The psychological reality of picture name agreement

Evangelia Balatsou* Simon Fischer-Baum† Gary M. Oppenheim‡

Draft: April, 2019

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

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 and Tyrell, 1995; Alario et al., 2004). *Name*

*School of Psychology, Bangor University

†Department of Psychology, Rice University

‡School of Psychology, Bangor University, Bangor, Gwynedd, LL57 2AS, UK, g.m.oppenheim@bangor.ac.uk

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.

1.1 Name agreement as a predictor of individual-level cognitive processes

But, 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 and 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, Duñabeitia, Blitsas, & Carreiras, 2009), Japanese (Nishimoto, Ueda, Miyawaki, Une, & Takahashi, 2012) and Persian (Bakhtiar, Nilipour, & Weekes, 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, Shaffer, & Hennrikus, 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 behaviour 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

[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)

to directly reflect the processes that occur within each 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 behaviour.

1.3 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 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.

2. Methods

2.1 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).

2.2 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.

2.3 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.,

[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.

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 centre of the screen for 50 ms. Next, a picture (422 x 422 pixels) appeared at the centre 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.

2.4 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 analysed 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. Results

Excluding 899 trials (3.4%) in which a voicekey error ended the trial early (< 300ms post stimulus onset) leaves 25,351 total picture naming attempts for our analyses (12,644 in the first session and 12,707 in the second session), summarised in Table 1.

3.1 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 considered correspon-

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	-	-	-	-	

dence 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.

3.2 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.

As described in the Method 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 behaviour. 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 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

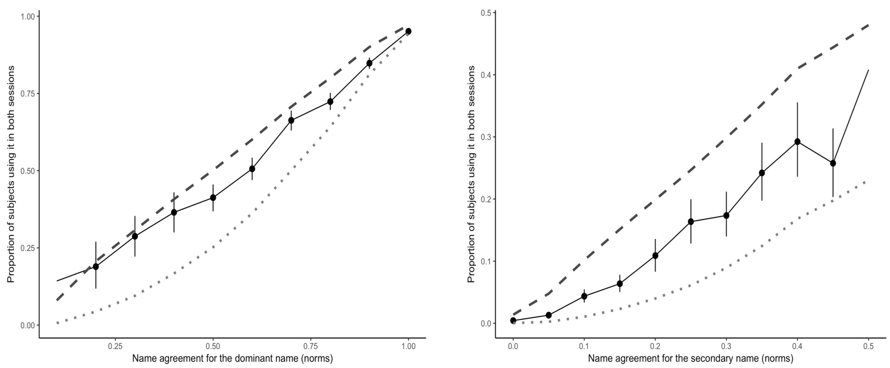

Figure 1: Logistic mixed effects regressions modeling the likelihood of participants using the same dominant name 1b) and secondary name (1b) in both sessions. The y axis represents the proportion of subjects who used the dominant (1a) and secondary name (1b) twice and the x axis represents the population-level name agreement for the dominant (1a) and secondary names (1b) from Oppenheim's (in prep.) Bangor IPNP norms.

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 focussed 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.

3.3 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 or 84.6%.

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. 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.

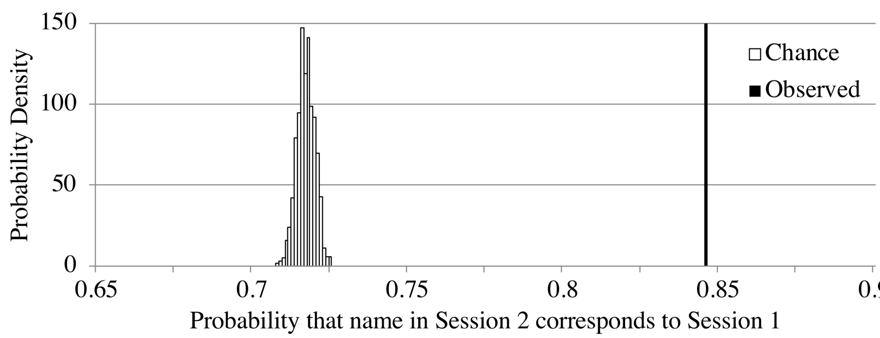

Figure 2: Results of the Monte Carlo analysis comparing the observed proportion of trials in which the name in Session 2 corresponded to name produced in Session 1 (black bar), compared to a distribution of the results of the Monte Carlo analysis that estimated the proportion expected by chance (white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

4. Discussion

Picture name agreement has long been associated with robust behavioural, 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 behaviour 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.

4.1 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 behaviour 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 behaviour 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, 2018), 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).

4.2 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 behaviour 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.

5. 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.

References

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.

Bakhtiar, M., Nilipour, R., & Weekes, B. S. (2013). Predictors of timed picture naming in Persian. *Behavior Research Methods*, *45*(3), 834-841.

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.

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.

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.

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*.

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.

Belke, E., Brysbaert, M., Meyer, A. S., & Ghyselinck, M. (2005). Age of acquisition effects in picture naming: evidence for a lexical-semantic competition hypothesis. *Cognition*, *96*(2), B45-B54.

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.

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.

Bose, A., & Schafer, G. (2017). Name agreement in aphasia. *Aphasiology*, *31*(10), 1143-1165.

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*.

Butterfield, G. B., & Butterfield, E. C. (1977). Lexical codability and age. J*ournal of Memory and Language*, *16*(1), 113.

Cameron-Jones, C. M., & Wilshire, C. E. (2007). Lexical competition effects in two cases of non-fluent aphasia. *Brain and Language*, *1*(103), 136-137.

Caramazza, A. (1997). How many levels of processing are there in lexical access?. *Cognitive neuropsychology*, *14*(1), 177-208.

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.

Cave, C. B. (1997). Very long-lasting priming in picture naming. *Psychological Science*, 8, 322–325.

Cheng, X., Schafer, G., & Akyürek, E. G. (2010). Name agreement in picture naming: an ERP study. *International Journal of Psychophysiology*, *76*(3), 130-141.

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.

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.

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.

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.

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.

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.

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.

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.

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.

Finkbeiner, M., Gollan, T. H., & Caramazza, A. (2006). Lexical access in bilingual speakers: What's the (hard) problem?. *Bilingualism*, *9*(2), 153.

Francis, W. S. (2014). Repetition priming in picture naming: Sustained learning through the speeding of multiple processes. *Psychonomic bulletin & review*, *21*(5), 1301-1308.

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.

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

Gilhooly, K. J., & Gilhooly, M. L. (1979). Age-of-acquisition effects in lexical and episodic memory tasks. *Memory & Cognition*, *7*(3), 214-223.

Gupta, P., & Cohen, N. J. (2002). Theoretical and computational analysis of skill learning, repetition priming, and procedural memory. *Psychological review*, *109*(2), 401.

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.

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.

Indefrey, P. (2011). The spatial and temporal signatures of word production components: a critical update. *Frontiers in psychology*, *2*.

Indefrey, P., & Levelt, W. J. (2004). The spatial and temporal signatures of word production components. *Cognition*, *92*(1), 101-144.

Janssen, N., Schirm, W., Mahon, B. Z., & Caramazza, A. (2008). Semantic interference in a delayed naming task: evidence for the response exclusion hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(1), 249.

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.

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.

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.

Lachman, R. (1973). Uncertainty effects on time to access the internal lexicon. *Journal of Experimental Psychology*, *99*(2), 199.

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).

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.

LaGrone, S., & Spieler, D. H. (2006). Lexical competition and phonological encoding in young and older speakers. *Psychology and Aging*, *21*(4), 804.

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.

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.

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.

Luce, R. D. (1959). *Individual Choice Behavior: A Theoretical Analysis*. New York: Wiley.

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.

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*.

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

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.

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.

Nozari, N., & Hepner, C. R. (2018). To select or to wait? The importance of criterion setting in debates of competitive lexical selection. *Cognitive neuropsychology*, 1-15.

Oldfield, R. C., & Wingfield, A. (1965). Response latencies in naming objects. *Quarterly Journal of Experimental Psychology*, *17*(4), 273-281.

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. (in preparation). Competition in normal production.

Oppenheim, G. M., & Balatsou, E. (2019). Lexical competition on demand. *Cognitive neuropsychology*, 1-4. DOI: 10.1080/02643294.2019.1580189

Oppenheim, G. M., Dell, G. S., & Schwartz, M. F. (2010). The dark side of incremental learning: A model of cumulative semantic interference during lexical access in speech production. *Cognition*, *114*(2), 227-252.

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.

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), 1583-1594.

Piai, V., & Roelofs, A. (2013). Working memory capacity and dual-task interference in picture naming. *Acta psychologica*, *142*(3), 332-342.

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.

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.

Rodríguez-Ferreiro, J., Menéndez, M., Ribacoba, R., & Cuetos, F. (2009). Action naming is impaired in Parkinson disease patients. *Neuropsychologia*, *47*(14), 3271-3274.

Roelofs, A. (1992). A spreading-activation theory of lemma retrieval in speaking. *Cognition*, *42*(1-3), 107-142.

Roelofs, A. (2003). Goal-referenced selection of verbal action: modeling attentional control in the Stroop task. *Psychological review*, *110*(1), 88.

Roelofs, A. (2018). A unified computational account of cumulative semantic, semantic blocking, and semantic distractor effects in picture naming. *Cognition*, *172*, 59-72.

Roelofs, A. P. A., & Piai, V. (2015). Aspects of competition in word production: Reply to Mahon and Navarrete.

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.

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.

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.

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.

Snodgrass, J. G., & Yuditsky, T. (1996). Naming times for the Snodgrass and Vanderwart pictures. *Behavior Research Methods*, 28(4), 516-536.

Szekely, 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.

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.

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.

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.

Vitkovitch, M., & Tyrrell, L. (1995). Sources of disagreement in object naming. *The Quarterly Journal of Experimental Psychology*, *48*(4), 822-848.

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.

Wiggs, C. L., & Martin, A. (1998). Properties and mechanisms of perceptual priming. *Current opinion in neurobiology*, *8*(2), 227-233.

Wixted, J. T., & Ebbesen, E. B. (1991). On the Form of Forgetting. *Psychological Science*, *2*(6), 409–415. http://doi.org/10.1111/j.1467-9280.1991.tb00175.x