the role of verbal fluency. *International Journal of Bilingual Education and Bilingualism*, *18*(5), 548-560. https://doi.org/10.1080/13670050.2015.1027143

Scaltritti, M., Navarrete, E., & Peressotti, F. (2015). Distributional analyses in the picture–word interference paradigm: Exploring the semantic interference and the distractor frequency effects. *Quarterly Journal of Experimental Psychology*, *68*(7), 1348-1369. https://doi.org/10.1080/17470218.2014.981196

Sharbrough, F. (1991). American Electroencephalographic Society guidelines for standard electrode position nomenclature. *J clin Neurophysiol*, *8*, 200-202. https://doi.org/10.1097/00004691-199104000-00007

Schmitt, B. M., Münte, T. F., & Kutas, M. (2000). Electrophysiological estimates of the time course of semantic and phonological encoding during implicit picture naming. *Psychophysiology*, *37*(4), 473-484. https://doi.org/10.1111/1469-8986.3740473

Schnur, T. T. (2014). The persistence of cumulative semantic interference during naming. *Journal of Memory and Language*, *75*, 27 44. https://doi.org/10.1016/j.jml.2014.04.006

Schnur, T. T., Schwartz, M. F., Brecher, A., & Hodgson, C. (2006). Semantic interference during blocked-cyclic naming: Evidence from aphasia. *Journal of Memory and Language*, *54*(2), 199-227. https://doi.org/10.1016/j.jml.2005.10.002

Schriefers, H., Meyer, A. S., & Levelt, W. J. (1990). Exploring the time course of lexical access in language production: Picture-word interference studies. *Journal of memory and language*, *29*(1), 86-102. https://doi.org/10.1016/0749-596x(90)90011-n

Shao, Z., Meyer, A. S., & Roelofs, A. (2013). Selective and nonselective inhibition of competitors in picture naming. *Memory & cognition*, *41*(8), 1200-1211. https://doi.org/10.3758/s13421-013-0332-7

Shao, Z., Roelofs, A., Martin, R. C., & Meyer, A. S. (2015). Selective inhibition and naming performance in semantic blocking, picture-word interference, and color–word Stroop tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(6), 1806. https://doi.org/10.1037/a0039363

Shao, Z., Roelofs, A., Acheson, D. J., & Meyer, A. S. (2014). Electrophysiological evidence that inhibition supports lexical selection in picture naming. *brain research*, *1586*, 130-142. https://doi.org/10.1016/j.brainres.2014.07.009

Shitova, N., Roelofs, A., Schriefers, H., Bastiaansen, M., & Schoffelen, J.-M. (2017). Control adjustments in speaking: Electrophysiology of the Gratton effect in picture naming. *Cortex*, 92, 289–303. https://doi.org/10.1016/j.cortex.2017.04.017

Smith, M. E., & Guster, K. (1993). Decomposition of recognition memory event-related potentials yields target, repetition, and retrieval effects. *Electroencephalography and Clinical Neurophysiology*, *86*(5), 335-343. https://doi.org/10.1016/0013-4694(93)90046-x

Spitzer, M., Kwong, K. K., Kennedy, W., Rosen, B. R., & Belliveau, J. W. (1995). Category-specific brain activation in fMRI during picture naming. *Neuroreport: An International Journal for the*

Rapid Communication of Research in Neuroscience. https://doi.org/10.1097/00001756-199511000-00003

Snodgrass, J. G., & Yuditsky, T. (1996). Naming times for the Snodgrass and Vanderwart pictures. *Behavior Research Methods*, 28(4), 516-536. https://doi.org/10.3758/bf03200540

Schendan, H. E., & Maher, S. M. (2009). Object knowledge during entry-level categorization is activated and modified by implicit memory after 200 ms. *Neuroimage*, *44*(4), 1423-1438. https://doi.org/10.1016/j.neuroimage.2008.09.061

Schendan, H. E., & Kutas, M. (2007). Neurophysiological evidence for the time course of activation of global shape, part, and local contour representations during visual object categorization and memory. *Journal of Cognitive Neuroscience*, 19(5), 734-749. https://doi.org/10.1162/jocn.2007.19.5.734

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of experimental psychology: Human learning and memory*, *6*(2), 174. https://doi.org/10.1037/0278-7393.6.2.174

Schneider, R., Yurovsky, D., & Frank, M. (2015, July). Large-scale investigations of variability in children's first words. In *CogSci* (pp. 2110-2115). https://doi.org/10.31234/osf.io/cg6ah

Schriefers, H., Meyer, A. S., & Levelt, W. J. (1990). Exploring the time course of lexical access in language production: Picture-word interference studies. *Journal of memory and language*, 29(1), 86-102. https://doi.org/10.1016/0749-596x(90)90011-n Actions

Spalek, K., Damian, M. F., & Bölte, J. (2013). Is lexical selection in spoken word production competitive? Introduction to the special issue on lexical competition in language

production. *Language and Cognitive Processes*, *28*(5), 597-614. https://doi.org/10.1080/01690965.2012.718088

Starreveld, P. A., & La Heij, W. (1995). Semantic interference, orthographic facilitation, and their interaction in naming tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(3), 686. https://doi.org/10.1037/0278-7393.21.3.686

Stemberger, J. P. (1985). An interactive activation model of language production. *Progress in the psychology of language*.

Strijkers, K., & Costa, A. (2011). Riding the lexical speedway: A critical review on the time course of lexical selection in speech production. *Frontiers in psychology*, *2*, 356. https://doi.org/10.3389/fpsyg.2011.00356

Strijkers, K., Costa, A., & Thierry, G. (2010). Tracking lexical access in speech production: electrophysiological correlates of word frequency and cognate effects. *Cerebral cortex*, 20(4), 912-928. https://doi.org/10.1093/cercor/bhp153

Sutton, S., Braren, M., Zubin, J., & John, E. R. (1965). Evoked-potential correlates of stimulus uncertainty. *Science*, *150*(3700), 1187-1188. https://doi.org/10.1126/science.150.3700.1187

Székely, A., Jacobsen, T., D'Amico, S., Devescovi, A., Andonova, E., Herron, D., ... & Federmeier, K. (2004). A new on-line resource for psycholinguistic studies. *Journal of memory and language*, *51*(2), 247-250. https://doi.org/10.1016/j.jml.2004.03.002

Székely, A., D'amico, S., Devescovi, A., Federmeier, K., Herron, D., Iyer, G., ... & Bates, E. (2003). Timed picture naming: Extended norms and validation against previous studies. *Behavior Research Methods, Instruments, & Computers*, *35*(4), 621-633. https://doi.org/10.3758/bf03195542

Tatum, W. O. (2014). Ellen r. grass lecture: Extraordinary eeg. *The Neurodiagnostic Journal*, *54*(1), 3-21.

Thompson-Schill, S. L., D'Esposito, M., Aguirre, G. K., & Farah, M. J. (1997). Role of left inferior prefrontal cortex in retrieval of semantic knowledge: a reevaluation. *Proceedings of the National Academy of Sciences*, *94*(26), 14792-14797. https://doi.org/10.1037/e536982012-664

Torrance, M., Nottbusch, G., Alves, R. A., Arfé, B., Chanquoy, L., Chukharev-Hudilainen, E., ... & Madjarov, G. (2018). Timed written picture naming in 14 European languages. *Behavior Research Methods*, *50*(2), 744-758. https://doi.org/10.3758/s13428-017-0902-x

Van Heuven, W. J., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). SUBTLEX-UK: A new and improved word frequency database for British English. *Quarterly journal of experimental psychology*, *67*(6), 1176-1190. https://doi.org/10.1080/17470218.2013.850521

Van Maanen, L., van Rijn, H., & Borst, J. P. (2009). Stroop and picture—word interference are two sides of the same coin. *Psychonomic bulletin & review*, *16*(6), 987-999. https://doi.org/10.3758/pbr.16.6.987

Van Scherpenberg, C., Fieder, N., Savage, S., & Nickels, L. (2019). The relationship between response consistency in picture naming and storage impairment in people with semantic variant primary progressive aphasia. *Neuropsychology*, *33*(1), 13. https://doi.org/10.1037/neu0000485

Valente, A., & Laganaro, M. (2015). Ageing effects on word production processes: an ERP topographic analysis. *Language, Cognition and Neuroscience*, *30*(10), 1259-1272. https://doi.org/10.1080/23273798.2015.1059950

Valente, A., Bürki, A., & Laganaro, M. (2014). ERP correlates of word production predictors in picture naming: a trial by trial multiple regression analysis from stimulus onset to response. *Frontiers in neuroscience*, *8*, 390. https://doi.org/10.3389/fnins.2014.00390

Van Heuven, W. J., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). SUBTLEX-UK: A new and improved word frequency database for British English. *Quarterly journal of experimental psychology*, *67*(6), 1176-1190. https://doi.org/10.1080/17470218.2013.850521

van Scherpenberg, C., Fieder, N., Savage, S., & Nickels, L. (2019). The relationship between response consistency in picture naming and storage impairment in people with semantic variant primary progressive aphasia. *Neuropsychology*, *33*(1), 13. https://doi.org/10.1037/neu0000485

Van Turennout, M. (1998). Brain activity during speaking: From syntax to phonology in 40 milliseconds. *Science*, *280*(5363),572-574. https://doi.org/10.1126/science.280.5363.572

Van Turennout, M., Hagoort, P., & Brown, C. M. (1997). Electrophysiological evidence on the time course of semantic and phonological processes in speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*(4), 787. https://doi.org/10.1037/0278-7393.23.4.787

Van Turennout, M., Hagoort, P., & Brown, C. (1999). The time course of grammatical and phonological processing during speaking: Evidence from event-related brain potentials. *Journal of Psycholinguistic research*, *28*(6), 649-676. https://doi.org/10.1016/s0167-8760(97)85574-2

Verhoef, K., Roelofs, A., & Chwilla, D. J. (2009). Role of inhibition in language switching: Evidence from event-related brain potentials in overt picture naming. *Cognition*, *110*(1), 84-99. https://doi.org/10.1016/j.cognition.2008.10.013

Vigário, R. N. (1997). Extraction of ocular artefacts from EEG using independent component analysis. *Electroencephalography and clinical neurophysiology*, *103*(3), 395-404. https://doi.org/10.1016/s0013-4694(97)00042-8

Vigliocco, G., Vinson, D. P., Damian, M. F., & Levelt, W. (2002). Semantic distance effects on object and action naming. *Cognition*, *85*(3), B61-B69. https://doi.org/10.1016/s0010-0277(02)00107-5

Vitkovitch, M., & Tyrrell, L. (1995). Sources of disagreement in object naming. *The Quarterly Journal of Experimental Psychology*, *48*(4), 822-848. https://doi.org/10.1080/14640749508401419

Vromans, R. D., & Jongman, S. R. (2018). The interplay between selective and nonselective inhibition during single word production. *PloS one*, *13*(5), e0197313.https://doi.org/10.1371/journal.pone.0197313

Weekes, B. S., Shu, H., Hao, M., Liu, Y., & Tan, L. H. (2007). Predictors of timed picture naming in Chinese. *Behavior Research Methods*, *39*(2), 335-342. https://doi.org/10.3758/bf03193165

Whelan, R. (2008). Effective analysis of reaction time data. *The Psychological Record*, *58*(3), 475-482. https://doi.org/10.1007/BF03395630

Wheeldon, L. R., & Monsell, S. (1994). Inhibition of spoken word production by priming a semantic competitor. *Journal of memory and language*, *33*(3), 332-356. https://doi.org/10.1006/jmla.1994.1016

Wicha, N. Y. Y., Bates, E., Moreno, E., & Kutas, M. (2000). Grammatical gender modulates semantic integration of a picture in a Spanish sentence. *Journal of Cognitive Neuroscience, Supplement*, *1*, 126. ttps://doi.org/10.1016/s0304-3940(03)00599-8

Wiggs, C. L., & Martin, A. (1998). Properties and mechanisms of perceptual priming. *Current opinion in neurobiology*, *8*(2), 227-233. https://doi.org/10.1016/s0959-4388(98)80144-x

Wixted, J. T., & Ebbesen, E. B. (1991). On the form of forgetting. *Psychological science*, *2*(6), 409-415. https://doi.org/10.1111/j.1467-9280.1991.tb00175.x

Wold, S., Esbensen, K., & Geladi, P. (1987). Principal component analysis. *Chemometrics and intelligent laboratory systems*, *2*(1-3), 37-52. https://doi.org/10.1016/0169-7439(87)80084-9

Wong, A. W.-K., Chiu, H.-C., Wang, J., Cao, J., Wong, S.-S., & Chen, H.-C. (2017). An early locus of associative and categorical context effects in speech production: Evidence from an ERP study using the picture–word interference paradigm. *Language, Cognition and Neuroscience*, 32(10), 1305–1319. https://doi.org/10.1080/23273798.2017.1355060

Yuan, J., Liberman, M., & Cieri, C. (2006). Towards an integrated understanding of speaking rate in conversation. In *Ninth International Conference on Spoken Language Processing*. https://doi.org/10.1121/1.4777950