The psychological reality of picture name agreement.

Evangelia Balatsou   Simon Fischer-Baum   Gary M. Oppenheim*

Abstract

Picture name agreement is commonly used as both a control variable and independent variable in studies of language production. It describes the proportion of participants who volunteer a picture's modal name in a norming study—a population-level descriptor—but researchers often assume that name agreement also indexes cognitive processes that occur within individuals. For instance, if norms show that 50% of speakers name a picture as *couch*, then each time a person tries to name the picture, they might 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 (*e.g.*, 50% of speakers prefer the name *couch*), continually developed through experience. 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 Experiment 1, 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. Experiment 2 replicated this result and further showed that prior selections predicted Session 3 outcomes better than those in Session 2, in line with an incremental learning account. This is the first direct demonstration that picture name agreement has some psychological validity, but also reveals that it does not directly index within-participant lexical competition as previously assumed.

1. Introduction

To what extent is word production a probabilistic process? Although word errors (e.g., 'dog' 'cat') in spontaneous speech (e.g., Garrett, 1975) imply that lexical selection cannot be entirely deterministic, and phenomena like cumulative semantic interference suggest that the production system regularly coactivates semantically related words (e.g., Howard et al., 2006), identifiable errors are rare enough that it would seem reasonable to characterise lexical selection as a battle between a single 'correct' target and a host of incorrect competitors. And although computational models often use assumptions of stochasticity to predict lexical selection times (Roelofs, 1992), error rates (Dell et al., 1997), or both (Oppenheim et al., 2010), they typically avoid the question of how a speaker might choose between multiple acceptable options (e.g., Levelt et al.'s, 1999, treatment of the 'hyperonym problem'). For instance, when a speaker of American English attempts to describe a line drawing of a soft, multiple-occupancy, seating object, either 'couch' or 'sofa' would be a reasonable response, and both terms are commonly used in their linguistic community. But it would be premature to ask *how* speakers choose between multiple acceptable options without first asking *if* speakers choose between multiple acceptable options, so that second question will be the focus of this paper.

Seemingly relevant empirical data on variability in the outcomes of successful word retrieval comes

*School of Psychology, Bangor University, Bangor, Gwynedd, LL57 2AS, UK, g.m.oppenheim@bangor.ac.uk
Thanks to Annie Blazer for assistance with data collection and transcription in Experiment 2.

from picture naming norms, in the form of an item characteristic known as *picture name agreement*. Name agreement is an empirically derived measure of the proportion of speakers who independently produce a 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 49 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, 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 speakers tended to name pictures with high name agreement 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 termed 'codability', proved an even stronger predictor (Gilhooly & Gilhooly, 1979; Lachman, 1973; Lachman & Lachman, 1980; Lachman, Shaffer, & Hennrikus, 1974), and though it can be correlated with other such factors, its independent effect has been repeatedly confirmed via multiple regression, metanalysis, and factorial experiments (see Peret & Bonin, 2019). 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; Johnston et al., 2010; Szekely et al., 2003), 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 pictures with low name agreement induce some form of challenge within individual speakers, requiring a more time-consuming decision about which name to use (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 merely predict aggregate group behaviour to an 'off-label' use for 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 & Brown, 1988), phonological encoding (LaGrone & Spieler, 2006) and repetition priming in picture naming tasks in both children and adults (Lorsbach & Morris, 1991; Mitchell & Brown, 1988), and name agreement effects on naming errors have been described as evidence of semantic and lexico-semantic impairments in Alzheimer's disease (Harley & Grant, 2004; Rodríguez-Ferreiro et al., 2009).

Neuropsychological and neurophysiological studies have more specifically used name agreement to assess the dynamics of lexical selection in word production. For instance, observations that people with aphasia appear especially error-prone when naming low-agreement pictures, compared to matched controls, have led to claims that they have greater difficulty selecting among competing alternatives (Laiacona et al., 2001; Kremin et al., 2001; Cameron-Jones & Wilshire, 2007; Bose & Schafer, 2017), and fMRI-based reports of greater Left Inferior Frontal Gyrus (LIFG) activity when naming low- compared to high-agreement pictures have been cited as key evidence that the LIFG specifically mediates such selection (Kan & Thompson-Schill, 2004; Thompson-Schill, et al., 1997). Similarly, electrophysiological differences naming between high- and low-agreement pictures have been described as both general evidence for the time course of lexical selection (Cheng et al., 2010) and specific evidence for the recruitment of selective inhibition mechanisms to suppress alternative names (Shao et al., 2014). Researchers have thus used name agreement effects to assess both the cognitive processes and neural substrates of word production.

Such uses of name agreement typically tie it to word selection, most often framing its effects within the theory of competitive lexical selection (e.g., Bates et al., 2003; Bose & Schafer, 2017; LaGrone & Spieler, 2006; Nozari & Hepner, 2019). 'Competition' in this sense refers to the controversial idea that within-speaker co-activation of candidate names (e.g., 'couch' versus 'sofa') makes the process of lexical selection more time-consuming (Levelt et al., 1999; Roelofs, 1992; 2003; Howard et al., 2006). According to the competition narrative, naming a high-agreement picture of a dog, for instance, imposes no difficulty because no other names exist or compete for selection, but naming a low-agreement picture of a couch should impose great difficulty because it can also be named as 'sofa' or 'settee', frequent alternatives identified via picture naming norms. Each individual should consider the additional names identified by picture naming norms from other members of their linguistic community. Though such a narrative has intuitive appeal, it faces the basic problem that name agreement is an empirical measure of group-level tendencies, *prima facie* unsuited for use as a predictor of individual-level cognitive processes. Using name agreement to predict or manipulate word production difficulty therefore require four major implicit assumptions:

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 ($\alpha / \sigma(\alpha)$; 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 norming studies 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 substantial 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.** Such serial independence has recently been proposed as the basis for a competition model of semantic interference effects in word production (Roelofs, 2018). Although perhaps less obvious than the preceding points, this point also directly follows from the assumption that group-level norms index individual-level cognitive processes: each speaker carries a distinct history of language use; if that history de-synchronises a speaker's semantic-to-lexical mappings from those of the group—for instance, by allowing them to accumulate a preference for sofa over couch—the group-derived distributions will no longer index their individual distributions. This point is especially important 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, by relaxing the serial independence assumption, it is easy to imagine that 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 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. These adjustments provide momentum to select and reinforce the same target in the future, explaining persistent biases in much the same way that they explain perseveration errors (*ibid*, Simulation 4). All else equal, such adjustments should accumulate into stable speaker-specific tendencies to use particular words: idiolects. Low name agreement in norming studies, 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 picture name agreement should predict individual-level competition, because, as a measure of (*between participants*) population-level variation, it is unclear whether name agreement is even a psychologically valid predictor of the underlying (*within participants*) lexical co-activation.

Although it is possible that the between-participants variation that is measured by picture naming norms does indeed index the range and relative strengths of the names that each individual considers (henceforth, 'the Luce choice account'), it is also possible that the between-participants variation that is measured by picture naming norms simply reflects between-participants variation (henceforth, 'the idiolect account'). Because traditional norming studies ask individuals to name a set of pictures just once, they cannot distinguish between these possibilities.[1] In Experiment 1, we distinguish between these accounts—and finally assess the psychological reality of name agreement and the stochasticity of lexical selection—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-participant 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. Experiment 2 then extends this approach to three sessions to test a prediction of the incremental learning-based idiolect account: if individuals accumulate robust biases to choose particular words, then these reinforced preferences should affect their name choices more strongly in the third session than in the second.

[1] 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.

2. Experiment 1

2.1 Methods

2.1.1 Overview

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).

2.1.2 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 our previous norming study (Oppenheim, 2021). The study was approved by Bangor University Ethics Committee and participants received course credit or cash compensation.

2.1.3 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 participants. 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; the voicekey was used to ensure comparability with similar timed naming paradigms. 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 200 ms, followed by a blank screen for 500ms. Next, a picture (422 x 422 pixels) appeared at the centre of the screen for 3000 ms or until the voicekey triggered, followed by a variable ITI of 900-1900ms. Short self-paced rests followed each 105-trial block. One to two weeks later, the participant returned to repeat the full procedure.

2.1.4 Analytical approach

Responses were initially transcribed on-line and later confirmed via audio recordings. Our recent norms from the same population (Oppenheim, 2021) 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 v1.1-27.1 library (Bates et al., 2021) in R 4.0.3 (R Core Team, 2020). 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. All complete regression tables are provided in the Supplementary Material section.

2.2. Results

Excluding 16 trials (0.06%) in which a voicekey error ended the trial early (< 300ms post stimulus onset) leaves 25,984 total picture naming attempts for our analyses (12,985 in the first session and 12,999 in the second session), summarised in Table 1.

2.2.1 Population-level name agreement

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, 2021). 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, 2021, and Session 1 of the current experiment: r= .93, p < .001) and secondary name agreement (excluding 63 items without a secondary name: r= .87; p < .001). Such by-item correspondences also remained in Session 2, for both the dominant name (r= .91, p < .001) and the secondary name (r= .84; p < .001). By-item response frequencies also correlated well between Session 1 and 2 within this experiment, for both dominant (r= .92, p < .001) and secondary (r= .85, p < .001) name agreement. Thus, considered at the population level, name selections were consistent with previous norms and appear relatively stable across sessions.

2.2.2 Individual-level name agreement

We can also ask whether the same individuals tended to use the same names across sessions. For instance, Table 1 indicates that 79% of participants named items using their dominant names in Session 1. If this proportion simply reflects a sampling of individuals and their preferred names—79% of our participants happened to prefer these pictures' dominant names, as described in our 'idiolect' account—then we would expect that *the same* 79% 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, .79. On the other hand, if they were stochastically selecting among responses each time, as described in our 'Luce choice' account, then only 79% of the original 79% should

Table 1: Response frequencies for each session in Experiments 1 and 2. Frequencies from Oppenheim's (in prep.) UK norms, calculated in the same way, are also provided for comparison.

	Norms	**Experiment 1**				**Experiment 2**					
		Session 1		**Session 2**		**Session 1**		**Session 2**		**Session 3**	
	Mean	Mean	N	Mean	N	Mean	N	Mean	N	Mean	N
Dominant	0.79	0.79	10287	0.80	10445	0.80	10377	0.82	10658	0.83	10795
Secondary	0.11	0.10	1261	0.10	1305	0.09	1207	0.09	1139	0.09	1135
Other	0.07	0.08	996	0.07	935	0.07	927	0.06	762	0.05	713
Omission	0.03	0.03	441	0.02	314	0.04	468	0.03	367	0.03	347
Total	1	1	12985	1	12999	1	12979	1	12926	1	12990

use the dominant name in both sessions. Thus the probability of a person using the dominant name in both sessions would be .79ˆ2=.62. Figure 1 visually illustrates the distinction between these functions.

To statistically evaluate an analogue of this idea, we first excluded 113 items that invariably elicited their dominant names in Session1 (leaving 407 items and 10,174 trials for this analysis), and then, as described in the Method section, 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 (2021) 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). Our approach kept the correlations of fixed effects below $|0.03|$. Basically, any individual speaker has some proportion of items for which they previously used the dominant name and some proportion for which they did not: the normed name agreement effect can be assessed within each subset for each participant. Similarly, every item has only one normed name agreement value, but some participants who previously named it using the dominant name and some who did not, so comparing their probabilities, for each item, allows the model to estimate the effect of individuals' prior use (and estimates the within-participants effect in much the same way, by essentially assessing it within bands of items with similar name agreement estimates). The key question for the analysis is whether these factors individually contribute to predicting response likelihoods.

First considering our Luce choice account, if between-participants measures of dominant name agreement predict the within-participants 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: 85.14:1, $\beta_{\text{DominantNameAgreement}}= 4.44$, $SE= 0.21$, $p < .001$).

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: 10.56:1 , $\beta_{\text{UsedDominantInSession1}}= 2.36$, $SE= 0.12$, $p < .001$). Thus we find support for both for the

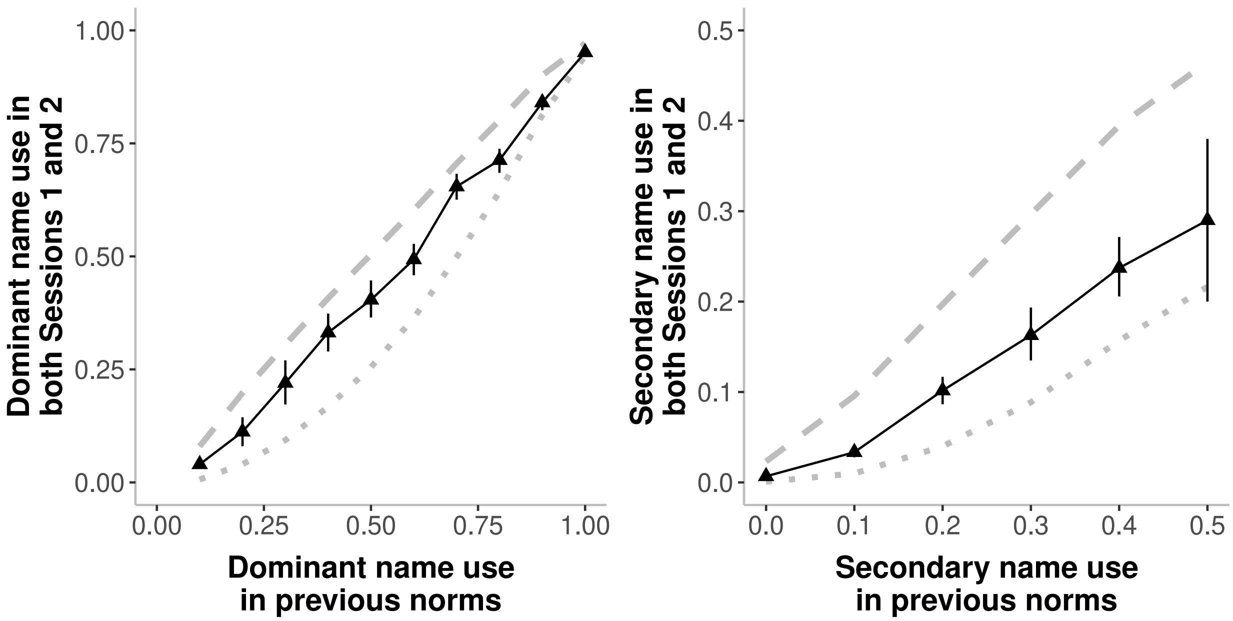

Figure 1: Observed name consistency in Experiment 1 compared to predictions from the idiolect account (dashed grey line), assuming only cross-speaker heterogeneity, and the Luce choice account (dotted grey line), assuming only within-speaker stochastic selection. Panel A depicts consistency for each pictures dominant name; Panel B depicts consistency for each pictures second most commonly used name. The x-axis represents the stimulus-contingent frequency of the specified name in a recent norming study with the same population (Oppenheim, 2021). The y-axis represents the frequency with which participants in Experiment 1 produced the specified name in both Session 1 and Session 2.

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 a reduced form of the above model to a relevant subset of the data (namely, the 1263 trials, of those listed above, that participants had initially named using an items second most common name) 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: 38.25:1 , $\beta_{\text{DominantNameAgreement}}$= 3.64, *SE*= 0.43, $p < .001$).

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 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 (2021) 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 207 items that no participant had named using the secondary name in Session 1; this leaves 313 items and 7824 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: 10.81:1, $\beta_{\text{SecondaryNameAgreement}}$= 2.38 , *SE*= 0.16, $p <$ .001) and participants' own productions in Session 1 (odds ratio: 718.74:1 , $\beta_{\text{UsedSecondaryInSession1}}$= 6.58, *SE*= 0.53 $p < .001$). 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 regress 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 secondary name than to a weaker secondary name in Session 2 (odds ratio: 8516.83:1 , $\beta_{\text{SecondaryNameAgreement}}$= 9.05, *SE*= .66, $p < .001$). 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. A Monte Carlo analyses, presented in the Supplementary Material, shows that this trend further holds among even lower ranking responses.

3. Experiment 2

Experiment 1 showed that individuals' own preferences and population-level trends jointly affect their word selections, thus providing support for both the Luce choice account and the idiolect account. In the Introduction, we motivated the idiolect account as a corollary an incremental learning model of word production (*viz*, Oppenheim et al., 2010). According to that account, speakers continually re-tune their semantic-to-lexical mappings, reinforcing those that have proven most useful in the recent past (*i.e.*, those supporting target names) and weakening those that have proven less useful (those supporting alternatives). This incremental learning should produce long-lasting priming of particular names, so one possible interpretation of the results in Experiment 1 is that support for the idiolect predictors in Session 2 reflected, at least in part, persistent priming of responses from Session 1, one week earlier.

An alternative interpretation that preserves the assumption of serial independence is that speakers may vary in their word preferences—for instance, as a result of having acquired ones language from a distinct subcommunity—but maintain these stable variations throughout their lives. This interpretation could be compatible with theoretical accounts that explicitly reject the possibility of continual incremental learning in semantic-to-lexical mappings (e.g., Roelofs, 2018), that could be seen as analogous to a critical period hypothesis.

To distinguish between these accounts, in Experiment 2 we replicate Experiment 1 with 25 new participants, but now extended it to three sessions. Again, we focus on the likelihood of a participant producing the same name for a picture in two sequential sessions. However, with three sessions, we can now compare the likelihood of a participant producing the same name in Sessions 1 and 2 to that in Sessions 2 and 3. From the accounts described above, we can derive the following contrasting predictions:

1. If speakers continually develop their word preferences, per the incremental learning account, then these modifications should accumulate in much the same way as cumulative semantic interference. The idiolect predictor (*UsedNameInPreviousSession*) should therefore affect word selections more strongly in Session 3 than in Session 2.

2. If speakers simply maintain variations that they developed during an earlier acquisition process, then the idiolect predictor should affect word selections to a similar degree in Sessions 2 and 3.

3. Finally, if speakers do not maintain distinct word preferences (the simple Luce choice account), then the main effect of the idiolect predictor should simply fail to replicate.

3.1 Methods

3.1.1 Overview

Methods for Experiment 2 exactly followed those of Experiment 1, except that each participant now completed three sessions instead of two.

3.1.2 Participants

Twenty-five Bangor University students (18 female, Mean age : 19.4 years, *SD*= 0.9) participated in exchange for course credit. 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 Experiment 1 nor Oppenheim's (2021) previous norming study. The study was approved by Bangor University Ethics Committee and participants received course credit or cash compensation.

3.1.3 Materials, apparatus and procedure

Each session exactly followed the methods of Experiment 1. One to two weeks after the second session, the participant returned to repeat the full procedure a third time.

3.1.4 Analytical approach

Transcription and response coding exactly followed those in Experiment 1. Statistical analyses apply confirmatory logistic mixed effects regression, as in Experiment 1. Complete regression tables are provided in the Supplementary Material.

3.2. Results

Excluding 105 trials (0.2%) in which a voicekey error ended the trial early (< 300ms post stimulus onset) leaves 38,895 total picture naming attempts for our analyses (12,979 in Session 1, 12,926 in Session 2, 12,990 in Session 3). General response characteristics, summarised in Table 1, were comparable to those in Experiment 1.

Extending the analytical approach from Experiment 1, we used maximal logistic mixed effects regression to predict participants' likelihood of producing a picture's dominant name in Session 2 or 3 as a function of (1) its population-level name agreement from Oppenheim's (2021) recent Bangor norming study (a continuous measure from 0:1, centered);

(2) whether the individual participant produced the dominant name in the previous session (a binary measure {0,1}, centered). To test changes in these effects, we now added to the model Session number (an integer {2,3}, centered) and its interactions with the population and individual predictors. To estimate random slopes within items, we excluded the 162 items for which every participant produced the dominant name in either Session 1 or Session 2, leaving 358 items and 17,842 trials for this analysis.

Main effects in this analysis replicated several basic patterns from Experiment 1. First, participants were again more likely to produce dominant names that were more frequently used in population-level norms (odds ratio: 61.02:1, $\beta_{\text{DominantNameAgreement}}$= 4.11, *SE*= 0.21 , $p < .001$), and they grew more likely to produce dominant names across consecutive sessions (odds ratio: 1.15:1, β_{Session}= 0.14 , *SE*= 0.55, $p = .012$). Second, participants were also more likely to produce a dominant if they themselves had produced it in the previous session (odds ratio: 9.75:1, $\beta_{\text{UsedDominantInPreviousSession}}$= 2.28 , *SE*= 0.91, $p < .001$).

The key question, illustrated in Figure 2a, is how these effects develop across sessions. The effect of population-level dominant name agreement did not significantly change across sessions (odds ratio: 1.20:1, $\beta_{\text{Session x DominantNameAgreement}}$ = 0.19 , *SE*= 0.28, $p = .51$). In line with the prediction from the incremental learning account, however, speakers' individual preferences became significantly more pronounced (odds ratio: 1.74:1, $\beta_{\text{Session x UsedDominantInPreviousSession}}$= 0.55 , *SE*= 0.09, $p < .001$).

As in Experiment 1, similar patterns also held in an analogous analysis of non-dominant name use. To allow the inclusion of within items random effects, we excluded 279 items with secondary names that participants never used in Session 1 or 2, leaving 12,009 trials for this analysis. Participants were overall more likely to produce secondary names that were more commonly observed in our norming studies (odds ratio: 274.44:1, $\beta_{\text{SecondaryNameAgreement}}$= 5.61, *SE*= 0.43, $p < .001$), and though they grew significantly less likely to produce secondary names across sessions (odds ratio: 0.80:1, β_{Session}= -0.22, *SE*= 0.07, $p = .003$), they remained more likely to produce secondary names that they had used themselves in the previous session (odds ratio: 10.02:1, $\beta_{\text{UsedSecondaryInPreviousSession}}$= 2.30 , *SE*= 0.10, $p < .001$). And though the effect of population-level secondary name agreement did not significantly change across sessions (odds ratio: 2.54:1, $\beta_{\text{Session x SecondaryNameAgreement}}$ = 0.93 , *SE*= 0.52, $p = .07$), individual preferences again became significantly more pronounced by this measure (odds ratio: 1.42:1, $\beta_{\text{Session x UsedSecondaryInPreviousSession}}$= 0.35, *SE*= 0.12, $p = .003$; see Figure 2b).

Finally, it is worth noting that analyses of name-switching patterns provide results that are consistent with both those already previously described for Experiment 1 and those just reported for Experiment 2. As in Experiment 1, participants were more likely to switch from secondary names to dominant names that were more frequently used in population-level norms (odds ratio: 60.89:1 , $\beta_{\text{DominantNameAgreement}}$= 4.11, *SE*= .45, $p < .001$), and this trend became more pronounced in later sessions (odds ratio: 3.96:1, $\beta_{\text{Session x DominantNameAgreement}}$ = 1.38 , *SE*= 0.69, $p = .046$), as name switching became less frequent overall (odds ratio: 0.81:1, β_{Session}= -0.22 , *SE*= 0.11, $p = .041$). Analyses of switches from dominant to secondary names showed similar influences: participants were more likely to switch from dominant names to secondary names that were more frequently used in population-level norms (odds ratio: 4380.43:1 , $\beta_{\text{SecondaryNameAgreement}}$= 8.38, *SE*= 0.46, $p < .001$), though this trend did not become significantly more pronounced in later sessions (odds ratio: 2.01:1, $\beta_{\text{Session x SecondaryNameAgreement}}$ = 0.74 , *SE*= 0.61, $p = .22$), and name switching became less frequent overall (odds ratio: 0.78:1, β_{Session}= -0.25 , *SE*= 0.10, $p = .013$).

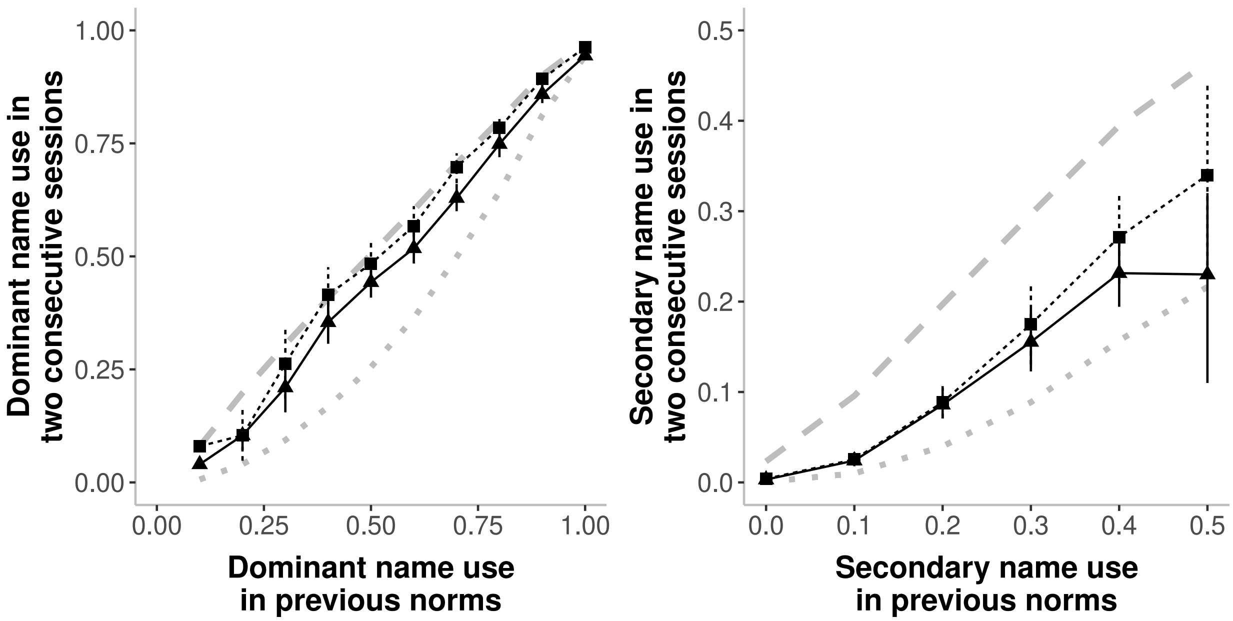

Figure 2: Observed name consistency for sequential sessions in Experiment 2 compared to predictions from the idiolect account (dashed grey line), assuming only cross-speaker heterogeneity, and the Luce choice account (dotted grey line), assuming only within-speaker stochastic selection. Panel A depicts consistency for each pictures dominant name; Panel B depicts consistency for each pictures second most commonly used name. The x-axis represents the stimulus-contingent frequency of the specified name in a recent norming study with the same population (Oppenheim, in prep). The y-axis represents the frequency with which participants in Experiment 2 produced the specified name in two consecutive sessions: Sessions 1 and 2 (solid black line) or Sessions 2 and 3 (dotted black line).

4. Discussion

Returning to our original question, is word production a probabilistic process in the sense that speakers stochastically choose between multiple acceptable options each time they select a word? On the assumption that population-derived estimates index the range of names that individual speakers consider during naming tasks, picture name agreement has been associated with behavioural, neuroimaging, and electrophysiological effects, which have in turn been characterised as evidence of that assumed conflict. However, the basic question has until now remained untested, so in this paper we used a repeated picture naming task to assess both whether population-derived norms predict within-speaker variation in naming behaviour and whether speakers might accumulate robust idiosyncratic preferences for particular words despite variation in their linguistic communities. To our knowledge, this is the first systematic investigation of picture name consistency in unimpaired[2] adults and, remarkably, our

[2]There has been recent interest in response stability in the neuropsychological literature (van Scherpenberg et al., 2019), but without comparison to neurally intact populations; thus this study provides a useful baseline.

results support both propositions: speakers' word selections in Session 2 of Experiment 1 and Sessions 2 and 3 of Experiment 2 were jointly predicted by the distribution of names in their linguistic community and their own previous responses, suggesting that they are sensitive to the linguistic variation observed in their communities, but nonetheless develop and maintain their stable word preferences across naming episodes.

Before delving into these questions, though, it is worth noting that our group-level correlations support the validity of picture name agreement for its on-label use, that is, 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 this study corresponded well with those observed in recent norming studies from the same population (Oppenheim, 2013), and with each other. Thus, such population-level norms have demonstrated utility for predicting the distributions of names across speakers, consistent with their traditional use in selecting materials for new experiments, particularly for predicting those distributions in the first naming instance of an experiment. However, their utility is more limited for predicting responses or variations across repeated trials.

4.1 Population-level norms predict within-speaker variability

The first major finding from this study is evidence for the Luce choice-inspired stochastic selection account, in the form of both predictable name selections, but more importantly predictable naming switches. First, within-participants analyses in both experiments demonstrated that speakers were more likely to use names that were more frequently attested in the population-derived naming norms, regardless of whether they themselves had used those names initially, and this pattern held for both dominant names and alternatives. Logistic regressions of within-participants name switching further demonstrated that population-level name agreement predicts name co-availability within individual speakers, 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 measures index the distribution of names within individual speakers), it confirms that speakers both maintain multiple candidate names and tend to switch to the names that other speakers use more frequently to describe the same stimuli. Second, the fact that population-level norms also predict speakers' likelihood of spontaneously switching to non-dominant names further demonstrates that such name changes cannot simply be explained as moves toward a single 'correct' response.

Our name use and name switching measures thus indicate that picture name distributions from norming studies predict at least the co-availability of responses within individual speakers, which is a crucial precondition for the common, if controversial, interpretation of name agreement effects as reflecting response competition. Though these data cannot directly show that speakers necessarily coactivate multiple labels within the same trial, that assumption is common to both competitive (Howard et al, 2006; Roelofs, 2018) and noncompetitive (Oppenheim et al, 2010) accounts of word production effects (see e.g., Nozari & Pinet, 2020). 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).

Whether such effects specifically require competitive lexical selection processes, in the sense of, e.g. Levelt, et al. (1999), is a separate matter. The past fifteen years have brought considerable criticism of both the hypothesis of lexical selection by competition and the specific empirical evidence that researchers have claimed to support it. For instance, although semantic picture word interference is empirically robust, it is not clear that it reflects the same kind of selection processes that speakers would engage for typical communicative production (e.g., Mahon et al., 2007; Oppenheim & Balatsou, 2019). And though cumulative semantic interference is similarly robust, it can be sufficiently explained by incremental learning processes without imposing strong constraints on within-trial selection processes (Oppenheim et al., 2010 *et passim*). Many of the studies of name agreement effects that we described in the Introduction were published before the competition claim was broadly questioned: an era when demonstrating that a picture elicited many different names *and* long naming latencies was sufficient to prove that competition from those names *caused* the long naming latencies. Although demonstrating within-speaker lexical co-activation is a crucial first step, future work along these lines will need to more carefully distinguish evidence for lexical selection by competition from evidence of mere lexical co-activation.

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; 2021). Under a competitive selection 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 a 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 norming studies systematically overestimate within-speaker variability. In each experiment, within-participants, within-items analyses demonstrated that speakers develop and maintain preferences for even non-dominant names. These robust individual differences imply that population-level name agreement also reflects individuals' stable word preferences, and Experiment 2 traced the development of these idiosyncrasies over three sessions, each approximately a week apart, to demonstrate that they represent very long-lasting accumulations of experience. Such persistent changes as a result of word production can be readily identified as incremental learning: small, likely implicit, experience-driven adjustments to the semantic-to-lexical connections that support the retrieval of particular names. We have argued elsewhere that such incremental learning underpins a range of empirical phenomena that had otherwise been attributed to within-trial competition (cf. Roelofs, 2018), including long-lag perseveration errors (Fischer-Baum, Irons, Oppenheim, 2018; Oppenheim, et al., 2010). By the same token, an incremental learning model of word production would predict the accumulation of such idiosyncrasies if left unchecked. In Oppenheim et al.'s (2010) Dark Side model, a supervised learning algorithm provides such a check against runaway errors: following a "dog" "cat" slip, a corrective process strengthens the connections supporting "dog" and weakens those supporting "cat". Speakers are demonstrably able to detect and correct such frank errors (e.g., Levelt, 1983), so it is reasonable to assume that such self-supervision could provide a basis for corrective learning. But when there is little reason to prefer one name over an alternative (e.g., "couch", "sofa") there is no reason to expect such a check (see Nozari & Hepner, 2019, for a related point), allowing any reinforcement of a chosen response to simply persist until it affects the next retrieval. Thus, even if a 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 idiosyncratic 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 nevertheless 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 abandoned the latter. 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 (e.g., 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 flexibility and interacting with speakers with divergent preferences may provide the necessary impetus to regularly switch between similarly appropriate names and thus maintain them in relative equilibrium. And though a thorough discussion is beyond the scope of the present work, such voluntary name changing may thereby provide the basis for a simple accessibility-based account of several behavioural and neural correlates of picture name agreement.

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, predicts 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 communities. But we have also shown that individual speakers continually develop and reinforce changes to their semantic-to-lexical mappings that put their word preferences in conflict with their likely interlocutors. 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 effects. This efficacy is somewhat surprising, but not too surprising, because it is still probably the case that pictures that have multiple names elicit less target lexical activation and more lexical co-activation, even if population-based measures of name agreement are not the perfect way to predict 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.

Credit statement

Evangelia Balatsou: Investigation, Formal analysis, Writing - Original draft; Simon Fischer-Baum: Conceptualization, Formal analysis, Writing - Reviewing and Editing; Gary Oppenheim: Conceptualization, Methodology, Formal analysis, Supervision, Validation, Visualization, Writing - Original draft, Writing - Reviewing and Editing.

Acknowledgements

Thanks to Annie Blazer for assistance with data collection and transcription in Experiment 2. This work was supported in part by NSF BCS1752751.

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., et al. (2021). lme4: Linear mixed-effects models using Eigen and S4. CRAN.

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.

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.

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.

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.

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., 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.

Fischer-Baum, S., Irons, S., & Oppenheim, G. M. (2018). Breaking the dark side: A computational neuropsychological approach. Poster presented at the *10th International Workshop on Language Production*, July 1-5, Nijmegen, Netherlands.

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.

Garrett, M. F. (1975). The analysis of sentence production. In G. H. Bower (Ed.), *The psychology of learning and motivation: advances in research and theory* (pp. 133–177).

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.

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

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.

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. M. (1983). Monitoring and self-repair in speech. *Cognition*, *14*(1), 41–104.

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.

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. (2019). To select or to wait? The importance of criterion setting in debates of competitive lexical selection. *Cognitive neuropsychology*, 36(5-6 193-207.

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*(April 2019), 100875.

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 presented at Architectures and Mechanisms of Language Processing 2017, Lancaster, UK.

Oppenheim, G. M. (2021). Lexical selection in normal language production. Submitted for publication.

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.

Perret, C., & Bonin, P. (2019). Which variables should be controlled for to investigate picture naming in adults? A Bayesian meta-analysis. *Behavior Research Methods*, 51(6), 2533–2545.

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

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.

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

Szekely, A, D'Amico, S., Devescovi, A., et al. (2003). Timed picture naming: Extended norms and validation against previous studies. *Behavior Research Methods, Instruments, and Computers, 35(4), 621-633*.

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.