Do words compete as we speak? A systematic review of picture-word interference (PWI) studies investigating the nature of lexical selection

Article

Published Version

Korko, M., Bose, A. ORCID: https://orcid.org/0000-0002-0193-5292, Jones, A., Coulson, M. and de Mornay Davies, P. (2024) Do words compete as we speak? A systematic review of picture-word interference (PWI) studies investigating the nature of lexical selection. Psychology of Language and Communication, 28 (1). pp. 261-322. ISSN 2083-8506 doi: 10.58734/plc-2024-0011 Available at https://centaur.reading.ac.uk/119764/

Identification Number/DOI: 10.58734/plc-2024-0011 <https://doi.org/10.58734/plc-2024-0011>

Publisher: Sciendo

Psychology of Language and Communication 2024, Vol. 28 No. 1

DOI: 10.58734/plc-2024-0011

Małgorzata Korko[1], Arpita Bose[2], Alexander Jones[3], Mark Coulson[4], and Paul de Mornay Davies[3]

[1] The Institute of Education, The Maria Grzegorzewska University, Poland

[2] School of Psychology and Clinical Language Sciences, University of Reading, United Kingdom

[3] Department of Psychology, School of Science and Technology, Middlesex University, United Kingdom

[4] University of West London, School of Human and Social Sciences, United Kingdom

Do words compete as we speak? A systematic review of picture-word interference (PWI) studies investigating the nature of lexical selection

This review synthesizes findings from 117 studies that have manipulated various picture-word interference (PWI) task properties to establish whether semantic context effects reflect competitive word retrieval, or are driven by noncompetitive processes. Manipulations of several PWI task parameters (e.g., distractor visibility) have produced contradictory findings. Evidence derived from other manipulations (e.g., visual similarity between targets and distractors) has been scarce. Some of the manipulations that have furnished reliable effects (e.g., distractor taboo interference) do not discriminate between the rival theories. Interference from nonverbal distractors has been shown to be a genuine effect dependent on adequate lexicalization of interfering stimuli. This supports the swinging lexical network hypothesis and the selection-by-competition-with-competition-threshold hypothesis while undermining one of the assumptions of the response exclusion hypothesis. The contribution of pre-lexical processes, such as an interaction between distractor processing and conceptual encoding of the target to the overall semantic context effect is far from settled.

Key words: picture word interference, lexical selection, language production,competition

Address for correspondence: Małgorzata Korko

The Institute of Education, The Maria Grzegorzewska University, Warszawa, Poland.

E-mail: mkorko@aps.edu.pl

A Brief History of The Picture-Word Interference Paradigm

In the last forty-five years, the picture-word interference (PWI) paradigm has been one of the main experimental tools for investigating the competitive nature of word production (e.g., Rosinski, 1977; Rosinski et al., 1975). In the task, the speaker sees a picture of an object (e.g., a table) accompanied by a distractor – usually a visually or aurally presented word (e.g., "chair"), and is asked to name the object as quickly as possible while ignoring the distractor. A common finding is that when the name of the object and the distractor word are semantic co-ordinates (e.g., TABLE-"chair"), objects are named more slowly than when these items are unrelated (e.g. TABLE-"cloud"). This decrement in performance, which amounts to about 21 milliseconds (Bürki et al., 2020), is described as the picture-word interference effect (see also the *semantic interference effect*, *semantic context effect, semantic category effect*, and *context effect*) and has been taken as evidence for the activation of nontarget words that potentially delay the production of the target word, cause some other type of interference in the form of slips of the tongue (e.g., saying "chair" when naming a picture of a table), and/or decreased fluency (e.g., hesitations, repetitions, false starts and/or sound prolongations). However, an increasing number of observations of polarity reversals from interference to facilitation following manipulations of the various aspects of the task has challenged the original account, fueling the development of alternative, noncompetitive hypotheses concerning the origins of the effects.

To date, it is not entirely clear where in the process of recognizing and naming a pictured object that is accompanied by a distracting stimulus (either a word or a picture), interference (or facilitation) takes effect, or how such an effect comes about. The answers are hard to find because results are often conflicting, methodologies vary, and the PWI literature is voluminous. Several helpful reviews have been published to elucidate the origins of semantic context effects, and thus illuminate the cognitive processes involved in spoken word production, but these focused either on advancing or challenging a particular account (see Abdel Rahman & Melinger, 2019, for the swinging lexical network hypothesis; Mahon et al., 2007, for the response exclusion hypothesis; Mulatti & Coltheart, 2012 for a critique of the response exclusion hypothesis), or presented a mere overview of selected methodologies, their respective findings, and proposed interpretations (e.g., Spalek et al., 2013).

Bürki et al. (2020) have recently undertaken a qualitative and meta-analytic review of PWI studies, with a focus on task design and materials properties that are thought to modulate the semantic interference effect (i.e., the stimulus onset asynchrony, familiarization, number of target words, number of semantic categories, number of repetitions, mean naming times in the unrelated condition, distractor frequency, distractor length, and distractor picturability). The task properties analysed by Bürki et al. (2020) lent themselves well to a quantitative analysis, but few of them have been directly manipulated in the reviewed studies, that is, the authors of those studies made no a priori assumptions about the effects of these task characteristics on naming latencies.

In contrast, the current review investigated PWI task characteristics (distractor properties, target-distractor relationships, task types) that have been directly manipulated in order to address the question of whether or not lexical selection (and word production, more broadly) is a competitive process. See Figure 1 for the main sites of manipulations in PWI studies. For the purpose of this review, the word "competitive" is used in its technical sense, namely, as a delay in production caused by the activation of target-related, but context-irrelevant lexical representations.

The current review organizes, summarizes, and interrogates the findings of PWI research spanning 45 years. By doing so, it aims to: (a) establish what effects (interference, facilitation, or no effect) have been observed following the various manipulations of PWI task parameters. Are the reported patterns of results consistent across the PWI studies and is the degree of consistency dependent on the site of manipulation under review? (b) Evaluate the feasibility of competitive and noncompetitive accounts of lexical selection to account for the reviewed sets of data. What are the loci of the semantic context effects and what mechanisms possibly contribute to the emergence of these effects? Is the observed effect genuinely due to lexical competition, occurring at the level of lexical encoding and due to the activation of nontarget words, or does it reside outside the language system, occurring either prelexically (before a lemma is selected) or postlexically (after a lemma is selected), and is better characterized as the outcome of other, noncompetitive processes? (c) Assess the risk of bias in each study included in this review, (d) provide recommendations about the design of PWI studies, and (e) identify gaps in PWI research and provide future directions.

Competitive and Noncompetitive Accounts of Lexical Selection Derived from Picutre-Word Interference Studies

Competitive Accounts

Proponents of the competitive view of lexical selection hold that the speed and ease with which a word is produced depends on the coactivation of nontarget lexical representations (e.g., Abdel Rahman & Melinger, 2009, 2019; Levelt et al., 1999; Piai et al., 2012; Starreveld & La Heij, 1995). The higher the activation of the lexical competitors, the more time is needed and the harder it is to produce the sought-after word.

Selection-by-Competition. The most prominent model of lexical selection, the selection-by-competition account, assumes that the semantic interference effect reflects competition between the target word and the co-activated but context-inappropriate lexical representations (Bloem & La Heij, 2003; Bloem et al., 2004; Caramazza & Costa, 2000; Levelt, 1989; Levelt et al., 1999; Roelofs, 1992; Starreveld & La Heij, 1995; Vigliocco et al., 2002; Vigliocco et al., 2004). In the context of the PWI task, the effect is thought to occur due to higher activation of categorically related distractor words (e.g., DOG-"mouse") relative to their unrelated controls (e.g., DOG-"table"). Essentially, distractors that belong to the same semantic category as targets are activated both directly by the distractor word itself ("mouse"), and indirectly by the target picture through the process of spreading activation (i.e., the processing of DOG activates its related semantic nodes, such as ANIMAL, CAT, MOUSE, LIVES ON A FARM, BARKS, etc., which in turn activate their corresponding lexical representations, "animal," "cat," "mouse," etc.). The lexical node "mouse" thus receives activation from two sources (the target and the distractor). In contrast, an unrelated distractor (e.g., "table") receives activation from a single source, notably, the distractor word alone (DOG is unlikely to spread activation to "table"). A categorically related distractor is therefore a stronger competitor than an unrelated distractor, which results in a greater delay in the selection and production of the target word.

Selection-by-Competition with a Competition Threshold. An important extension of the original competitive model of lexical selection (e.g., Roelofs, 1992; Levelt et al., 1999) is the idea that non-target representations must exceed a threshold to enter into competition with the target word (Piai et al., 2012). The notion of a competition threshold was introduced to account for the elusive nature of the effect, which is detected under one set of experimental conditions, but which disappears under another (Roelofs, 1992). There are several ways in which to boost the lexical activation of competitors in the PWI paradigm so that they reach a critical threshold. One way is to increase the number of exemplars from the same semantic category as targets. Other methods are outlined in the Distractor Format section. In natural word production, it is conceivable that certain non-target words will accrue sufficient activation to compete for selection by virtue of either being recently heard, more frequently used, or emotional in content, thereby hampering the production of the target word.

Swinging Lexical Network Hypothesis (SLNH). A core idea within the swinging lexical network hypothesis (Abdel Rahman & Melinger, 2009, 2019), developed specifically to address the issue of polarity reversals (interference turned into facilitation), is that the semantic context effect reflects a trade-off between facilitation at the level of conceptual encoding (activation spreading within the semantic network) and interference at the level of lexical processing (activation spreading to lexical representations, only one of which can be selected). The framework is based on two assumptions. One, activation flows bi-directionally within- (a concept will activate its related semantic nodes and a lexical node will activate its related lexical representations) and across the levels (concepts will activate their corresponding lexical representations and lexical representations will activate their respective concepts). Two, whether picture naming is delayed or precipitated depends on the cohort size of the activated lexical representations and their relative strength of activation.

In the case of categorically related target-distractor pairs (e.g., DOG-"mouse"), multiple related concepts (e.g., DOG, CAT, MOUSE) and their corresponding lexical nodes may be accessed due to an overlap in semantic features (both have four legs, have a tail, are furry, live on a farm). In addition, a similar set of semantic and lexical nodes (e.g., DOG, CAT, MOUSE, FOX) may be activated by the shared superordinate category node (ANIMAL). In both cases, target-related concepts and their affiliated lexical nodes are thought to boost the activation of the distractor word (e.g., involuntary activation of CAT may activate "mouse"). This type of recursive activation constrains the activated cohort, increasing resonance within the network. In effect, the distractor word accrues enough activation at the lexical level to outweigh facilitation at the conceptual level, with the net result of interference.

In contrast, recursion is less likely in the case of associatively related target--distractor pairs (e.g., DOG-"leash") for two reasons. One, there is little or no semantic feature overlap. Two, because the items do not share the same superordinate category node (one belongs to ANIMALS, the other to ACCESSORIES), the ANIMALS node activated by the picture of a DOG and spreading activation to its related concepts (e.g., DOG, CAT, MOUSE, FOX) and their respective lemmas/lexemes is unlikely to strengthen the activation of "leash." Greater diffusion within semantic and lexical networks results in weaker activation of the distractor word, which is unable to offset facilitation at the conceptual level, with the net result of facilitation.

Noncompetitive Accounts

Noncompetitive accounts of spoken word production offer alternative explanations of PWI effects. They claim that a lexical item is selected (or retrieved) once it has reached an activation threshold or after a certain time delay (dictated by external factors such as speech rate) irrespective of the co-activation of related but irrelevant representations (e.g., Caramazza, 1997; Dell & O'Seaghdha, 1991; Mahon et al., 2007; Miozzo & Caramazza, 2003).

Response Exclusion Hypothesis (REH). According to the response exclusion hypothesis (Mahon et al., 2007; see also response selection, Finkbeiner & Caramazza, 2006; Miozzo & Caramazza, 2003; and response plausibility, Lupker & Katz, 1981), the interference effect commonly observed in the PWI task arises due to the confound of response relevance (some stimuli appear to be more plausible as responses than others) rather than the direct manipulation of semantic relatedness (and thereby activation spread and competition) between targets and distractors. The delay in naming a pictured object is thus a procedural artefact that has little to do with how words are selected and produced during picture naming. The hypothesis is based on two assumptions. One, distractor words have the articulatory advantage over target pictures (naming a word is quicker than naming a picture). As such, they are more likely to access the articulatory buffer first and must be removed from it, if a target picture's name is to be articulated. Two, the speed with which a distractor response is cleared from buffer depends on how quickly the system can decide whether or not the response satisfies some "response relevance criteria;" the more plausible the response associated with the distractor, given the task requirements, the longer it takes to remove it from the buffer to produce the desired word. For example, in the case of categorically related (e.g., DOG-"mouse") and unrelated (DOG-"table") target-distractor pairs, the unrelated word "table" will be rejected sooner than the related word "mouse" because the former violates an implicit semantic category criterion of naming an animal.

In the case of associatively related stimuli that are not co-ordinates, the cost of removing distractor responses from the articulatory buffer is similar across the related and unrelated conditions because both distractor responses are equally implausible given the task requirements. For example, when one is naming a picture of a SHIP, both "anchor" and "button" in the target-distractor pairs SHIP--"anchor" versus SHIP-"button" can be easily dismissed as potential responses because both denote parts in a task in which participants name whole objects. Similarly, when naming a picture of a BED, it is comparatively easy to reject the verb "sleep" (BED-"sleep") and the verb "drive" (BED-"drive") as viable responses because both distractor words violate an implicit rule of naming an object (using a noun) as opposed to naming an action (using a verb). This minimal cost associated with removal of associatively related distractor responses from the response buffer is outweighed by the benefit of semantic priming (e.g., the concept of ANCHOR primes SHIP), resulting in overall quicker picture naming in the related than unrelated condition.

Verbal Self-Monitoring. Despite the claims that the semantic context effects are merely procedural artefacts (REH), the task of articulatory buffer clearing could in fact be part of the speech production process, with a speech-monitoring system a likely candidate for the job (Dhooge & Hartsuiker; 2010, 2012). Essentially, the delay in production could be understood in terms of the extra time needed by the verbal self-monitor to perform checks on a response that has been selected (the most highly activated word enters the articulatory buffer) but not yet articulated (and is blocking the articulatory buffer). The monitor would have to decide whether or not the response that is blocking the buffer fulfils certain contextual constraints and should actually be uttered. These could be implicit response relevance criteria (the account is compatible with REH), lexicality criterion (whether or not the response sounds like a word), appropriateness criterion (whether or not the response is socially appropriate), and so on, or explicit criteria such as task demands (e.g., use a basic noun form, not a diminutive). Indeed, empirical findings suggest that the self-monitoring mechanism is biased towards nonwords (i.e., the speaker is less likely to produce nonsense speech; Baars et al., 1975), is sensitive to the social context (i.e., the speaker is less likely to produce a swearword than an incorrect neutral word; Severens et al., 2012) and needs time to perform the required operations (i.e., the activity of the monitoring mechanism is compromised under time pressure; Gehring et al., 1993). In light of this explanation, the lexical selection itself is not a competitive process since a word is selected regardless of the activation of nontarget words. A delay in naming occurs because the selected word undergoes further checks post-lexically by a self-monitor after a lemma has been selected to determine whether or not it fits the communicative context and should in fact be articulated.

Concept Exclusion Hypothesis (CEH). It is possible that semantic information supplied by the distractor word interacts with picture processing from very early on, at the object recognition stage (within the access of stored structural knowledge), or during the process of mapping of a structural representation onto a relevant conceptual representation (object identification). There is indeed evidence that conceptual knowledge modulates early visual processing of objects (e.g., Bar, 2004; Gauthier et al., 2003). Within the context of the PWI paradigm, several authors have indicated a prelexical locus of the semantic context effects. Lupker and Katz (1981) argued that a delay in picture naming observed when the target and distractor share the same semantic category (and thus activate closely related concepts) could be attributed to the process of conceptual disambiguation. Jescheniak et al. (2014) and Matushanskaya et al. (2016) acknowledged that their data do not rule out an early, prelexical locus of the effect. In other words, it is plausible that the processing of the distractor hampers target concept selection. Costa et al. (2003) suggested that the PWI effect reflects the ease with which a concept is selected for lexicalization (semantic selection account), placing the locus of the observed effects before lexical access. Although these assertions are not couched in any single, coherent account, they will be discussed in the current review under the umbrella term of "concept exclusion hypothesis" (CEH). Crucially, even though, figuratively speaking, structural or conceptual representations may compete for selection, or may need to be blocked once wrongly selected by a cognitive system, the interference effects are traced to stages that do not directly involve lexical access and are a procedural artefact (the speaker finds it harder to recognize/identify a table as a table when it is accompanied by the distractor "chair" than when it is accompanied by the distractor "cloud," which causes an overall delay in production). Hence, they are not competitive in the narrow lexical sense of the word.

Scope and Organization of the Review

Only chronometric studies with unimpaired adult native speakers were included in the current review. Neuroscientific, neuropsychological, translation, and bilingual studies, as well as studies that used target stimuli other than pictures of objects were excluded. Detailed inclusion and exclusion criteria are provided in the Methodology section.

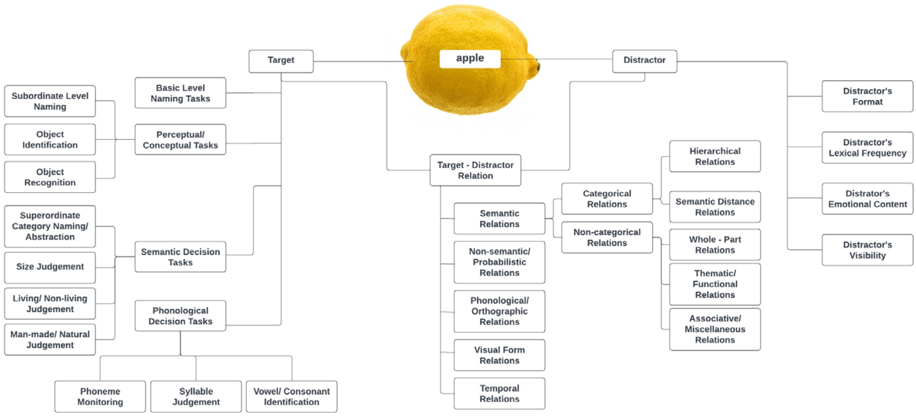
The review has been organized around the main sites of manipulation in the PWI task: the distractor, the task (or what one has to do with the target) and the target-distractor relation (see Figure 1). Although the time course of the PWI effect, or the temporal relation between target and distractor, has also been systematically varied and analyzed in previous studies, due to extensive coverage elsewhere (e.g., Bürki et al., 2020; Glaser & Düngelhoff, 1984; Schriefers et al., 1990), this was not included in the current review. Stimulus onset asynchrony (SOA) of 0 milliseconds (i.e., when target and distractor are presented simultaneously) serves as a default, but results are reported for other SOAs when relevant to the research question. The current review presents findings of studies in which a single variable (pure effect, e.g., distractor frequency) as well as pairs or even triplets of phenomena (joint effect, e.g., distractor frequency and semantic relatedness) were concurrently manipulated.

Terminology and Notation Used in the Review

A final note concerns terminology and notational conventions used in the current review. The word *target* refers to a stimulus to be named. The word *distractor* denotes a stimulus to be ignored. Occasionally, the word *distractor* is substituted by the more descriptive term *interfering stimulus*, which is found in the literature along with other alternative terms, such as *context word/picture*, *context object,* and *context stimulus*. Target-distractor relation or relatedness between targets and distractors is understood in terms of similarity between the two types of stimuli with regards to a specific feature, for example, visual overlap. The term *PWI effect* is used interchangeably with other terms mentioned in the introduction. This is dictated both by convention and by the lack of a more precise term. For although *context effect* seems to fit well in the context of PWI findings, it is also applied to other picture naming paradigms, such as semantic blocking and continuous naming tasks. The effect is quantified as a difference between the mean reaction time (RT) of the *unrelated condition* and the mean RT of the *related condition*; if the effect is positive (+), the distractor slows down naming and is said to produce *interference* or an *inhibitory effect*; when the effect is negative (-), the distractor speeds up naming, and thus results in *facilitation*. Stimulus onset asynchrony (SOA) refers to the point in time at which the distractor is presented relative to the target. If SOA is negative, for example, an SOA of -150 milliseconds, distractor presentation *precedes* the target onset by 150 milliseconds. If SOA is positive, for example, an SOA of +150 milliseconds, the distractor is displayed 150 milliseconds *after* the target onset; when SOA equals 0 milliseconds, target and distractor are presented simultaneously.

The terms *prelexical* and *postlexical* refer to the stages of processing at which semantic context effects may have their origin. An effect that occurs *prelexically* is one that arises before a lemma/lexeme is selected. It can result from a distractor interfering with the perceptual and/or conceptual encoding of the target picture (object recognition/identification) or it can be the outcome of an early attentional mechanism operating on the distractor itself, with some distractors being inherently harder to ignore than others. A semantic context effect that occurs postlexically, on the other hand, is one that can be traced to the lemma/lexeme postselection stage, such as when the selected lemma/lexeme is occupying the

Figure 1. *Main Sites of Manipulation in the Picture-Word Interference Task*



response buffer, awaiting articulation. Crucially, pre- and postlexical effects are PWI task artifacts that provide little or no information about the nature of lexical selection itself.

When providing stimulus examples from the reviewed studies, concepts are given in upper case, and distractor words in lower case italics. Target answers appear in inverted commas.

Methodology

Search Details

Three electronic databases (PsychINFO, PubMed, and Web of Science) were searched using the following combinations of keywords: "picture word interference" OR ("semantic interference" OR "context effect*" OR "semantic category" AND "picture naming") AND compet* NOT bilingual*. Additional records were identified by cross-checking the reference lists of core PWI articles to ensure all relevant papers were considered for review. The search covered the period up to July 2019. In total, 229 references were located by PsycINFO after the following filters were applied: peer-reviewed journal articles, adulthood (18 years and older), published in English, experimental studies. Web of Science generated 345 references, using the following filters: psychology experimental, articles, English, of which 184 were novel to PsycINFO. The search with PubMED generated 70 references (with English language as the sole filter), of which 22 were new. Nine additional articles were identified through manual search. For the detailed systematic review procedure, see the flowchart in Figure 2.

Study Selection

Two raters (MK and PD) independently assessed the eligibility of the identified titles and abstracts for review. Full-text articles were retrieved when either reviewer decided that the article was potentially eligible. Studies were considered for review if they fulfilled the following inclusion criteria: (a) the studies presented original research written in English (i.e., reviews, book chapters, conference proceedings, commentaries were not accepted), (b) the participants were adult native speakers with no history of cognitive or neurological impairment (studies conducted in idiosyncratic languages, such as Chinese or sign language were excluded), (c) the studies utilized the PWI paradigm to investigate the competitive nature of single word production (i.e., studies using other tasks, such as the Stroop task or word reading, as well as studies investigating syntactic or morphological processing, such as grammatical gender, number, verb, modifier processing, were excluded), (d) studies that were based on behavioral data (i.e., studies using neuroscientific methods, such as electroencephalography, neuroimaging, transcranial direct current stimulation, etc., were excluded), and (e) the studies used an experimental design with picture naming latency as the main dependent variable (i.e., studies based exclusively on correlational, distributional, or error analyses, or those incorporating computational models, were excluded). Any discrepancies between the raters were resolved through discussion and consensus.

Sixty articles met these criteria with a total of 117. Note that because each selected article presents (often contradictory) findings from across a series of experiments, for practical reasons, each experiment has been treated as an independent study. The number of selected articles therefore does not match the number of reported studies (experiments, abbreviated to Exps).

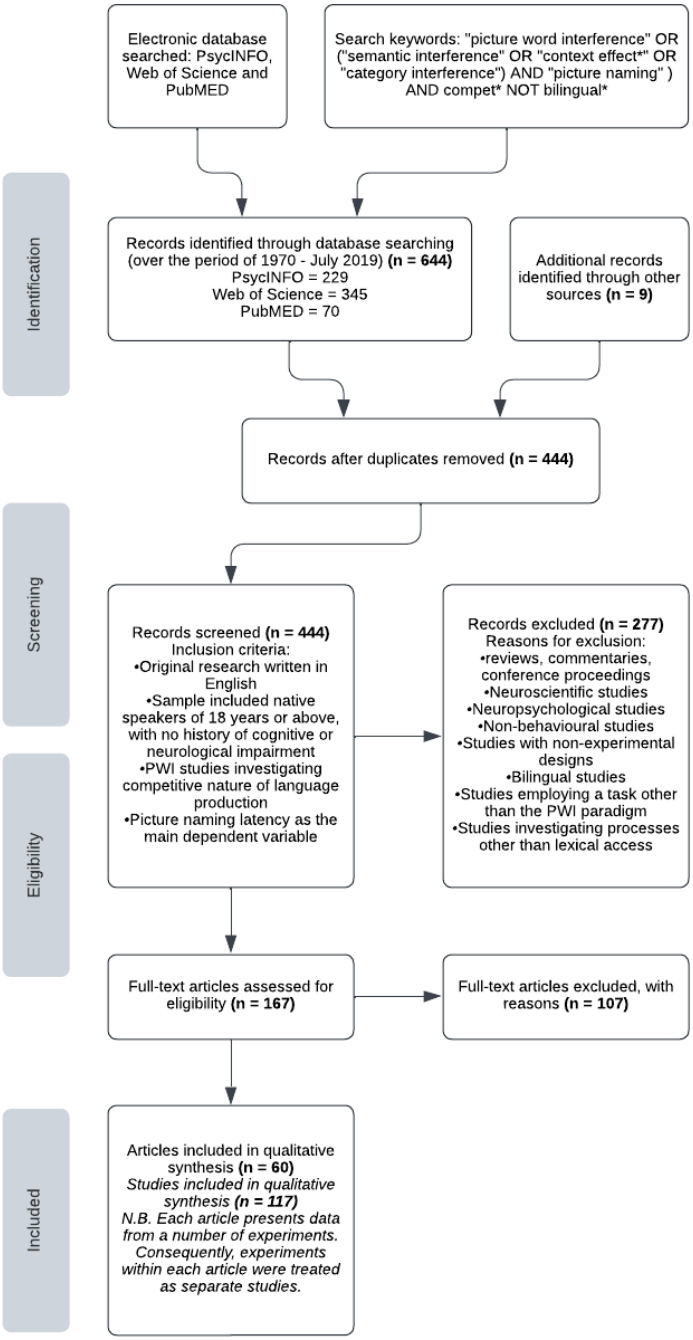
Data Extraction

The following data were extracted from the accepted articles: language in which the study was administered, number of participants, number of items, the task and the target answer, SOA, the findings, and the statistics reported.

Risk of Bias Assessment

Risk of bias in the included studies was assessed following the guidelines by the Cochrane Collaboration (Higgins & Green, 2011). Six domains were evaluated: sequence generation and allocation concealment (selection bias), blinding of participants and personnel (performance bias), blinding of outcome assessment (detection bias), incomplete outcome (data attrition bias), reporting bias (selective reporting), and other sources of bias (other bias). Since each domain addresses distinct issues which may be difficult to rate unambiguously, the risk of bias assessment was performed for individual entries within each domain.

Figure 2. *PRISMA Flowchart of the Systematic Review Procedure*



Furthermore, due to the specificity of psycholinguistic research, three separate entries were added to other sources of bias. These were selective reporting of demographic information, verification of response accuracy and timing, and matching of item sets across experimental conditions. The risk of bias assessment was conducted by five independent raters. The first (MK) rated all the included studies. The remaining four raters rated one quarter of the included studies each. A judgement of low, high, or unclear risk was made for each entry. Discrepancies in the ratings were resolved by discussion.

Results and Discussion

The following subsections present the rationale for the experimental manipulation of individual PWI task parameters, followed by the results of these manipulations. We specifically examined whether interference, facilitation, or no effect was observed. Each subsection concludes with a discussion of the results within both competitive and noncompetitive frameworks. The supplementary materials referred to in this section are available on the OSF page at https://osf.io/7qkh4.

Site of Manipulation: Distractor

Distractor Format

The primary rationale for substituting distractor words with nonverbal distractors was to challenge one of the assumptions of the response exclusion hypothesis (REH). Specifically, the REH posits that distractor words, due to their phonological well-formedness, reach the articulatory buffer more quickly than target picture names that only need to be phonologically encoded. The distractors must be removed from the buffer, based on some implicit and/or explicit response viability criteria, before the target picture name can be produced. It is reasonable to assume that if interference is observed with categorically related nonverbal distractors, which lack privileged access to the response buffer, this finding could be used to challenge the REH.

Twenty-one of the 117 reviewed PWI studies manipulated distractor format. Of these, 16 used pictures or photographs rather than words as distractors. These studies will hereafter be referred to as picture-picture interference (PPI) studies. Four of the 21 PPI studies utilized environmental sounds (Mädebach et al., 2017, Exps 1, 2, 3, and 4), and one used pseudo-words (Dhooge & Hartsuiker, 2012, Exp. 1) as distractors. See stimuli examples and results in the Supplementary Materials in Table A1.

One study of the 21 reviewed PPI studies reported facilitated picture naming in the presence of categorically related distractor pictures (La Heij et al., 2003, Exp. 2). In five of the 21 PPI studies, distractor pictures belonging to the same semantic category as target pictures showed no facilitation, but neither did they reliably interfere with picture naming (Bölte et al., 2015, Exps 1 and 2; Damian & Bowers, 2003; Geng et al., 2013, Exp. 2; Navarrete & Costa, 2005, Exp. 1B). Two further studies reported null results, but only under specific task conditions: (a) when distractors were not included in the response set and (b) when their position or sequential order was highly predictable (Mädebach et al., 2017; Exp. 4; Matushanskaya et al., 2016; Exp. 2). In 12 of the 21 studies with distractor format manipulation, categorically related distractors in the form of pictures or environmental sounds were found to interfere with picture naming (Aristei et al., 2012; Bölte et al., 2015, Exp 3; Dean et al., 2001; Humphreys et al., 1995, Exps 1 and 4B; Glaser & Glaser, 1989, Exp. 6; La Heij et al., 2003, Exp. 1; Mädebach et al., 2017, Exp. 1, 2 and 3; Matushanskaya et al., 2016, Exp. 1). In the one study (Dhooge & Hartsuiker, 2012, Exp. 1) that used pseudowords as distractors, prolonged naming was observed in the context of unrelated distractor words (e.g., ASHTRAY-flower) relative to unrelated pseudo-words (e.g., ASHTRAY-cromth).

Contrary to commonly held assumptions, a sizeable portion of studies demonstrated semantic interference with categorically related nonverbal distractor stimuli. By contrast, facilitation with such stimuli appears to be a rare, and so far, nonreplicable observation (reported only in La Heij et al., 2003, Exp. 2). These findings challenge the core prediction of the REH. Here, categorically related nonverbal distractors had no privileged access to the response buffer, and yet delayed naming was observed. What the studies reporting interference with nonverbal distractors appear to have in common is that their experimental conditions allowed for enhanced activation of distractors at the lexical level. In other words, once the distractor's lexical representation was adequately strengthened (possibly exceeding an activation threshold), the chances of it interfering with target naming increased. There are several ways in which to boost the lexical activation of nonverbal distractor stimuli. One is to include them in the response set (i.e., use them as targets). Indeed, studies in which the same stimuli routinely served as both targets and distractors (e.g., Dean et al., 2001, Exp. 1; Glaser & Glaser, 1989, Exp. 6; Humphreys et al., 1995, Exp. 1) or in which the response set membership was intentionally manipulated (e.g., Mädebach et al., 2017, Exp. 4), consistently reported interference. Naming of distractors when these appeared as targets in the task seemed to have primed their lexicalization even when they had to be ignored. Inclusion of multiple exemplars from the same semantic category as experimental stimuli, using small item sets and repeated distractor naming (e.g., Glaser & Glaser, 1989, Exp. 6) can equally contribute to spontaneous activation of the distractor's name.

Another factor that may inadvertently boost lexical activation of interfering stimuli is target uncertainty. In situations in which the signaling of the target is ambiguous, for example, the temporal succession of targets and distractors is so rapid that it makes target selection confusable for the speaker, lexical access may be initiated for all pictured concepts. Similarly, when task difficulty is high and cue onset prolonged (e.g., Humphreys et al., 1995), the speaker may strategically prepare verbal responses for both stimuli in advance of cue presentation. The speaker may also involuntarily lexicalize a distractor's concept when an expectation regarding either its spatial position (Jescheniak et al., 2014) or temporal order (Matushanskaya et al., 2016, Exp. 1) is first induced but then violated (but note the temporal head start given to distractor pictures in both studies, with SOAs of -67 ms and -200 ms, which may equally allow for an alternative, REH-compatible interpretation). Finally, overtly naming both targets and distractors, which was made possible when novel compound nouns were produced to refer to the displayed objects in a study by Aristei et al. (2012) ensured distractors got a fair share of lexical input.

Interference can thus be induced reliably by categorically related nonverbal distractor stimuli under experimental conditions that promote their lexicalization. These findings support the SLNH and lexical selection-by-competition with a competition threshold hypothesis suggesting that the interference effect can be traced back to the lexical processing stage. One cannot rule out the possibility, however, that the delay occurs partly or exclusively at a prelexical stage, when the concept activated by the distractor picture interferes with target picture recognition, identification, and/or the process of conceptual-lexical mapping, although the observation that the interference effect is not modulated by structural similarity between targets and distractors (Humphreys et al., 1995, Exp. 1) would argue against such an explanation.

The finding of no effect reported in a subset of studies could be interpreted as a trade-off between facilitation at the conceptual level (i.e., the distractor picture precipitates identification of the target concept through the spread of activation within the semantic network) and interference at the lexical level (i.e., the distractor's name enters into competition with the target's name, slowing down naming). The reason why no interference was obtained in the reviewed PPI studies despite evidence of lexical processing of distractors (e.g., Bölte et al., 2015, Exps 2 and 3) is that lexical activation in the PPI paradigm proceeds indirectly through conceptual encoding and may therefore be too weak to overshadow facilitation at the conceptual level. This idea is consistent with the SLNH and lexical selection with a competition threshold.

The only study that showed facilitation with semantically related picture distractors (La Heij et al., 2003, Exp. 2) has been called into question because categorical relatedness of the stimuli was confounded with associative relations. That associative (noncategorical) relations can reverse the polarity of the effect, leading to facilitation rather than interference, has been well documented using both the picture-word and picture-picture interference paradigm (e.g., Bölte et al., 2015, Exp.1, Costa et al., 2005; Geng et al., 2013, Exp. 1; Mahon et al., 2007). Moreover, although La Heij et al. (2003) appeared to have resolved the issue of target uncertainty (confusability about which picture to name) by reducing the duration of distractor presentation to 50 milliseconds while keeping target presentation duration at 300 milliseconds, the very brief distractor exposure could have allowed for only a limited amount of distractor processing. If the distractor was indeed conceptually encoded, but its lexical representation was not sufficiently activated, the procedure could have well led to quicker target identification (through semantic priming), and thereby faster naming responses, an idea consistent both with the REH and SLNH.

The only study that used pseudo-words as distractors (Dhooge & Hartsuiker, 2012, Exp 1) reported prolonged naming latencies in the context of unrelated distractor words (e.g., ASHTRAY-flower) relative to unrelated pseudo-words (e.g., ASHTRAY-cromth). Both competitive and noncompetitive accounts offer plausible explanations for this effect. Competitive accounts argue that legitimate distractor words accumulate more activation than nonwords, making them stronger competitors. Alternatively, the effect can also be seen as supporting the REH and self-monitoring accounts, which propose that pseudo-words (e.g., cromth) are dismissed as illegitimate responses more quickly than real words (e.g., flower), resulting in shorter naming times. Greater interference for legitimate words relative to pseudowords could also be explained by the CEH account, however, which places the locus of the effect at conceptual encoding. According to this explanation, legitimate words would supply semantic information which interacts with information about the target picture, adversely affecting its recognition and subsequently naming, unlike pseudowords, which hardly evoke any meaning, and are thus unlikely to interfere with the object recognition and/or identification process. Finally, the pseudoword distractor effect can also be attributed to an early attentional mechanism (Roelofs, 2003; Roelofs et al., 2011) that operates on the distractor before it can interfere with object identification or reach the articulatory buffer. Specifically, nonword distractors are identified as to-be-ignored stimuli more quickly than legitimate distractor words, which engage the speaker's attention for a longer duration.

Distractor Frequency

Of the 117 reviewed PWI studies, 13 manipulated the lexical frequency of distractor words. The rationale behind this manipulation is that words with higher activation levels - those that occur more frequently in language use - should be stronger competitors, or at the very least, not weaker than distractor words with lower activation levels. While unrelated high-frequency distractor words (e.g., APPLE-chair) were originally shown to induce greater interference than their low-frequency counterparts (e.g., APPLE-stool; Klein, 1964), the reverse pattern, namely, slower picture naming for unrelated low-frequency distractors, was reported in twelve separate studies (Catling et al., 2010; Dhooge & Hartsuiker, 2010; Geng et al., 2014; Hutson et al., 2013; Miozzo & Caramazza, 2003; Starreveld et al., 2013). One study reported no difference in picture naming response times between high- and low-frequency distractor conditions (Geng et al., 2014, Exp. 2). See Table A2 in the Supplementary Materials for results of distractor frequency manipulation.

Prolonged picture naming times, observed as the distractor frequency effect in the PWI task with infrequently encountered distractor words, align with the postlexical REH account. According to this account, high-frequency words enter the articulatory buffer sooner and can be dismissed as inappropriate responses more rapidly than low-frequency words (e.g., Miozzo & Caramazza, 2003; Finkbeiner & Caramazza, 2006; Mahon et al, 2007).

An alternative account proposed to accommodate the counterintuitive finding of slower picture naming in the presence of unrelated low-frequency distractors introduces a perceptual reactive blocking mechanism (Roelofs, 2003; Roelofs et al., 2011). This mechanism's role is to discern between relevant and irrelevant information at the input level. Because high-frequency words are recognized more quickly (e.g., Balota & Chumbley, 1984), the reactive blocking mechanism comes into play earlier, enabling a faster decision regarding which stimulus to name and which to ignore.

Greater interference with low-frequency distractors could also be explained by their attentional capture, which is compatible with the reactive blocking mechanism proposed by Roelofs (2003). The relatively infrequent use of low-frequency words and thereby their lower concept familiarity may make them "pop out," diverting resources away from the primary task of target picture naming. In the visual domain, low frequency distractors have indeed been shown to interfere more with visual search than their more frequent controls (e.g., Müller et al., 2009). Relatedly, concept retrieval for low-frequency words can be proportionally slower because their semantic nodes are thought to be more widely dispersed compared to more focused, densely connected nodes of high-frequency words (e.g., Steyvers & Tenenbaum, 2005). Semantic information from low-frequency distractor words may therefore be extracted later than that supplied by high-frequency words, delaying target object identification and consequently slowing down naming.

In one study in which no distractor frequency effect was observed (Geng et al., 2014, Exp. 2), the speed of target picture processing was experimentally reduced. This reduction was achieved by including only four target pictures in the response set and increasing the number of target repetitions to 20. The absence of interference for low-frequency distractor words under conditions of facilitated target naming aligns with the late, postlexical locus account of the effect. It is possible that with practice, the names of the few target pictures used in the experiment become so readily accessible that they reach the articulatory buffer before it can be blocked by distractor words. However, the data do not definitively rule out the role of an early attention mechanism because it is not certain whether target processing occurred swiftly enough to pre-empt the adverse effect of distractors at the input level.

Nine of the 13 PWI studies also included factorial combinations of distractor frequency with other variables, such as phonological relationship, semantic relationship, and distractor visibility. However, these cross-manipulations did not yield consistent results. In one study, the distractor frequency effect disappeared when targets and distractors were phonologically related (Miozzo & Caramazza, 2003, Exp. 7). The interaction between distractor frequency and phonological overlap suggests a common processing stage, namely, that the delay occurs at the level at which phonological information is accessed. Such a pattern of results favors a late rather than an early locus account of the effect. Dhooge and Hartsuiker (2010, Exp. 2) found no traces of interference from low-frequency distractor words under reduced visibility conditions. As the masking procedure is assumed to prevent phonologically well-formed distractor responses from entering the articulatory buffer, the absence of interference for masked low-frequency distractor words would indicate that distractor responses must indeed enter the post-lexical stage for the distractor frequency effect to be observed.

In three of the factorial studies, distractor frequency was cross-manipulated with stimulus onset asynchrony (SOA). The results were mixed. In two of these (Dhooge & Hartsuiker, 2010, Exp. 3; Starreveld et al., 2013, Exp. 1) the effect was absent at early SOAs, but got progressively larger with increasing distractor onset latencies. Direct comparisons and interpretation of data appear to be problematic, however, because one study stopped short of including late onset distractor latencies (SOAs > 0 ms, Dhooge & Hartsuiker, 2010, Exp. 3); while in the other (Starreveld et al., 2013, Exp. 1), the significant interaction between the SOA and distractor frequency was not followed up with simple effect analyses. To complicate the picture further, a contradictory finding of no interaction between SOA and distractor frequency was reported by Miozzo and Caramazza (2003, Exp. 5). Here, the distractor frequency effect was equivalent in size at all SOAs (-100, 0, and +150 ms). So, although the distractor frequency effect would seem to mimic the time course of the phonological facilitation effect, and could thus be said to have a post-lexical basis, more research is needed to resolve the existing discrepancies.

Other studies in which distractor frequency was factorially crossed either with semantic relatedness ($n = 2$) or with the task the participant was given to perform (n = 2) also produced conflicting results. In terms of semantic relatedness manipulation, low-frequency distractors produced interference of an equivalent size irrespective of whether they were semantically related or unrelated to the targets (Miozzo & Caramazza, 2003, Exp. 5), suggesting that the effects are driven by two separate processes, which would be at odds with the competitive account of lexical access. This is in contrast to a significant interaction between distractor frequency and semantic relatedness reported by Starreveld et al. (2013, Exp. 1). The distractor frequency effect was larger when distractors were categorically related than when they were unrelated to targets, a finding that suggests a shared processing mechanism in support of competitive lexical selection. In terms of task manipulation, even though the distractor frequency effect persisted when picture naming was replaced by manual phonological decision tasks (a finding which would argue against a postlexical locus), the results were clear-cut only for a syllable judgement task (Hutson et al., 2013, Exp. 1), but fell short of significance in a phoneme monitoring task (Hutson et al., 2013, Exp. 2).

Distractor Visibility

To better understand how and at what point in the process of naming a depicted object in the presence of a distracting stimulus semantic context effects come about, seven of the 117 reviewed PWI studies have manipulated distractor visibility (see Table A3 in the Supplementary Materials). This type of manipulation typically involves a masking procedure, in which the distractor is obscured by a forward (a string of symbols preceding the distractor, e.g., #####) and a backward mask (a string of symbols following the distractor, e.g., NGCFRLNHS). The distractor itself is displayed for a very brief duration, typically 53 ms. Most participants are unaware of subliminally presented primes as evidenced in subjective postexperimental reports, in which only a few declare seeing stimuli other than the target pictures themselves (e.g., Dhooge & Hartsuiker, 2010) as well as in more objective post-hoc visibility tests (e.g., Damian & Spalek, 2014), in which participants have to identify masked distractors on individual trials under the masking conditions that are identical to the ones in the experimental task.

The rationale for using masked distractors in the PWI paradigm is predicated on the assumption that the masking procedure effectively strips the stimuli of their lexical privilege to enter the articulatory buffer before target responses, while preserving their potential to affect earlier processing stages (Finkbeiner & Caramazza, 2006). Because distractors are no longer consciously perceived, they do not trigger the formulation of covert verbal responses, which would otherwise block the articulatory buffer and which would need to be removed to give way to target responses. Masking thus eliminates the need to engage response-exclusion processes, which should result in facilitation (a categorically related distractor which is conceptually, but not phonetically encoded may help to activate the target concept through a process of spreading activation) or at the very least produce a null result.

Facilitation was indeed observed for subliminally presented, categorically related distractors in three of the seven studies that manipulated distractor visibility (Finkbeiner & Caramazza, 2006, Exps 1 and 2; Dhooge & Hartsuiker, 2010, Exp. 2). This finding supports a postlexical explanation in line with the REH for the semantic context effect. However, two of the seven studies reported no difference in naming latencies between masked related and unrelated distractor conditions (Damian & Spalek, 2014; Piai et al., 2012, Exp.1), and one study found longer naming times in the related condition relative to the unrelated condition when distractors were masked and highly activated (Piai et al., 2012, Exp. 2).

The absence of an effect as well as the observation of interference can be interpreted in support of lexical selection with a competition threshold. Furthermore, the fact that an inhibitory effect was obtained (Piai et al., 2012, Exp. 2) when the lexical activation levels of masked distractors were experimentally increased, for example, by using many exemplars from a few semantic categories and including distractors in the response set, demonstrates that once sufficiently activated at the lexical level, masked distractors can interfere with the production process despite their "invisibility."

Special attention should be given to the joint manipulation of distractor visibility and frequency, on the one hand, and distractor visibility and emotional content, on the other. Both the distractor frequency effect (see the previous section) and the taboo interference effect (see the next section) have been ascribed a postlexical locus according to the REH. Here, the results diverge again. The distractor frequency effect disappeared when distractors were masked (Dhooge & Hartsuiker, 2010, Exp. 2), which is in accordance with the REH. The taboo interference effect, on the other hand, was preserved under masked distractor conditions (Hansen et al., 2017, Exp. 4). The lack of interaction between the emotional content of distractor words and distractor visibility would argue for two separate loci of the effects. If taboo interference is attributed to response exclusion processes operating post-lexically (e.g., Dhooge & Hartsuiker, 2010, 2012), then the persistence of the effect under masked condition would either question the validity of the masking procedure or suggest the effect has an earlier locus.

Emotional Content of Distractors

Thirteen of the 117 selected PWI studies manipulated the emotional content of distractor words. Taboo words (e.g., c*nt) were employed in thirteen studies, four of which additionally utilized nontaboo negative (e.g., maggot), and two of which also used positive (e.g., friend) verbal stimuli (see Table A4 in the Supplementary Materials). Manipulating the emotional content of distractors leads to different expectations from supporters of competitive and noncompetitive accounts of semantic context effects. In the context of taboo distractor words, REH and self-monitoring proponents would anticipate faster naming times because these words are expected to be intercepted more quickly by the self-monitoring mechanism and removed from the articulatory buffer sooner than neutral words. This is based on a social appropriateness criterion, according to which obscene words are deemed less plausible as responses in a task implicitly requiring socially appropriate labels.

However, when dealing with negative and positive distractor words, it becomes unclear which response relevance criterion would be applied by a monitoring system to remove emotionally charged stimuli from the articulatory buffer sooner than their neutral counterparts. In this case, both REH and self-monitoring theories would predict comparable naming times.

On the other hand, interference in the context of taboo, negative and/or positive distractor words aligns with competition-based activation threshold accounts because highly emotional words could theoretically accrue sufficient activation to be more potent competitors then neutral distractor words.

Of the 13 studies which manipulated the emotional content of distractors, 12 found a reliable taboo interference effect – picture naming latencies were substantially longer in the context of semantically unrelated obscene words (e.g., LEAF-c*nt) than when presented with their neutral counterparts (e.g., LEAF-vest, Dhooge & Hartsuiker, 2011, Exp. 2; Hansen et al., 2017, Exps 1, 2, 3, and 4; Mädebach et al., 2018, Exp. 1; White et al. , 2018, Exps 1 and 2; White et al., 2017, Exps 1 and 2; White et al. , 2016, Exps 1 and 2). The effect also generalized, although to a lesser degree, to negative distractor words (e.g., WEB-demon) in three of four studies (White et al., 2016, Exp. 2; White et al., 2018, Exps 1 and 2), but was eliminated when target pictures were accompanied by positive distractor words in two out of two studies (e.g., WEB-blossom, White et al., 2016, Exps 1 and 2).

That taboo and negative distractor words delay picture naming compared with neutral distractor words invites several interpretations. According to the competition threshold hypothesis (Piai et al., 2012), due to their highly arousing nature, emotional words are more likely to accumulate sufficient activation to compete with target names for selection. For the same reason, they may be harder to block by an early selective attention mechanism (Roelofs, 2003; Roelofs et al., 2013). This explanation is analogous to the attentional capture account, according to which the detection of arousing verbal stimuli involuntarily shifts attention away from the primary task (picture naming), slowing preparation of target response at either or both the prelexical and lexical processing stages. It is difficult to see how the interference effect for emotionally salient distractor words would reflect the operation of a post-lexical self-monitoring mechanism since taboo words should be less plausible as response than neutral words in the task in which the speaker names pictures of neutral objects, but see Dhooge and Hartsuiker's, (2011) interpretation, according to which the self-monitoring mechanism is performing more stringent checks when the articulatory buffer is occupied by socially inappropriate words, an operation which is time-consuming.

In two studies, discrepancies emerged in relation to the time course of the taboo interference effect, with the effect being present at early (SOA = -150 ms), simultaneous (SOA = 0 ms), and late distractor presentation (SOA = +150 ms) in one study (White et al., 2017, Exp. 2), but absent at a SOA of +150 ms in another (Hansen et al., 2018, Exp. 1). The persistence of the effect at a late distractor onset (150 milliseconds after target presentation) is difficult to reconcile with the post--lexical REH account, because it is unlikely that the distractor would occupy the output buffer given the target's temporal head start.

In one study which cross-manipulated emotional content of distractors with distractor visibility, the taboo interference effect persisted under masked distractor conditions (Hansen et al., 2017, Exp. 4). In two studies that also used a factorial design, the taboo interferences effect was attenuated, but not eliminated or reversed, as is typically found with semantically related distractors, under phonological overlap conditions (White et al., 2016, Exp. 1; Hansen et al., 2017, Exp. 3). While the former finding argues against a postlexical locus of the effect (insofar as the masking procedure effectively prevents the distractors from reaching the articulatory buffer), the latter allows for both pre- and postlexical interpretation. The overall picture is obscured again by contradictory findings. Taboo distractors that were phonologically related to target pictures' names (e.g., BIN-b*tch) were named faster than their unrelated counterparts (e.g., BIN-tw*t). The phonological facilitation effect was equivalent in magnitude for taboo and for neutral words in one study (Hansen et al. 2017, Exp. 3), but significantly larger for taboo words than for neutral (or negative and positive) words in another (White et al., 2016, Exp. 1). The absence of interaction between phonological relatedness and emotional valence of distractors (additivity of the effects) suggests that the effects arise at two different processing stages within the language production system, and so is in line with an early locus of the effect. Significant interaction, on the other hand, suggests a single (postlexical) source of the taboo interference effect.

Two of the 13 studies investigated how emotional valence of distractors affects phonological and semantic decision making (Mädebach et al., 2018, Exps 1 and 2). A taboo interference effect emerged when basic-level naming was replaced by a manual phoneme monitoring task (e.g., indicate whether the target's name starts with a *b* or a *k*) as well as by a size judgement task (e.g., indicate whether the displayed object is larger or smaller in real life than a shoe box), but only when the processing demand associated with target identification was high (degraded visual input). Since the emotional content of distractors had an adverse effect on decision times in both tasks, and neither of these were assumed to entail preparation of articulatory responses, it was concluded that the effect arose at the lexical processing stage and possibly also at the level of conceptual encoding, at least when the cognitive load associated with target recognition was high.

Site of Manipulation: Target-Distractor Relations

An important, if not the main site of manipulation in the PWI paradigm, is the relationship between the target and the distractor. There are several ways in which targets and distractors can relate to each other. Semantic relatedness is the most obvious relation, followed by nonsemantic (probabilistic), visual, phonological (orthographic), and temporal (stimulus onset asynchrony) relations. The current review covered the first three of these because manipulations of phonological overlap are not pertinent to the current research question, while manipulations of temporal relations have been widely discussed elsewhere (e.g., Glaser & Düngelhoff, 1984; Levelt, 1992; Schriefers et al., 1990).

Semantic Relations

PWI studies typically distinguish between categorical and noncategorical semantic relations. In the former group, targets and distractors are members of the same semantic category (e.g., vehicles) and are usually co-ordinates, or cohyponyms, (i.e., they are derived from the same level of specificity, e.g., CAR-train), but can also include hierarchical (hypernymic, e.g., CAR-vehicle versus hyponymic, e.g., CAR-Audi) and semantic distance relations (distant, e.g., HORSE-frog versus close, e.g., HORSE-buffalo). In the noncategorical group, the semantic relations are more heterogeneous and can include: whole-part (CAMEL-hump), thematic (BENCH-park), functional (BRUSH-paint), and other associative (miscellaneous) relations. Their definitions vary and individual relations may be difficult to disentangle.

Categorical Relations. In a standard PWI task, distractors are normally derived from the same level of abstraction (or specificity) as targets, and the stimuli are said to be co-ordinates (cohyponyms). Numerous studies have shown interference induced by categorically related co-ordinates and these constitute a backdrop against which other types of semantic relations are discussed. A group of studies have manipulated the hierarchical relations between targets and distractors, employing distractor stimuli drawn either from a superordinate or a subordinate level of representation. Their findings are presented in the first instance, followed by findings from a different set of studies with semantic distance manipulation.

Hierarchical Relations (Hypernymy and Hyponymy). In ten of the 117 reviewed PWI studies, participants named objects with their preferred basic-level names (e.g., a picture of a CAR as "car") while ignoring distractor words drawn from a different level of abstraction (or specificity) than targets. Following Rosch et al. (1976), Hantsch et al. (2005) defined the basic level as a level of specificity between a category label used to refer to a collection of objects (e.g., vehicle) and a label denoting a specific instance of a member of such a category (e.g., Audi). Of the 10 PWI studies with hierarchical relations manipulation, six used hyponyms, or subordinate-level names (e.g., FISH-shark), and four used hypernyms, or superordinate-level names (e.g., FISH-animal) as distractors. For results, see Table A5 in the Supplementary Materials.

Of the four PWI studies with hypernymic distractor names, two studies (Costa et al., 2003, Exps 3 and 4) reported interference for unrelated basic-level distractor names (e.g., BICYCLE-horse) relative to unrelated superordinate-level distractor names (e.g., BICYCLE-weapon). This within-level interference effect (i.e., distractor from within the same level of abstraction causing delay in production relative to distractors from a different level of abstraction than targets) was taken to reflect a decision process in which a relevant semantic representation is chosen for lexicalization (the "semantic selection" account). According to this account, when a speaker is asked to produce a basic-level name, the selection system considers available basic-level semantic representations as possible candidates for lexicalization, which delays naming. An alternative explanation to the within-level interference effect observed by Costa et al. would be the REH account, according to which both basic- and superordinate-level distractor words would occupy the articulatory buffer before the target name, but basic-level distractor words would be harder to reject as potential responses because they are more relevant to the task at hand (naming an object at the basic-level of abstraction). By contrast, two other studies with superordinate words as distractors (Kuipers et al., 2006, Exp. 1A; Roelofs, 1992) reported no effect of the level of specificity from which distractors are drawn. Contrary to Costa et al. (2003), naming times for target pictures with unrelated (Kuipers et al., 2006) and related (Roelofs, 1992) superordinate-level distractor names and for those with unrelated basic-level names were comparable. The source of these discrepancies is unknown.

In five of the six PWI studies utilizing hyponymic distractor names, semantically related subordinate-level distractors (e.g., FISH-shark) resulted in slower naming times than their unrelated counterparts (e.g., FISH-Barbie), a finding which was constrained to early rather than late SOAs (the effect disappeared at SOAs of +200 and +300 ms, Bölte et al., 2013, Exp. 2; Hantsch et al. 2005, Exps 1, 2 and 4; Hantsch et al., 2012, naming condition). This adds credibility to the claim that distractor words do not have to be drawn from the same level of specificity to interfere with picture naming. Interference for categorically related subordinate-level distractor names is in line with both competitive (selection-by-competition, SLNH) and noncompetitive (both REH and CEH) accounts. An effect in the opposite direction was reported in one of the six studies (Roelofs, 1992). When distractor words were presented before targets (SOA = -100 ms), semantically related subordinate-level distractors (CAR-jeep) induced facilitation relative to their unrelated counterparts (CAR-dagger). No clear effects were observed at SOAs of 0 and +100 milliseconds, which is in conflict with the vast majority of PWI studies utilizing subordinate-level names as distractors. The reason for this discrepancy is not fully understood.

Semantic Distance Relations. Semantic distance (also referred to as semantic gradient or semantic similarity) between targets and distractors is another aspect of the target-distractor relationship that has been systematically varied to establish the origin of the semantic context effect and thus elucidate the nature of lexical selection. Semantic distance is the degree of semantic overlap between a pair of items, which can be measured using a variety of methods. One way is to gather subjective semantic similarity ratings from subjects, who estimate the degree of relatedness between individual concepts (e.g., SPIDER-FLY, HOUSE--SWAN) using Likert-type scales. Another is to draw on published feature production norms (e.g., McRae et al., 2005; Vigliocco et al., 2004), which are derived from lists of attributes generated in response to a given concept (e.g. KNIFE "is sharp," "used for cutting," and "found in the kitchen"). Yet another is to employ more objective techniques, such as latent semantic analysis (LSA; Landauer et al., 1998), with semantic similarity scores derived from large corpora of text, and the normalized Google distance (NGD; Cilibrasi & Vitányi, 2007), with semantic similarity values based on the number of hits returned by the Google search engine for a given set of words. Understanding which method was applied to a given stimulus set is crucial because it can determine the polarity of the PWI effect.

Competitive and noncompetitive accounts would yield different predictions when manipulating semantic distance. According to competitive accounts, distractors closely related to targets should induce greater interference than distantly related distractors. This is because the former are more likely to spread activation to lexical representations that overlap with those of distractors. In contrast, proponents of the noncompetitive view would expect the opposite effect. From their perspective, not only would distractor words have an equal status in terms of response relevance, but also semantically close distractors would spread activation at the conceptual level, resulting in facilitation of target naming. Thirteen of the 117 reviewed PWI studies have directly manipulated the semantic distance between distractors and targets, producing contradictory findings (see Table A6 in the Supplementary Materials).

Of the 13, four studies (Mahon et al. 2007, Exps 5, 5b, 6, and 7) found that target-distractor pairs that were closely related (e.g., HORSE-zebra) interfered less with picture naming than distantly related pairs (e.g., HORSE-whale). The results were not always clear-cut, with discrepancies emerging within and across the experiments. Facilitation was observed when targets and distractors were presented simultaneously (SOA = 0 ms) in Experiments 5, 5b, and 6, but using the same SOA, null results were reported in Experiments 7 and 7b, with facilitation constrained to an early SOA (-160 ms) in Experiment 7. The method used (subjective semantic similarity ratings) allowed assessment of the effect of within-category semantic distance, but may have underestimated the role of distinctive features of items (e.g., stripes for zebra). For example, an examination of the semantically close target-distractor pairs (e.g., HORSE-zebra) in Experiment 5 revealed that the majority of them were characterized by distinguishing features, which could have driven the facilitatory effect. Other factors, such as proportion of related trials, may also have contributed to the ease with which pictures in the context of semantically close distractors were named. For example, when as many as half of the trials were related, the semantic distance effect emerged at the SOA of -160 milliseconds. Greater relatedness proportion increases the chances of strategy development, and so, presenting a semantically close distractor ahead of the target could have led to expectancy generation, facilitating naming. When the number of trials was reduced to 38% (Experiment 7b) and the relation between the stimuli became less predictable, no semantic distance effect was observed.

Four of the 13 studies (Hutson & Damian, 2014, Exps 1 and 2; Lupker, 1979, Exp. 3; Mahon et al., 2007, Exp. 7b) reported no effect of semantic distance. Here too, however, the studies were not free from methodological problems. In Lupker (1979), the matching of items on the psycholinguistic variables known to affect naming speed across the two semantic distance conditions, close or typical (e.g., BANANA-peach) and far or atypical (e.g., BANANA-lime), and the size of the item set (n = 10) were less than optimal. Hutson and Damian (2014) failed to replicate the facilitation effect for semantically close distractors reported by Mahon et al. (2007) and the semantic gradient effect (decreased interference with increasing semantic distance) reported by Vigliocco et al. (2004, Exp. 3). Removing the more problematic items from the analysis did not affect the null results. Although there was an indication of interference for semantically close items when the NGD method was applied to the data (Hutson & Damian, 2014, Exp. 2), the lack of correlation between semantic distance scores computed with this method and LSA scores and subjective ratings of semantic similarity undermines its construct validity.

The finding that picture naming is increasingly delayed with diminishing semantic distance between targets and distractors was reported in five of the thirteen studies (Aristei & Abdel Rahman, 2013; Rose et al., 2019; Vieth et al., 2014a, Exps 1 and 2; Vigliocco et al., 2004, Exp. 3). A significant linear trend was observed in Vigliocco et al. (2004, Exp. 3) at the SOA of -150 ms, with shorter naming latencies for smaller feature overlap. It is not certain, however, if the trend would have persisted had the unrelated ("far") distractor items been removed from the analysis. The design of the study also suffered from several flaws, for example, unequal repetition of distractors within (e.g., hatchet repeated three times in the "close" condition) and across semantic distance conditions, with some pairs being associatively (e.g., TROUSERS-belt) and phonologically related (e.g., BANANA-broom). Vieth et al. (2014a, Experiment 1) reported an inhibitory semantic distance effect, but only with an early distractor onset (SOA = -160 ms). Picture naming latencies decreased with diminishing conceptual feature overlap (Vieth et al., 2014a, Exp. 2). Semantic similarity between target and distractor was the most reliable predictor of picture naming latencies, when distractors were presented aurally (SOA = -100 ms), accounting not only for the effects of categorical relatedness, but also for the effects of response relevance (Aristei & Abdel Rahman, 2013). Rose et al. (2019) observed interference for within-category semantically close distractors at the SOA of 0 milliseconds despite a high relatedness proportion (67%) and with distractors included in the response set.

Facilitation for semantically close items is compatible with the REH account, which assumes larger conceptual priming for a greater degree of semantic overlap. Since both distractors are words (i.e., they enter the articulatory buffer ahead of the picture's name) and both are equally plausible as responses (the REH does not differentiate based on a semantic distance criterion, and so exclusion times for close and distant distractor names should be identical), no interference ensues, and the net result is facilitation. The finding that picture naming is slowed when the semantic distance between targets and distractors increases is problematic for competitive accounts as well as the CEH hypothesis.

The null results are hard to reconcile with the predictions of the REH account, which assumes a facilitatory effect - facilitation due to semantic priming between distractors and targets, which is greater in the case of semantically close items. The null results are similarly problematic for the selection-by-competition account because semantically close distractors are bound to be more strongly activated by the target picture than their semantically distant counterparts. The SLNH also predicts a net inhibitory effect for semantically close distractors because these tend to activate smaller semantic cohorts with greater resonance within the lexical network and stronger activation of distractor words.

Interference observed in the presence of distractors that are semantically close to targets is consistent with the assumptions of competitive accounts of lexical selection, according to which the effect of semantically related words on lexical competition is enhanced by the strength of activation spread between concepts as a function of their semantic feature overlap. The finding is problematic for non-competitive accounts, which predict greater facilitation for greater feature overlap. The inhibitory effect could also potentially originate during prelexical stages, in support of the CEH account. For example, in the case of a close semantic pair, (ANT-spider) as opposed to a distant semantic pair (ANT-beaver), deciding whether one sees an ANT or a SPIDER would take longer than recognizing the picture as an ANT with a structural or conceptual representation of BEAVER in one's mind. The fact that interference for semantically close pairs in the majority of studies was restricted to early SOAs would additionally argue for a prelexical (structural and/or conceptual) locus of the semantic distance effect.

Noncategorical Relations.

Miscellaneous Associative Relations. Associative target-distractor relations in PWI studies form a highly heterogeneous group and their operational definitions vary depending on source. For example, they may be understood in terms of the frequency of co-occurrence in language use (how often two words appear as close neighbors in written and/or spoken texts, e.g., Spence & Owens, 1990), or be determined through free association norms obtained from subjects' verbal associations generated to lexical cues (e.g., Alario et al., 2000; Sailor et al., 2009). Two associatively related items may also be defined as loosely belonging to the same semantic field but being derived from different semantic categories (e.g., Abdel Rahman & Melinger, 2007). Due to the lack of a unanimous definition, the relations will be discussed further under the umbrella term of miscellaneous associative relations, as distinct from associative relations that reflect whole-part relations (e.g., SHIP-anchor) and functional/thematic relations (e.g., DESERT-camel).

Competitive and noncompetitive accounts of semantic context effects predict different patterns of results. Noncompetitive accounts, like the REH, expect facilitation because both associatively related distractors and unrelated distractors would be equally implausible as responses, yet some level of conceptual priming would still occur. In contrast, competitive accounts would require additional assumptions. They suggest that associatively related distractor words might activate interfering lexical representations, but with lesser magnitude or lower probability compared to categorically related items. Of the 117 reviewed PWI studies, 18 examined whether miscellaneous associates exert different effects on picture naming than nonassociates, with mixed results (see Table A7 in the Supplementary Materials).

While there are practically no reports of interference induced by noncategorical associates (i.e., associatively related distractors from a different semantic category than targets, MOUSE-cheese), in five of the 18 studies, they were found neither to precipitate nor to delay picture naming (Alario et al., 2000, Exp. 1b; Bölte et al., 2015, Exp. 2; Cutting & Ferreira, 1999, Exp. 3A, non-homophone condition; Mahon et al., 2007, Exp. 6c; Lupker, 1979, Exp. 1).

The null results are difficult to reconcile with the REH account, which predicts facilitation through semantic priming (e.g., CARROT primes RABBIT) unless it is offset by interference arising from the confound of response relevance. Both the associated distractor (e.g., in RABBIT-carrot) and its non-associated counterpart (e.g., in RABBIT-station) are equally implausible as responses in the task of naming an animal (RABBIT). The response relevance confound is therefore eliminated and facilitation should thus remain the dominant force. The null results find a reasonable explanation in the SLNH, however. The latter assumes no interference if the lexical activation of the associated distractor is stronger than that of its nonassociated control, but not strong enough to outweigh facilitation at the conceptual level. In the case of associated distractors (RABBIT-carrot), the spread of activation is thought to be more widely dispersed, with relatively little recursion within the semantic and lexical networks, and, therefore, weaker lexical activation of interfering stimuli compared to categorical co-coordinates (RABBIT-horse). The failure to detect an effect can also be attributed to some procedural details. For example, associated distractors produced comparable effects to their unrelated controls but only under brief stimuli exposure (Alario et al., 2000. Exp. 1b; Bölte et al., 2015. Exp. 2). A facilitatory effect emerged in two other studies by the same authors (Alario et al., 2000. Exp. 2b; Bölte et al., 2015. Exp. 1) when the timing of stimuli exposure was prolonged. Methodological factors could also have played their part in one of the studies by Mahon et al. (2007, Exp. 6c), in which association norms were exceptionally obtained from a word association test rather than published free association norms. In three additional studies, associated distractors failed to induce any effects solely at later SOAs (0, +150, and +300 ms; Brooks et al., 2014, Exp. 1; Sailor et al., 2009, Exps 1 and 2).

In two of the 18 studies, adding associative strength to categorical relations conferred neither advantage nor disadvantage to picture naming (La Heij et al., 1990, Exp. 2; Lupker, 1979, Exp. 2). Naming reaction times in the strongly associated co-ordinate condition (e.g., HAND-foot) were equivalent to naming reaction times in the non-associated co-ordinate condition (e.g., HAND-ankle), indicating that the effect of associative relation was negligible. However, methodological flaws (e.g., inadequate matching of item sets, small item sets with repetitive naming of targets) and insufficient statistical power may have contributed to type II error.

In 12 of the 18 studies utilizing noncategorical associates as distractors, miscellaneous associative relations were shown to speed up naming relative to their unrelated controls, irrespective of their modality (visual or auditory), but primarily with early stimulus onset asynchronies (i.e., SOAs < 0 ms). Six of these reported faster picture naming for associated distractors when these were presented prior to target onset (Alario et al., 2000, Exp. 2b; Bölte et al., 2015, Exp. 3; Brooks et al., 2014, Exp. 1; La Heij et al., 1990, Exp. 2; Sailor et al., 2009, Exps 1 and 2). In the remaining six studies, facilitation was also observed when targets and distractors were presented simultaneously (SOA = 0 ms), however, this finding was restricted to analysis by subjects only in half of these studies (Sailor et al., 2009, Exp. 2; Bölte et al., 2015, Exp. 1; Brooks et al. 2014, Exp. 1; Damian & Spalek, 2014; Mahon et al., 2007, Exps 1, 2 and 2b).

Competitive and noncompetitive views have furnished their own interpretations of these results. Facilitatory effects for associatively related distractors are in line with the SLNH and are thought to reflect the trade-off between conceptual priming and lexical competition. The net outcome is facilitatory because noncategorical associates in the activated cohort are more dispersed than categorically related items resulting in relatively less recursion and weaker activation of the distractor at the lexical level. This is supported further by the changing polarity of the effect that is facilitatory for strongly associated co-ordinates (e.g., LEG-arm), but disappears for weakly associated co-ordinates (e.g., LEG-head) at early SOAs, and that is nonexistent for strongly associated co-coordinates at SOA of 0 ms, reappearing as interference for weakly related co-ordinates. Facilitation for associatively related distractors is also consistent with the REH account because both associated and nonassociated distractors fulfil the response relevance criterion to the same extent (both are nonviable responses), so they are cleared from the articulatory buffer at equivalent times, causing no delay. Facilitatory effects are neither at odds with the CEH account because structural information supplied by both types of distractors (associated and non-associated) should result in the same level of interference, which appears to be negligible, or at the very least insufficiently strong to cancel out facilitation due semantic priming.

Whole-Part Relations (Meronymy). As shown by the examples in the previous section, the materials in PWI studies with associative target-distractor relation manipulation typically contain a mixture of different types of associations. Ten of the 117 reviewed PWI studies have taken a more systematic approach to stimuli selection, examining the effects of solely whole-part relations (e.g., CAMEL--hump) on picture naming (Costa et al., 2005, Exps 1 and 2; Muehlhaus et al., 2013, Exps 1 and 2; Vieth et al., 2014b, Exps 1, 2 and 3, see Table A8 in the Supplementary Materials). Whole-part relations can be understood as relationships in which the target represents a whole, and the distractor its part. The latter is either a constituent of the object (e.g., FISH-gills; Costa et al., 2005), or denotes the material from which the object is made (e.g., CANDLE-wax; Muehlhaus et al., 2013). Whole-part relations can thus be said to include "has-a," "is-part-of," and "consists-of" relationships.

Facilitation for distractor words denoting parts was observed in five of the 10 PWI studies with whole-part relations manipulation. In Costa et al. (2005, Exps 1 and 2) and Muehlhaus et al. (2013, Exp. 1), this was the case when targets and distractors were presented simultaneously (SOA = 0 ms), while in Sailor and Brooks (2014, Exps 1 and 3), facilitatory effects were evident only at the SOA of -300 milliseconds, and at the SOA of -150 milliseconds when part terms were also strongly associated with targets (e.g., AMBULANCE-siren). Shorter response latencies reported by Costa et al. (2005) and Muehlhaus et al. (2013, Exp. 1) for distractors denoting parts could therefore be an epiphenomenon of strong association between the items. The distractor words in Muehlhaus et al. were indeed selected based on free association norms. While this was not explicitly stated in Costa et al. (2005), closer inspection of their materials suggests that distractors denoting parts were strong associates of their targets (e.g., PEN-ink; CHURCH-pew).

In five of the 10 studies with part-term distractors, whole-part relations did not reliably differ from their unrelated controls (Sailor & Brooks, 2014, Exps 1, 2 and 3; Vieth et al., 2014b, Exps 2 and 3). This was the case in all five studies when distractors were presented concurrently with the target (SOA = 0 ms) or after the target onset. No clear effects were registered at these SOAs, irrespective of whether the parts denoted by distractor words were visible in target pictures or not (visibility as a covariate in a post-hoc analysis by Sailor & Brooks, 2014, did not change the result), or of whether they denoted distinctive or nondistinctive parts of objects (e.g., AMBULANCE-dashboard, Sailor & Brooks, 2014, Exp. 3). An exception should be noted in Vieth et al. (2014b, Exp. 3) however, in which distractors denoting nondistinctive parts which were also visible in target pictures induced interference at the SOA of -150 milliseconds. Eliminating the association from target-distractor pairs by employing words denoting nondistinctive parts produced null effects and even led to polarity reversal in Sailor and Brooks (2014, Exp. 3).

Interference in the presence of whole-part relations was observed in three of the 10 studies (Sailor & Brooks, 2014, Exps 1 and 3; Vieth et al., 2014b, Exp. 3), but it only emerged for parts denoting nondistinctive features (e.g., DOG-nose) and was present either at the SOA of 0 milliseconds (Sailor & Brooks, 2014, Exps 1 and 3, in the latter, significant by items only) or at SOA of -150 milliseconds (Vieth et al., 2014b, Exp. 3, significant by subjects only). In Vieth et al. (2014b, Exp. 3) only nondistinctive (nonassociated) part terms that were also visible in target pictures induced interference when they were presented ahead of the target (SOA = -150 ms). This is different to Experiment 2, in which exactly the same materials were used, but the parts denoted by distractor words were not visible in the target picture.

The facilitatory effect found in PWI studies with whole-part relation manipulation invites two interpretations. The results reported in Costa et al. (2005), at least in Experiment 1, find a ready explanation in the REH account, which assumes default facilitation due to the automatic spread of activation at the conceptual level, unless it is counterbalanced by interference ascribed to the response relevance confound. The unrelated condition in Experiment 1 contained distractors which represented whole objects, in contrast to distractors in the related (whole-part) condition which included words denoting parts. According to the REH, in the task in which the implicit rule is to name a whole object, rejecting a distractor word denoting a part as a plausible response is easier and, therefore, less time-consuming than rejecting a distractor word denoting a whole. Even when both related and unrelated distractor words denote parts (e.g., Costa et al., 2005, Exp. 2), the REH can still account for faster picture naming in the context of part terms because these tend to prime targets, which in the absence of response relevance confound, should lead to facilitation. The results also fit well in the SLNH framework, which predicts a facilitatory effect if there is automatic spread of activation within the conceptual network and weak or nonexistent activation within the lexical network. In the example of BOTTLE-cork versus BOTTLE-gills, it is conceivable that the concept of a CORK will evoke the concept of a BOTTLE, whereas the concept of GILLS will not. The lexical node of cork may receive some activation from the concept of BOTTLE, which will, however, remain dominated by conceptual facilitation due to greater dispersion and limited recursion within the networks, resulting in faster naming.

The null results undermine the assumption of the REH because facilitation should be observed irrespective of whether a distractor is an associate or not. The findings of no effect could, however, be accommodated by the SLNH, particularly in the case of distractors denoting nondistinctive parts (e.g., AMULANCE--dashboard) as opposed to distinctive parts denoting unique features of objects (AMBULANCE-siren). AMBULANCE and DASHBOARD are likely to send activation to related concepts such as FIRE ENGINE and the superordinate category node (VEHICLES) which may all converge on the lexical activation of dashboard, making it a stronger lexical competitor capable of offsetting facilitation at the conceptual level.

Interference for distractors denoting target parts has two interpretations. An inhibitory effect for visible nondistinctive parts (e.g., CAMEL-knee) suggests little or no conceptual priming (KNEE is unlikely to activate CAMEL) in addition to interference which arises either due to prelexical processes (the visibility of KNEE in the picture creates temporary uncertainty as to what it is one has to name, in accordance with the CEH) or due to lexical competition (KNEE and CAMEL activate related concepts, e.g., HORSE, and a superordinate category node ANIMALS which are likely to converge on lexical activation of knee in accordance with the SLNH). Sailor and Brooks (2014) dismissed visibility of parts as a relevant factor, which undermines the CEH, leaving the competitive view unscathed.

Thematic Relations. Three of the 117 PWI studies have specifically investigated thematic relationships between targets and distractors (Abdel Rahman & Melinger, 2007, Exp. 3; de Zubicaray et al., 2013; Muehlhaus et al., 2013, Exp. 3, see Table A9 in the Supplementary Materials). Two items are said to be thematically related if they perform complementary roles in the same context or event (Estes et al., 2011). For example, DOG and LEASH are related by a walking theme. Estes et al. (2011) differentiate further between thematic relations that are spatial (e.g., DESERT-CAMEL), temporal (e.g., SUMMER-HOLIDAY), causal (e.g., WIND--EROSION), functional (e.g., HAMMER-NAIL), possessive (SURGEON-SCALPEL), and productive (COW-MILK). All three PWI studies manipulating thematic relations between targets and distractors reported faster picture naming in the context of related stimuli. A facilitatory effect for thematically related distractors was observed at the SOA of 0 milliseconds in one study (Muehlhaus et al., 2013, Exp. 1) and with an early distractor onset (SOA = -150 ms) in two other studies (Abdel Rahman & Melinger, 2007, Exp. 3; de Zubicaray et al, 2013).

The findings do not discriminate between the rival accounts of lexical access, however, especially when association strength is involved. In both de Zubicaray et al. (2013) and Muehlhaus et al. (2013), thematically related pairs were also associated as confirmed by published association norms. Although this was not explicitly stated in Abdel Rahman and Melinger (2007), it is not hard to imagine that association strength (e.g., FRENCHMAN-beret) could have been driving the reported effect. Facilitation is predicted by the REH on condition that both related and unrelated distractors fulfil the response relevance criterion to the same extent, which they do in the case of thematically related distractors, that is, the system does not differentiate between PARK and JUNGLE when the task is naming a picture of a bench. Facilitation is also the predicted net outcome of the SLNH. Because the distractor and target share few, if any semantic features (BENCH and PARK do not share internal features that LION and TIGER do), and activate different superordinate category nodes (i.e., furniture versus outdoor places), the spread of activation is more diffuse, and therefore, lexical activation of PARK may be insufficient to outweigh strong facilitation at the conceptual level. The CEH can also accommodate the reported data because structural and/or conceptual representations of PARK and JUNGLE should be equally easy/hard to exclude when one is looking at a picture of a long metal seat for several people.

Probabilistic Relations

As indicated in the previous section, whether facilitation or interference is observed in the PWI task may be strongly dependent on the association strength between targets and distractors, and therefore, the strength of activation within both conceptual and lexical networks. Associative relations are confounded by semantic relations, so measuring effects of pure association or the probability with which two items occur together, or two representations are coactivated, is not an easy task. The effect of purely associative relations on picture naming is an underexplored topic in the area of PWI research (no PWI study has employed distractors that would be purely associatively related to targets), but one that could provide additional insight into the processes underlying lexical selection. Research could utilize opaque compound nouns (e.g., HONEY-moon) or expressions that have entered parlance recently (e.g., FACE-book), which would potentially allow one to test for the effects of association strength independent of semantic relations as well as the effect of association directionality.

Visual Similarity

The role of visual form overlap between targets and distractors in the PWI paradigm has until recently received little attention, with only four of the 117 PWI studies directly manipulating this type of relationship (de Zubicaray et al., 2018, Exps 1 and 2; Humphreys et al., 1995, Exp. 1; Mahon et al., 2007, Exp. 6b, see Table A10 in the Supplementary Materials). Observing interference with semantically unrelated distractors that share visual features with targets would suggest that at least part of the semantic context effect is due to this confounding factor. This, in turn, could pose a problem for competitive accounts that place the locus of the effect at the lexicalization stage.

Visual similarity between pairs of items (e.g., ORANGE-ball) was determined either by subjective similarity ratings (de Zubicaray et al., 2018; Mahon et al., 2007) or measures of partonomic features and the degree of overlap between the outline contours of size-normalized drawings of objects (Humphreys et al., 1995). Two of the four studies (de Zubicaray et al., 2018, Exp. 1; Humphreys et al., 1995, Exp. 1) demonstrated a form-related interference effect in the absence of sematic relatedness, and one (Mahon et al., 2007) observed an effect in the same direction which approached significance by subjects but was nonsignificant by items. In the only two studies in which interference was reported, the same stimuli served as targets and distractors (i.e., they were members of the response set), which introduces a potential confound of covert lexicalization. The distractors that were named on trials in which they functioned as targets may have inadvertently been lexicalized on trials in which they were to be ignored. In addition, as discussed in the Distractor Format section, in the PPI task used by Humphreys et al. (1995), participants may have strategically prepared verbal responses for both pictures so an appropriate name could quickly be retrieved upon cue onset. In effect, the delay observed in the context of similar looking items could be related either to temporary difficulty in object recognition/identification, implicating a prelexical locus of interference, in accordance with the CEH, or lexical competition due to a covert lexicalization confound. Indeed, when the response set membership was manipulated (de Zubicaray et al., 2018, Exp. 2), visually similar distractors no longer exerted an effect on picture naming.

Site of Manipulation Task

The question of whether interference in the PWI task would persist if the cognitive load associated with a particular production stage (e.g., lexical selection) were to be reduced or shifted to a different processing stage (e.g., object recognition) has spurred some researchers into manipulating the task that the participant is given to perform. In 14 of the 117 reviewed PWI studies, task instructions were changed from basic-level naming to a nonnaming perceptual, semantic or phonological decision task. Additionally, in nine of the 117 selected studies, basic-level naming was replaced with subordinate-level naming. Studies in which participants were required to name the pictured object with a subordinate-level name were included in this section because of the additional perceptual processing load associated with this task.

Perceptual-Conceptual Decision Tasks

Perceptual tasks are primarily concerned with visual processing of objects and may involve object detection, discrimination, recognition and identification. These tasks allow researchers to isolate and investigate specific cognitive processes, such as visual processing, without the interference introduced by the articulatory and phonological processes involved in naming. By bypassing the lexicalization stage and postlexical processes, supporters of the competitive accounts would expect no interference. If interference were observed in a task requiring perceptual encoding, this would speak to the role of early processes in the semantic context effect.

Of the 117 reviewed PWI studies with task instructions manipulation, three utilized tasks requiring some degree of perceptual analysis, although performance on these tasks may also hinge on semantic processing and/or the process of integrating perceptual and conceptual information (see Table A11 in the Supplementary Materials). Of the three, two studies reported interference for semantically related distractors without the apparent involvement of lexical processes (Dean et al., 2001, Exp. 2; Lupker & Katz, 1981, Exp. 1) and one found no effect (Schriefers et al., 1990, Exp. 3). While interference in the color and object recognition tasks (Dean et al., 2001, and Lupker and Katz, 1981, respectively) was interpreted in favor of the prelexical locus of the PWI effect, its absence in the object recognition memory task (Schriefers et al., 1990) was argued to support a lexical basis. The conclusions in all three studies may, however, be premature. It is uncertain whether the interference effect observed by Dean et al. (2001) and Lupker and Katz (1981) was genuinely due to a nonlexical process or an outcome of strategic covert lexicalization. It is not clear either how much "visual" and how much "semantic" processing was involved in the task selected by Schriefers et al. (1990). If participants based their recognition decisions largely, or exclusively, on stored structural representations, responding "yes" when the same perceptual codes were activated at study and at test and "no" otherwise, without consulting semantic information, then the absence of interference reported by the authors does not preclude a semantic basis of the effect.

Subordinate-Level Naming

The vast majority of PWI studies required participants to name the depicted objects at the basic level of abstraction (e.g., TROUT as "fish"). In nine of the 117 reviewed studies, however, the task was replaced by subordinate-level naming. Findings from PWI studies in which participants were instructed to use a specific (as opposed to a general) name for the depicted object (i.e., TROUT as "trout") are reviewed in this section because subordinate-level naming is associated with higher perceptual demands (see Table A12 in the Supplementary Materials). Chronometric research has shown that objects are identified and named more slowly at the subordinate level than at the basic level of abstraction, even when controlling for potential lexical confounds, such as frequency of occurrence or word length of target names (e.g., Jolicoeur et al., 1984; Lin et al., 1997).

Two of the nine subordinate-level naming studies (Vitkovitch & Tyrrell, 1999, Exps 1 and 3) observed interference for related subordinate-level distractors (distractors representing the same semantic category and belonging to the same level of abstraction as targets) relative to their unrelated controls. Naming a picture of a MINI as "mini" was slower when it was accompanied by a semantically related distractor (e.g., jaguar) than when it was presented with an unrelated word (e.g., tulip). Such within-level interference is compatible with both competitive and noncompetitive views of lexical selection. According to the lexical-selection-by-competition view, the target picture of MINI is likely to activate the superordinate category node of CAR as well as related exemplar nodes (e.g., JAGUAR, AUDI), which in turn activate their corresponding lexical nodes (i.e., jaguar, audi), rendering the related subordinate distractor (jaguar) a stronger competitor than an unrelated distractor (tulip). The CEH would entail greater difficulty for concept selection in the context of related subordinate-level distractors than unrelated ones, not so much because the cognitive system has access to information regarding the abstraction level of the distractors, as implied by Costa et al. (2005), which in this case is equivalent for both types of stimuli, but because the semantic and possibly structural information activated by the related distractor creates greater confusion about the target's identity than the information extracted from the unrelated distractor, thereby prolonging the time needed to select the correct concept for lexicalization. The results can also be accommodated by the REH account because although both distractors should reach the articulatory output buffer at the same time, the related distractor (jaguar) is a more plausible response (satisfying some general response-relevant criteria, such as naming a car) and is therefore harder to exclude from the buffer than an unrelated word (tulip), also prolonging naming.

Seven of the nine subordinate-level naming studies used distractors denoting either the target object's basic-level name (identical basic distractors, e.g., MINI-car) or the target object's semantically related basic-level name (related basic distractors, e.g., MINI-train) versus their unrelated controls, with discrepant results reported both within- and across laboratories. Of the seven, two studies (Hantsch et al., 2009, Exp. 4; Vitkovitch & Tyrrell, 1999, Exp. 2) found between-level facilitation for identical basic-level distractors versus their unrelated controls. Naming a MINI as "mini" was faster when it was accompanied by a distractor word denoting the picture's basic-level name (i.e., car) than when it was presented with an unrelated basic-level distractor (e.g., flower). Facilitation for identical basic-level distractors relative to their unrelated controls cannot be explained by lexical competition alone because an opposite pattern of results would be expected. Similarly, the results appear to be in conflict with the REH, the predictions of which would entail either null results (due to conceptual priming and interference at the response level) or prolonged naming latencies for related basic-level distractors (e.g., the distractor car is more response relevant because the speaker must identify and name a specific type of car) which would be harder to exclude from the articulatory buffer than unrelated controls. The results are also problematic for the SLNH, which would predict an inhibitory net outcome, with facilitation at the conceptual level being outweighed by interference at the lexical level (the target MINI presumably activates the concept of a CAR, which in turn, activates its lexical node car making the distractor more competitive than its unrelated counterpart). There are at least two alternative accounts that are compatible with the facilitatory effect obtained for identical basic-level distractors, however. One is the CEH account – the word car activates a corresponding concept and possibly a structural representation of CAR. The semantic (and pictorial) features of CAR converge onto the semantic and structural representation of MINI, making both processes of object recognition and target concept selection easier relative to unrelated basic-level distractors. Two, facilitation is due to the confound of message congruency (Kuipers et al., 2006). Conceptually, CAR is not incongruent with the target response. CAR leads to the same response as MINI because the latter must first be identified as a car. This explanation is supported by Hantsch et al. (2009, Exp. 4), showing that identical basic-level distractors lead to facilitation in subordinate-level naming if the proportion of response-congruent trials is high and if only one exemplar per basic-level category is used. Facilitation can thus be understood as a net effect of two opposing forces – inhibition either due to lexical competition (SLNH) or the confound of response relevance (REH) and facilitation due to the confound of message congruency – in this case the latter wins.

In four of the seven subordinate-level naming studies (Hantsch et al., 2005, Exps 3 and 4; Hantsch et al., 2009, Exps 1 and 2), the reverse pattern of results (that of interference) for identical basic-level distractors was reported, however. Here, identical basic-level distractors interfered with subordinate naming relative to unrelated distractors. This observation applied to both visually (Hantsch et al., 2005, Exps 3 and 5) and auditorily presented distractors (Hantsch et al., 2009, Exps 1 and 2) as well as a range of SOAs (from -100 ms to +300 ms). Interference was taken as evidence for the notion that basic-level names become lexically activated during subordinate-level naming and that these basic-level names compete for selection with the subordinate-level target words, in accordance with the SLNH. The REH explanation was given a similar amount of credit by the authors.

Although Hantsch et al. (2009) suggest a number of factors that may have led to discrepant results (facilitation in Vitkovitch & Tyrrell, 1999, Exp. 2, and Hantsch et al., 2005, Exp. 4, and interference in Hantsch et al., 2009), such as distractor modality or the amount of pictorial information in the target picture, the authors fail to mention an important aspect of the experimental set-up, which may have effectively altered the nature of the task. Leaving statistical significance aside (e.g., in Hantsch et al., 2005, 2009, some comparisons would not have survived post-hoc corrections and some tests produced significant differences only by subjects), one major procedural difference between the studies by Vitkovitch and Tyrrell (1999, Exp. 2) and Hantsch et al. (2005, 2009) concerns the familiarization phase. While participants in Hantsch et al. (2005, 2009) were extensively trained on subordinate-level names of the experimental stimuli, in Vitkovitch and Tyrrell (1999), they were only familiarized with the task structure, without being pre-exposed to the experimental materials themselves. It is therefore reasonable to assume that in Vitkovitch and Tyrrell, the task was more likely to involve additional perceptual processing (with the focus on recognition and identification), whereas in Hantsch et al. (2005, 2009) because of participants' familiarity with the target pictures, perceptual load was significantly reduced, if not eliminated, with greater demand placed on correct name retrieval.

Two of the seven studies reported null results with subordinate-level naming (Vitkovitch & Tyrrell, 1999, Exp. 2, condition with alternative basic-level distractors, Hantsch et al., 2009, Exp. 3). Null findings find a ready explanation in at least three rival accounts. The effect could reflect an interplay of a facilitatory component due to semantic priming (e.g., in MINI-train, TRAIN primes CAR which primes MINI), on the one hand, and an interfering component, on the other. The latter may stem from either lexical competition (e.g., convergent activation on the lexical node of train), in accordance with the SLNH, interference at the pre-conceptual level (i.e., perceptual/semantic disambiguation), or interference at the post-lexical level (the REH account).

Semantic Decision Tasks

Eleven of the 117 PWI studies employed a semantic decision task in place of basic-level naming (see Table A13 in the Supplementary Materials). Semantic decision tasks make overt demands on retrieval of semantic knowledge and in the PWI studies reviewed in the current work have ranged from superordinate classification (i.e., making a binary decision as to whether an object belongs to a specific superordinate category), superordinate naming (i.e., naming the target's higher-level category), through size judgement (i.e., deciding whether the depicted object is larger or smaller in real life than a predefined object), to living/nonliving or natural/manmade classification. As was the case with perceptual-conceptual decision tasks, the rationale for using semantic decision tasks was to bypass the lexicalization and post-lexical stages. If interference were to be observed in these tasks, this would suggest that the semantic context effect is at least in part due to task artifact.

Of the four PWI studies employing nonabstraction tasks, one found facilitation for semantically related distractor pictures (Damian & Bowers, 2003), and one reported null results (Humphreys et al., 1995, Exp. 3B). In one, size judgement latencies were comparable across taboo and unrelated neutral distractor conditions (Mädebach et al., 2018, Exp. 1), and in another, taboo distractors interfered more with the nonlexical size decision performance than their neutral controls, but only when targets were visually degraded (Mädebach et al., 2018, Exp. 2). Although the absence of interference in a task that does not specifically require lexicalization could be interpreted as support for the lexical locus of the PWI effect, lack of adequate controls undermines the validity of at least two findings. The results in Damian and Bowers (2003) could have been unduly influenced by message congruency, with semantically related pairs (e.g., SHIRT-skirt) in the manmade/natural decision task being always response congruent (both the target and distractor are classified as man-made), and unrelated pairs (e.g., SHIRT-banana) being at least sometimes response incongruent (the target is classified as man-made but the distractor as natural, which leads to conflict at the response output level). The same problem concerns Humphreys et al. (1995), in which covert verbalization (due to task difficulty) could have resulted in interference, whereas response congruency during living/nonliving decision making may have led to facilitation, with the two cancelling each other out. The results reported by Mädebach et al. (2018) demonstrate that interference in the PWI task is possible without overt lexicalization and that it may arise relatively early in the process of spoken word production, when unwanted representations (such as emotionally charged words) divert attention away from semantic processing, consequently prolonging naming response times.

By far the most numerous group of studies employing semantic decision tasks (seven out of nine) utilized a form of an abstraction task (Costa et al., 2003, Exp. 1; Glaser & Düngelhoff, 1984, Exp. 2; Glaser & Glaser, 1989, Exp. 6; Hantsch et al. 2012; Humphreys et al. 1995, Exp. 3A; Lupker & Katz, 1981, Exp. 2; Smith & Magee, 1980). In four of these, basic-level naming was replaced by superordinate category naming (e.g., naming a picture of a dog as "animal"). In two studies, subjects were instructed to name target objects with a higher category name (i.e., naming a picture of a poodle as "dog"). One study required subjects to make a binary decision as to whether a depicted object belonged to a specific superordinate category (e.g., is DOG an animal?).

Four of the seven PWI studies employing an abstraction task reported facilitation when target pictures had to be assigned to a higher-level category, namely, naming a picture of a poodle as "dog" in the context of semantically related distractors (i.e., dachshund) relative to when they were paired with unrelated (i.e., tulip) distractors (Costa et al., 2003; Glaser & Düngelhoff, 1984, Exp. 2; Glaser & Glaser, 1989, Exp. 6; Hantsch et al., 2012). The effect emerged both for distractors denoting the target picture's name (identical distractor words, e.g., POODLE-poodle), and for those derived from the same semantic category as targets (alternative distractor words, e.g., POODLE-dachshund), although some discrepancies were observed for different SOAs. For example, while in Glaser and Glaser (1989, Exp. 6), facilitation was observed for all SOAs spanning a range of -300 ms to +75 ms, in Glaser and Düngelhoff (1984; Exp. 2), reliable facilitation was only found with long pre-exposure times (i.e., SOAs -400 ms and -300 ms). This facilitation effect has at least two sources. One, even if the name dachshund competes lexically with the correct response "dog" when one is categorizing and naming a picture of a POODLE, this interference may be overshadowed by strong facilitation due to semantic priming. The distractor dachshund is likely to activate the superordinate category node ANIMAL. Similarly, the structural information supplied by the distractor does not conflict with the information extracted from the target picture in a task in which participants need to classify the object to a higher-level category.

At the response level, both distractor (e.g., dachshund) and target (e.g., poodle) lead to the same response (i.e., "dog"), which again speeds up categorization relative to the unrelated condition, in which the target stimuli (poodle versus tulip) map on to divergent response codes, that is, "dog" and "flower." Thus, in the unrelated condition, the message incongruency likely causes a delay.

Three of the seven studies showed no difference in picture categorization between related and unrelated conditions (Humphreys et al., 1995, Exp. 3A; Lupker & Katz, 1981, Exp. 2; Smith & Magee, 1980, Exp. 1). However, also here the studies do not remain without criticism. For example, in the PPI task used by Humphreys et al. (1995), the decision to categorize the color-cued target picture is preceded by at least two processes – a decision of whether to name or to categorize (i.e., name when red, categorize when green), and a recollection of which picture was displayed in which color. Due to high processing demands, a strategy may be adopted according to which the green picture is covertly labelled with a category name while the red picture is covertly labelled with a basic-level name in anticipation of the cue onset. Smith and Magee (1980) provide no descriptive or inferential statistics, simply claiming that words "[…] do not cause nearly as much interference when the task is changed to picture categorization" (p. 379-380). However, graphically, the targets in the 100% congruent condition, in which all items were semantically related (e.g., SHOE-dress), appeared to be classified faster than those in the 0% congruent condition (SHOE-frog). Finally, in Lupker and Katz (1981), the related distractors denoting target names (CAR-car) as well as those denoting names of some other exemplars from the same semantic category (CAR-train) showed facilitation, which nevertheless fell short of statistical significance.

Phonological Decision Tasks

Of the 117 studies reviewed, four employed a phonological decision task in place of basic-level naming. Phonological decision tasks used in combination with the PWI paradigm have included phoneme monitoring (or phoneme detection), syllable judgement and vowel/consonant identification. In the phoneme monitoring task, a target picture's name is mentally scanned for the presence of a particular phoneme. The latter can be predefined via task instructions (e.g., indicate whether the target picture name begins with a /b/ or a /k/) or specified on a trial-by-trial basis (e.g., indicate whether the target picture's name contains the phoneme seen in a previous trial). The syllable judgement task involves making a decision about the number of syllables in the target picture's name (e.g., is ANCHOR mono- or disyllabic?). In the vowel/consonant identification task, participants make a judgement as to whether the final segment of a target picture's name is a vowel or a consonant. There is compelling evidence that tasks which require a decision based on either segmental information (i.e., individual segments, their order in a word) or metrical information (i.e., number of syllables, stress patterns) of the target picture's name involve conceptual, lexical and morpho-phonological access, but do not engage articulatory processes (e.g., Oppenheim & Dell, 2008; Wheeldon & Levelt, 1995). The rationale for using this type of manipulation was that if the PWI effect survives the elimination of articulatory preparation, this would suggest that the effect has a prearticulatory basis.

Of the 117 studies reviewed, four employed a phonological decision task in place of basic-level naming (Abdel Rahman & Aristei, 2010; Hutson et al., 2013, Exps 1 and 2; Mädebach et al., 2018, Exp. 1). See results in the Supplementary Materials in Table A14.

In the vowel/consonant identification task (Abdel Rahman & Aristei, 2010), participants took longer to manually classify the final letter of the target picture's name in the presence of categorically related distractors than when targets were accompanied by unrelated distractors. The persistence of the semantic interference effect in the absence of overt articulation was taken as evidence against the REH account, according to which interference should only be obtained in tasks in which the articulatory output buffer is occupied by a production-ready representation which must be cleared for the target name to be produced. Also at odds with the REH account is the finding by Hutson et al. (2013, Exps 1 and 2), who reported no interaction between task (naming versus phonological decision) and distractor frequency, which together with the fact that the phonological decision task involved manual responses, indicates a locus outside the articulatory output buffer (but note a marginally significant effect in Exp. 2). Similarly, a delay in phoneme detection was reported for taboo distractor words relative to their neutral counterparts (Mädebach et al., 2018, Exp. 1). Since the manual phoneme detection task assumes no preparation of articulatory codes, the locus is again placed before the articulatory output buffer. In addition, the taboo interference effect appeared to be attenuated in the phoneme detection task relative to the naming task, which would suggest that the two are underpinned by the same process. This, however, cannot be confirmed without an interaction between task (naming versus phonological decision) and emotional content of distractors being subjected to a statistical analysis.

Risk of Bias Assessment Results

This section presents the results of the risk of bias assessment, conducted to evaluate the quality and methodological rigor of the reviewed studies. The risk of bias was assessed for each of the ten domains: sequence generation, allocation concealment, blinding of participants, blinding of the experimenter, blinding of the outcome assessor, incomplete outcome data, selective reporting, demographic information, response accuracy and timing, and matching of item sets. Judgments were categorized as either low, high, or unclear based on the criteria applied to individual domains, as specified in the subsections below.

The risk of bias graph (see Figure 3) and the risk of bias summary table (see Table A15 in the Supplementary Materials) provide an overview of the assessments for each of the ten domains, revealing potential sources of bias across the selected studies. Drawing from the results of the risk of bias assessments, a set of recommendations regarding the PWI methodology was formulated. These recommendations are presented in the Recommendations About the PWI Methodology section.

In the Summary of Findings, Future Directions, and Main Summary and Conclusions sections, we summarize the evidence synthesized in this review, emphasizing gaps, inconsistencies, and suggestions for future studies.

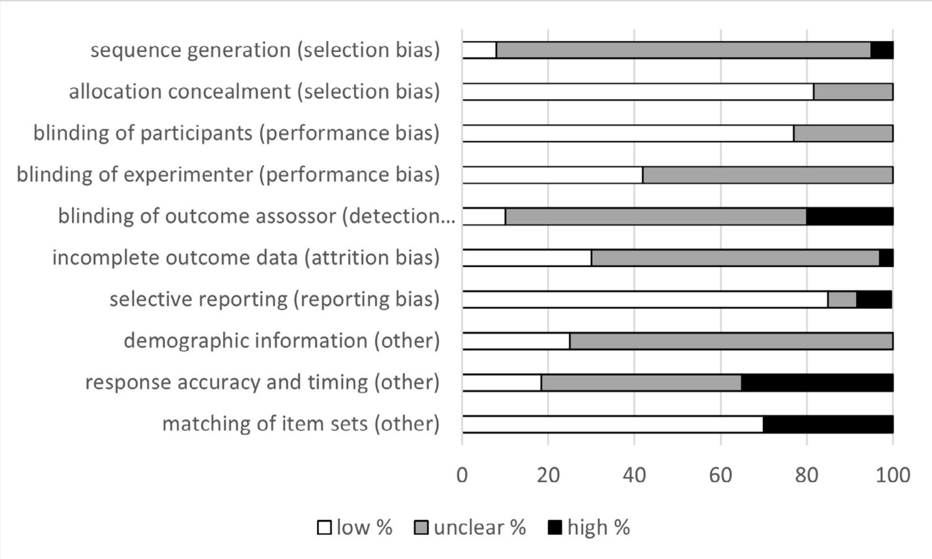
Random Sequence Generation

In studies with a between-subjects design, the risk of bias associated with the sequence generation domain was judged as low if assignment of participants to individual conditions was randomized. A rating of low risk was also given to within-subjects or mixed design studies in which both the generated sequence of blocks/conditions was counterbalanced across participants (e.g., using a Latin square design) and participants were allocated to each block/condition sequence by a random process. It was also applied to those studies in which a sequence of trials (and thereby conditions) was randomized separately for each participant. Only five (8%) studies met these criteria. The domain received a high-risk judgement if a nonrandom method of sequence allocation was used. This was the case in three studies (5%) in which participants were allocated to a block/condition sequence based on their time of arrival in the laboratory. The method of sequence generation and/or condition or sequence allocation was unreported in 44 of the 60 included studies, and hence their risk of bias was marked as unclear. Risk of bias was also deemed unclear if the sequence of blocks/conditions was unevenly distributed across participants. This was the case in seven studies (12%) in which the number of participants did not correspond to the number of generated block sequences. In three studies, allocation of keys to the right and left hand response was not counterbalanced (all were identified as unclear bias studies).

Allocation Concealment

The risk of selection bias due to inadequate allocation concealment refers to the extent to which knowledge of condition or sequence allocation can influence participant selection. The majority of studies (82%) employed a within-subjects design, and were therefore considered to have a low risk of bias in this particular

Figure 3. *Risk of Bias Graph. Proportion of Picture-Word Interference Studies with Each of the Judgements of the Risk of Bias*



domain. Where a between-subjects design was employed (18%) and no adequate measures were taken to conceal allocation, awareness of the forthcoming assignment could, in principle, have enabled the experimenter to delay the testing of a participant until the next "appropriate" assignment based on some prognostic factors (e.g., a young male participant might do better on a manual decision task) or to alter the allocation altogether. Since the risk of bias could not be ruled out in those studies and none of them reported whether or how allocation was concealed, their risk of bias was marked as unclear.

Blinding of Participants

In most studies, including those with between-subjects designs, participants' awareness of condition allocation (e.g., whether someone is asked to name a picture or make a size judgement of a pictured object) is in itself unlikely to lead to performance bias. However, in rare cases, blinding of participants may be undermined by the blocking or full randomization of conditions. For example, when emotional content of distractors is manipulated as a within-subject variable and taboo and neutral distractor conditions are blocked, participants can develop strategies to optimize their performance in that particular condition. Even when conditions are intermixed but their order is fully randomized, it is possible that trials from the same condition (e.g., related one) would appear in succession, allowing participants to generate expectancies and affecting their processing of subsequent stimuli. Studies in which such designs were employed and in which information on participant blinding was omitted were considered to have an unclear risk of bias (23%). In the remaining studies, the risk of bias was judged as low because participants' awareness of condition allocation was either unlikely to affect their performance or it was minimized by adequate randomization of conditions.

Blinding of Experimenter

Blinding of experimenter pertains to a situation in which the person conducting an experiment remains unaware of condition allocation throughout the duration of the experiment. Knowledge of condition allocation may affect an experimenter's attitude towards participants, leading to performance bias. Although none of the assessed studies explicitly stated whether or not the experimenter was blinded to condition allocation, this was unlikely to pose a risk in 42% of the assessed studies due to randomization of conditions. Inadequate blinding of experimenter could have been a source of performance bias in the remaining studies in which a between-subjects or a blocked within-subjects design was used, and hence, their risk of bias was marked as unclear.

Blinding of Outcome Assessor

Risk of bias due to inadequate blinding of outcome assessor refers to the extent to which measurement of an outcome is influenced by an assessor's knowledge of condition allocation. In the context of PWI studies, this pertains primarily to measurement of participants' vocal responses, which, unlike the recording of manual responses, allows some room for judgement (e.g., was the participant's response correct, was the voice key triggered prematurely by a nonspeech sound). None of the studies reported whether or not the person coding participants' responses was blinded to condition allocation. Their risk of detection bias was judged as low, however, if the study used a randomized within-subjects design and if response coding in that study was verified offline with an audio-recording (10% of studies). If no information was given about how participants' responses were evaluated, or a within-subjects design with intermixed conditions was used but offline response accuracy checking remained unspecified, the risk of bias was judged as unclear (70% of studies). Assessment of participants' responses was at a high risk of bias in studies which used between-subjects designs or within-subjects designs with condition blocking, in which the experimenter registered participant's responses online (i.e., while the experiment was in progress), and in which no objective record (e.g., an audio recording) was described to verify the accuracy of the experimenter's judgement (20% of studies).

Incomplete Outcome Data

In the context of PWI studies, incomplete outcome data handling pertains to identification of errors and reaction time outliers, their exclusion from data analysis, and subsequent justification of that exclusion. Outlier measurement, exclusion of problematic data points and reasons for their exclusion were adequately reported in one quarter of the assessed studies. In the overwhelming majority of studies (67%), which were considered to have an unclear risk of bias, at least two flaws of incomplete outcome data handling were identified. These include lack of clarity on outlier identification (e.g., whether reaction time outliers were calculated using global mean across all participants and conditions or individual conditional mean), clustering of all error types (e.g., incorrect responses, premature voice key triggering, voice key malfunction, disfluencies) into one category, and provision of a global percentage of excluded data without breakdown per condition. When there was no evidence of data screening, the risk of bias was marked as high (3%).

Selective Reporting

Bias may also be introduced through the selective use and reporting of statistical tests. Risk of bias appeared to be high in 8% of studies in which either no descriptive and inferential statistics were provided, a post-hoc test was performed but no post-hoc statistics were reported, or in which the reported effect would not have survived post-hoc corrections had those been applied. In 85% of studies, all prespecified outcomes were reported with adequate detail. A small number of studies (7%) received an unclear risk of bias judgement due to apparent discrepancies between descriptive and inferential statistics (e.g., where descriptive data would suggest an interaction, no statistically significant interaction was reported).

Other Bias: Selective Reporting of Demographic Information

Only one quarter of the assessed studies provided adequate detail on age, gender, and the first language of participants. The vast majority (75%) neglected to report either one, a combination of two, or all of the variables. Hence their risk of bias was marked as unclear.

Other Bias: Verification of Response Accuracy and Timing

Even with adequate blinding of outcome assessment, the procedure used to register participants' vocal responses is not error-proof. Often, the experimenter evaluates responses in real time, during a brief (typically a one- or two-second-long) interval before the onset of the next trial. This was a standard procedure in 35% of the assessed studies, which were marked as high risk. In 47% of studies, there was no indication of how responses were evaluated or whether or not their accuracy and timing were rechecked offline, and hence their risk of bias was deemed unclear. Verification of response accuracy and timing based on audio-recordings was reported only in 18% of studies. These were therefore judged to be at a low risk of bias.

Other Bias: Matching of Item Sets

The validity of findings also becomes questionable when the stimuli sets between conditions are not matched on relevant psycholinguistic properties. For example, distractor words across low- and high-frequency conditions may be matched on length, but not on other variables known to affect processing speed, such as age of acquisition or imageability. The standard procedure to avoid complications due to inadequate matching of items sets is to recombine pictures and distractors from the related condition into unrelated pairs of items. In studies in which this was possible and in those in which adequate matching was realized (70%), risk of bias was judged as low. Matching was not realized in 30% of studies, and these were judged to be at a high risk of bias.

Recommendations About the Picture-Word Interference Methodology

Based on the risk of bias assessment, the following recommendations can be made about the design and conduct of future PWI studies.

Many authors neglect to report how sequences of trials, blocks, and conditions were generated and whether or not participants were allocated to each sequence by a random process. Investigators should not only minimize selection bias by ensuring that participants are allocated to a generated sequence based on a method that includes an element of chance and that the generated sequences are adequately counterbalanced in blocked designs (i.e., an equal ratio of participants is allocated to each sequence, assignment of keys/buttons to responses is counterbalanced in manual tasks), but also communicate these efforts to the reader.

The risk of performance bias arising from inadequate blinding of participants can be reduced by the choice of a within-subjects design in which the conditions of interest are intermixed rather than blocked, by inclusion of filler trials (especially when relatedness proportion is high, in some cases reaching 67%), and by the use of pseudo-random rather than fully randomized sequences. These efforts can be complemented by the use of post-test awareness probes to gauge participants' awareness of any regularities or interdependencies in the performed tasks.

More clarity should be given of experimenter and outcome assessor blinding, especially when knowledge of condition allocation poses a genuine risk of performance or detection bias. Ideally, experiments should be conducted and participants' vocal responses evaluated by a person who is naïve to the aims of the study. If this is not feasible, future studies should incorporate within-subjects designs in which conditions are intermixed rather than blocked. Since evaluation of participants' vocal responses in real time is not fully objective or error-proof, an audio--recording should be obtained of the experimental session for off-line checking of response accuracy

Information about how and why the data were trimmed and about the amount of data removed from statistical analyses should be adequately reported. This includes provision of percentages of errors and RT outliers per condition, description of RT outlier identification (i.e., if absolute or standard deviation cut-offs were adopted; if the latter, whether these were calculated from global or individual condition means), and differentiation between error types (e.g., incorrect responses should be separated from voice key malfunction). Data analysts should ideally be blinded to the conditions of interest, so that the risk of "inconvenient" data suppression or manipulation is minimized.

There is also a need for a more thorough reporting of descriptive and inferential statistics. It is not uncommon for authors to omit measures of variability or post-hoc statistics. Where multiple comparisons have been made, post-hoc corrections should be applied.

The risk of bias assessment has indicated persistent poor reporting of demographic details. Appropriate background information such as age, gender and first language of participants should be clearly stated.

Relying entirely on the automatic detection of response onset times by a voice key may introduce error (Protopapas et al., 2007). Specialized software can be used to visually inspect a waveform for premature triggering of the voice key (e.g., by non-speech sounds and movement) and for voice key activation failures (e.g., due to insufficiently loud responses).

Matching of stimuli sets across experimental conditions on variables known to affect item processing should be optimized where re-assignment of pairs is not possible. Careful consideration should be given to factors affecting the speed with which both words (e.g., lexical frequency, age of acquisition, length, imageability) and pictures (e.g., complexity, name agreement, concept familiarity) are processed.

Summary of Findings and Future Directions

The current review has systematically reviewed 117 PWI studies which manipulated the various task parameters (i.e., distractor format, target-distractor relationship) to address the question of the competitive nature of word production. Below is a summary of the main findings from individual manipulations, their implications for the proposed accounts of lexical selection and where gaps have been identified, suggestions for future research.

Evidence From Distractor Format Manipulation

Contrary to the claim that categorically related distractor pictures facilitate naming, the majority of the reviewed studies have shown the opposite – that pictorial distractors belonging to the same semantic category as targets interfere more with picture naming than their unrelated controls. A similar observation was made for distractors in the form of environmental sounds. This was apparent under conditions which promoted lexical encoding of the nonverbal interfering stimuli. The fact that pictorial and environmental sound distractors produced interference, despite having no articulatory advantage over target picture names, presents a challenge to the REH account. There is also the possibility, in accordance with the CEH, that the structural and/or conceptual information activated by nonverbal distractors introduces temporary uncertainty about what it is that one sees (object recognition) and/or what it is that one needs to name (concept selection). On the other hand, if that were the only source of interference, an inhibitory effect would be obtained irrespective of whether or not pictorial (or environmental sound) distractors were adequately lexicalized. The evidence obtained from distractor format manipulation thus appears to favor a competitive view of lexical selection with competition threshold (the distracting lexical units must accumulate sufficient activation to interfere with production), with early prelexical decision processes potentially also contributing to the effect.

The only study that utilized pseudo-words as distractors reported facilitated picture naming in the context of unrelated illegitimate words compared to unrelated real words. Although this observation was originally explained in terms of a post-lexical self-monitoring mechanism that is fine-tuned to detect and eliminate meaningless words more quickly than real words (lexicality bias), it also fits well with alternative competitive (SLNH and selection-through-competition with competition threshold) and noncompetitive (REH and CEH) accounts of semantic context effects.

Evidence From Distractor Frequency Manipulation

The evidence produced by studies with distractor frequency manipulation does not allow for strong conclusions to be drawn about the locus of semantic context effects or the mechanisms from which they emerge. The distractor frequency effect (i.e., prolonged naming in the case of low-frequency words relative to high-frequency words), which has consistently been replicated across laboratories, has received several interpretations, which are consistent with the REH, CEH, perceptual reactive blocking, and attentional capture hypotheses. Joint manipulations of distractor frequency and other variables such as SOA, distractor visibility, or emotional content of distractors, which could help to identify important contributors to the effect, have produced mixed results. Their interpretation is further complicated by questionable presuppositions on which predictions have been made (e.g., the efficacy and extent of masking, the locus of taboo interference). Distractor frequency has been extensively studied within the PWI paradigm, while other intrinsic properties of distractors (e.g., imageability, concept familiarity) remain underexplored. Examining psycholinguistic variables known to exert their effects at different stages of information processing could provide further insight into the processes underlying semantic context effects. This line of PWI research would also benefit from more empirical data derived from cross--factorial designs in which both distractor characteristics and task demands are manipulated. Although some attempts have been made to address this, the results of how the distractor frequency effect is modulated by phonological decisions, for example, were not clear-cut. Other tasks in which perceptual or semantic processing load is increased could also be employed.

Evidence From Distractor Visibility Manipulation

Manipulations designed to prevent phonological responses from entering an articulatory buffer using a masking procedure have produced limited and equivocal evidence. The prediction that interference would turn into facilitation if the need for articulatory buffer clearing was eliminated was confirmed by just a handful of studies. However, other studies have either failed to replicate the polarity reversal or demonstrated interference irrespective of whether the distractors were masked or visible. The validity of the masking procedure was further called into question, as it is not clear what stages are effectively "turned off" under reduced visibility conditions. For example, according to Dehaene et al. (1998), the processing of subliminally presented primes extends all the way down to include the motor system. This cannot be easily reconciled with the claim made by Finkbeiner and Caramazza (2006) that masking successfully prevents formulation of phonological production-ready responses. Even if the stage of articulatory encoding is indeed effectively eliminated by the masking procedure, so could the stage of lexical processing – without distractor lexicalization, interference would be hard, if impossible, to find. Facilitation with masked distractors should not be at all surprising if the processing of subliminally presented distractor words was restricted to their conceptual encoding. It was proposed that the notion of conscious perception of distractors be replaced with the concept of distractor activation strength.

Future research could use a factorial design, in which semantic relatedness, lexical activation strength of distractors and distractor visibility would be concurrently manipulated. If facilitation is indeed due to low activation strength of distractors and not to their masking, interference should be observed for categorically related distractors that have been given adequate lexical boost both under visible and masked distractor conditions. This was partly shown by Piai et al. (2012), but with a between-subject design, a limitation that could be addressed in future PWI studies.

Evidence From Manipulations of Emotional Content of Distractors

Although the taboo interference effect is a robust phenomenon, its origins are not fully understood. The evidence obtained from concurrent manipulations of the emotional content of distractors and other variables such as SOA and phonological relatedness is inconclusive. There is compelling, albeit scant evidence from studies in which emotional content of distractors was factorially crossed with task demands, suggesting that the taboo interference effect is driven by early lexical (competition from co-activated lexical representations) and/or pre-lexical (attentional modulation) processes. Persistent interference in the absence of articulatory preparation (phonological decision task) undermined the role of the articulatory buffer clearing mechanism as the main driving force of the effect. The prelexical basis of the taboo interference effect could not be ruled out because the effect was preserved in the absence of lexical encoding (semantic decision task), although this was only observed under degraded input conditions. The taboo interference effect also contradicts the REH account because according to the response relevance logic, unrelated, socially inappropriate distractor words should be cleared from the articulatory buffer sooner than unrelated neutral distractor words, since taboo words do not meet the implicit criterion of producing socially appropriate speech. Socially inappropriate words should also be detected and eliminated more easily by the verbal self-monitor, leading to facilitation. The opposite pattern of results (i.e., greater interference for taboo words) was explained by more conservative checking of offensive, potentially embarrassing responses by the self-monitor, an activity which takes time. In either case, the lines of reasoning used to explain pseudoword facilitation and taboo interference appear to be contradictory. It is also unclear how a self-monitoring system or the REH can account for the findings of interference for negative distractor words.

Evidence From Hierarchical Relations Manipulations

Between-level interference, namely that distractors that bear hierarchical relations to targets (being drawn from either a subordinate or superordinate category) are capable of inducing interference as long as they are drawn from the same semantic category, appears to be a genuine effect. This does not allow differentiation between the rival accounts, however. Both the lexical selection-by-competition and the REH account are plausible candidates. Moreover, some discrepancies emerged which should ideally be resolved by future research. Interference for basic-level distractors relative to superordinate-level names in the absence of semantic relatedness was observed in two studies but discredited in two others. Interference for subordinate-level distractors that are semantically related to targets relative to their unrelated counterparts observed in several studies appeared as facilitation in another study. Furthermore, the lack of fully-crossed factorial designs in which semantic relatedness, SOA, and level of specificity of distractors with all levels of abstraction (basic-, subordinate-, and superordinate-level) were manipulated creates a need for more research. Although this was partially achieved by Kuipers et al. (2006), the manipulation was restricted to basic- and superordinate-level names only. The emerging picture thus raises the question of whether the absence of interaction between semantic relatedness and level of specificity would extend to subordinate-level distractors and whether picture naming in the presence of subordinate-level distractors would differ from that in the presence of basic- and superordinate-level distractors. Comparing picture naming performance for basic-, subordinate-, and superordinate-level distractors could have further implications for lexical selection accounts because each type is associated with different processes. For example, when comparing the following pairs, DOG-animal and DOG-Spaniel, the pictorial and semantic information supplied by both distractor words will interact differently with the pictorial and semantic information supplied by the targets – the superordinate category name animal should be rejected more quickly, in accordance with the CEH, because accepting a four-legged creature with a tail as an ANIMAL should take less time than deciding whether or not the picture depicts a SPANIEL.

Evidence From Semantic Distance Manipulations

It is too early to draw any strong conclusions about the robustness of the semantic distance effect. Roughly an equal number of studies have produced an inhibitory, facilitatory or no effect for semantically close distractors relative to their more distant controls. Direct comparisons are problematic because of the different measures used to operationalize semantic distance, in addition to potentially confounding variables such as relatedness proportion or inadequate matching of stimulus sets across experimental conditions.

Evidence From Manipulations of Associative (Miscellaneous) and Thematic Relations

The facilitatory effect for associatively and thematically related distractor words relative to their unrelated controls is a fairly well-established phenomenon, particularly at early SOAs. However, due to the miscellaneous nature of the associative relations, facilitatory effects find plausible explanations in both competitive and noncompetitive accounts. Future research could examine the effect of purely associative (probabilistic) relations on the speed of naming in the absence of semantic relatedness

Evidence From Manipulations of Whole-Part Relations

Based on the evidence from manipulations of whole-part relations in the PWI paradigm, two critical factors appear to determine the direction of the semantic interference effect: distinctiveness and visibility of distractor-denoted parts in target pictures. Distinctiveness (or strong association between items) appears to be the driving force of the facilitatory effect, which can reflect the interplay either between strong semantic priming and weak lexical interference (in accordance with the SLNH), or strong semantic priming and lack of interference due to articulatory buffer clearing (in accordance with the REH). Distractors denoting non-distinctive features of targets either have no effect on picture naming or interfere with production when presented after target onset. While the null results could be due to the absence of semantic priming, interference could be explained by lexical competition. Prelexical decision processes (e.g., uncertainty about what needs to be named) could also contribute to the net inhibitory effect, especially with distractors denoting parts that are visible in target pictures. This claim found support in one study which specifically manipulated the visibility of distractor--denoted parts in target pictures, but was discredited in a post-hoc analysis of another study, leaving the issue unresolved.

Evidence From Manipulations of Visual Similarity

Despite an indication that visual similarity between targets and distractors in the absence of semantic relatedness contributes to the net inhibitory effect observed in the PWI task, which would suggest a pre-lexical locus of the effect, its robustness is far from settled. This issue requires additional work to establish the origin and reliability of the visual similarity effect. This could be achieved, for example, by including a range of SOAs, particularly early ones, at which processes such as object recognition and identification have a better chance of manifestation, while using targets and distractors that do not share response set membership. It would also be interesting to broaden the spectrum of structural features of objects (e.g., color, size, shape, texture) used to manipulate visual similarity between targets and distractors. Different statistical methods, such as multiple hierarchical regression analyses, could be employed to gauge the relative contribution of visual similarity to the net PWI effect, above and beyond other relevant variables such as semantic relatedness or semantic distance between targets and distractors.

Evidence From Manipulations of Task Demands

PWI studies with task demand manipulations have not furnished any clear answers about the mechanisms that could drive semantic context effects or the loci at which these effects emerge.

Subordinate Level Naming

PWI studies utilizing both perceptual-conceptual tasks and subordinate-level naming tasks present contradictory findings, with a range of effects. In addition, each can be challenged on methodological or conceptual grounds, such as implicit lexicalization, selection of tasks in which individual processing demands are not fully understood, or inadvertent use of procedures that can alter the nature of the task from perceptually- to lexically-based. There is therefore scope for more research utilizing purely perceptual tasks, but where task relevance of distractors is preserved. It is easy to imagine a perceptual task, such as orientation judgement task, in which participants decide whether a target object is upright or tilted, and in which performance is not confounded by semantic or lexical processing, but which may be of little use because information supplied by the distractor (be it semantic or structural) is not relevant to the task at hand. Use of fully crossed factorial designs, for example, comparing PWI performance with identical basic-level distractors to that with alternative basic-level distractors in addition to concurrent manipulation of semantic relatedness or the level of abstraction from which distractors are drawn, would also be an advantage, as would elimination of potential confounds, such as familiarization with experimental materials and message congruency.

Semantic Decision Tasks

A range of effects has been reported with the few PWI studies employing nonabstraction semantic decision tasks (i.e., living/nonliving, natural/mademade and size classification). When the task was changed to higher-level category naming (abstraction task), categorization was generally faster in the context of semantically related distractors than their unrelated controls. Interpretation of results in both groups of semantic decision studies however, is complicated by the confound of response congruency (response congruent targets and distractors leading to facilitation). Mixed evidence is compounded by the dearth of studies in which task (semantic decision versus basic-level naming) and semantic relatedness have been concurrently manipulated.

Phonological Decision Tasks

There is fairly consistent, albeit scant, evidence that the interference effect is preserved even if the task does not explicitly require generation of articulatory codes. The findings undermine the role of response-competition as well as self-monitoring processes at the articulatory stage as a single source of interference. The data are not incompatible, however, with accounts that place the locus of interference at an early, prelexical stage. For example, the presence of the distractor frequency effect in phonological decision tasks could be explained by an attentional capture account, according to which low frequency words (by virtue of being rare) in comparison to high frequency words attract additional cognitive resources, diverting attention away from target processing. Similarly, although Abdel Rahman and Aristei (2010) argued for interference to arise at the level of lexical selection as the most parsimonious account of the semantic interference effect in the absence of overt articulation, the results do not rule out the possibility that the effect resides outside the lexical selection stage, being an epiphenomenon of concept selection (due to competing conceptual representations), or concept rejection (with a conflict detection mechanism intercepting and possibly blocking conceptual representations that have been wrongly selected).

Main Summary and Conclusions

This systematic review has produced some important findings. Contrary to previous claims, interference with nonverbal distractors (the picture-picture interference effect) has been repeatedly observed, but only under conditions of sufficient distractor lexicalization – a finding which undermines the response-exclusion hypothesis (REH), but which can nonetheless be explained by the swinging lexical network hypothesis (SLNH) and the selection-through-competition with competition threshold account. Contrary to the accepted view that naming gets faster as distractors become more semantically similar to targets (the semantic distance effect), a range of effects from facilitation through null results to interference has been reported across studies. The distractor frequency effect, a phenomenon that has been repeatedly observed, has received several interpretations consistent with the REH, CEH, perceptual reactive blocking, and attentional capture accounts. Evidence for a number of effects, especially in studies in which cross-factorial designs were used, remains inconclusive. Evidence for other effects is scant, with isolated findings in need of replication across laboratories. Evidence for some of the more well-established effects in turn, does not allow us to differentiate between the rival theories. The concept exclusion hypothesis (CEH) has not been given adequate consideration, even though early processes associated with picture naming could potentially be influenced by structural and semantic information supplied by the distractor. Several studies have implicated the prelexical stage as a potential locus of the PWI effect, but except for studies manipulating the visual similarity between targets and distractors, its contribution to the net semantic interference effect has not been directly assessed. It therefore remains a viable determinant of the speed with which pictures are named in the PWI task.

In relation to the main research question of whether lexical retrieval is delayed by activation of nontarget lexical representations, this cannot be determined based on the available PWI results. While some of the well-established effects (e.g., interference with non-verbal distractors and with taboo and negative distractor words) challenge the noncompetitive theories, specifically the REH and the self-monitoring hypotheses, weighing the evidence in favor of the competitive explanations, specifically the swinging lexical network hypothesis (SLNH) and the selection-through-competition with competition threshold account, the contribution of other non-lexical processes to the net semantic context effect, such as the concept exclusion hypotheses (the CEH) cannot be ruled out. From the reviewed data, it transpires that the semantic context effect is a function of several forces operating at different stages of information processing (with prelexical, lexical, and postlexical loci) and at various levels of intensity. These forces may additionally interact with one another, which leads to the cancellation of some of the effects and potentiation of others, with the net semantic context effect remaining difficult to unravel.

Acknowledgments

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. However, the first author was supported by a Middlesex University PhD scholarship.

Conflict of Interest Disclosure

The Authors report there are no competing interests to declare.

Funding

No funding was received for conducting this study.

Research Ethics Statement

The study was approved by the research ethics committee at Middlesex University.

Data Availability Statement

The datasets generated during and/or analyzed during the current study are available in the OSF repository: https://osf.io/7qkh4

Authorship Details

Małgorzata Korko: research concept and design, collection and/or assembly of data, data analysis and interpretation, writing the article, critical revision of the article, final approval of the article. Arpita Bose: collection and/or assembly of data, data analysis and interpretation, critical revision of the article. Alexander Jones: collection and/or assembly of data, data analysis and interpretation, critical revision of the article. Mark Coulson: collection and/or assembly of data, data analysis and interpretation, critical revision of the article. Paul de Mornay Davies: collection and/or assembly of data, data analysis and interpretation, critical revision of the article, final approval of the article.

References

Abdel Rahman, R., & Aristei, S. (2010). Now you see it ... and now again: Semantic interference reflects lexical competition in speech production with and without articulation. *Psychonomic Bulletin & Review, 17*(5), 657–661. https://doi.org/10.3758/PBR.17.5.657

Abdel Rahman, R., & Melinger, A. (2007). When bees hamper the production of honey: Lexical interference from associates in speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*(3), 604–614. https://doi.org/10.1037/0278-7393.33.3.604

Abdel Rahman, R., & Melinger, A. (2009). Semantic context effects in language production: A swinging lexical network proposal and a review. *Language and Cognitive Processes, 24*(5), 713–734. https://doi.org/10.1080/01690960802597250

Abdel Rahman, R., & Melinger, A. (2019). Semantic processing during language production: an update of the swinging lexical network. *Language, Cognition and Neuroscience, 34*(9), 1176–1192. https://doi.org/10.1080/23273798.2019.1599970

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

Aristei, S., & Abdel Rahman, R. (2013). Semantic interference in language production is due to graded similarity, not response relevance. *Acta Psychologica, 144*(3), 571–582. https://doi.org/10.1016/j.actpsy.2013.09.006

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

Baars, B. J., Motley, M. T., & MacKay, D. G. (1975). Output editing for lexical status in artificially elicited slips of the tongue. *Journal of Verbal Learning and Verbal Behavior, 14*(4), 382–391. https://doi.org/10.1016/S0022-5371(75)80017-X

Balota, D. A., & Chumbley, J. I. (1984). Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision stage. *Journal of Experimental Psychology: Human Perception and Performance, 10*(3), 340–357. https://doi.org/10.1037/0096-1523.10.3.340

Bar, M. (2004). Visual objects in context. *Nature Reviews Neuroscience, 5*(8), 617–629. https://doi.org/10.1038/nrn1476

Bloem, I., & La Heij, W. (2003). Semantic facilitation and semantic interference in word translation: Implications for models of lexical access in language production. *Journal of Memory and Language, 48*(3), 468–488.https://doi.org/10.1016/S0749-596X(02)00503-X

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

Bölte, J., Dohmes, P., & Zwitserlood, P. (2013). Interference and facilitation in spoken word production: effects of morphologically and semantically related context stimuli on picture naming. *Journal of Psycholinguistic Research, 42*, 255–280. https://doi.org/10.1007/s10936-012-9219-1

Bölte, J., Böhl, A., Dobel, C., & Zwitserlood, P. (2015). Investigating the flow of information during speaking: The impact of morpho-phonological, associative, and categorical picture distractors on picture naming. *Frontiers in Psychology, 6*, 1540. https://doi.org/10.3389/fpsyg.2015.01540

Brooks, P. J., Seiger-Gardner, L., & Sailor, K. (2014). Contrasting effects of associates and coordinates in children with and without language impairment: A picture-word interference study. *Applied Psycholinguistics, 35*(3), 515–545. https://doi.org/10.1017/S0142716412000495

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

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

Caramazza, A., & Costa, A. (2000). The semantic interference effect in the picture-word interference paradigm: Does the response set matter? *Cognition, 75*(2), B51–B64. https://doi.org/10.1016/S0010-0277(99)00082-7

Catling, J. C., Dent, K., Johnston, R. A., & Balding, R. (2010). Age of acquisition, word frequency, and picture-word interference. *The Quarterly Journal of Experimental Psychology, 63*(7), 1304–1317. https://doi.org/10.1080/17470210903380830

Cilibrasi, R. L., & Vitányi, P. M. B. (2007). The Google similarity distance. *IEEE Transactions on Knowledge and Data Engineering, 19*(3), 370–383. https://doi.org/10.1109/TKDE.2007.48

Costa, A., Alario, F.-X., & Caramazza, A. (2005). On the categorical nature of the semantic interference effect in the picture-word interference paradigm. *Psychonomic Bulletin & Review, 12*(1), 125–131. https://doi.org/10.3758/BF03196357

Costa, A., Mahon, B., Savova, V., & Caramazza, A. (2003). Level of categorization effect: A novel effect in the picture-word interference paradigm. *Language and Cognitive Processes, 18*(2), 205–233. https://doi.org/10.1080/01690960143000524

Cutting, J. C., & Ferreira, V. S. (1999). Semantic and phonological information flow in the production lexicon. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*(2), 318–344. https://doi.org/10.1037/0278-

7393.25.2.318

Damian, M. F., & Bowers, J. S. (2003). Locus of semantic interference in picture-word interference tasks. *Psychonomic Bulletin & Review, 10*(1), 111–117. https://doi.org/10.3758/BF03196474

Damian, M. F., & Spalek, K. (2014). Processing different kinds of semantic relations in picture-word interference with non-masked and masked distractors. *Frontiers in Psychology, 5*, 1183. https://doi.org/10.3389/fpsyg.2014.01183

Dell, G. S., & O'Seaghdha, P. G. (1991). Mediated and convergent lexical priming in language production: A comment on Levelt et al (1991). *Psychological Review, 98*(4), 604–614. https://doi.org/10.1037/0033-295X.98.4.604

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

de Zubicaray, G. I., McLean, M., Oppermann, F., Hegarty, A., McMahon, K., & Jescheniak, J. D. (2018). The shape of things to come in speech production: Visual form interference during lexical access. *Quarterly Journal of Experimental Psychology*, *71*(9), 1921–1938. https://doi.org/10.1080/17470218.2017.1367018

Dean, M. P., Bub, D. N., & Masson, M. E. J. (2001). Interference from related items in object identification. *Journal of Experimental Psychology: Learning Memory and Cognition, 27*(3), 733–743. https://doi.org/10.1037/0278-7393.27.3.733

Dehaene, S., Naccache, L., Le Clec'H, G., Koechlin, E., Mueller, M., Dehaene-Lambertz, G., Van De Moortele, P. F., & Le Bihan, D. (1998). Imaging unconscious semantic priming. *Nature, 395*(6702), 597–600. https://doi.org/10.1038/26967

Dhooge, E., & Hartsuiker, R. J. (2010). The distractor frequency effect in picture–word interference: Evidence for response exclusion. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*(4), 878–891. https://doi.org/10.1037/a0019128

Dhooge, E., & Hartsuiker, R. J. (2011). How do speakers resist distraction? *Psychological Science, 22*(7), 855–859. https://doi.org/10.1177/0956797611410984

Dhooge, E., & Hartsuiker, R. J. (2012). Lexical selection and verbal self-monitoring: Effects of lexicality, context, and time pressure in picture-word interference. *Journal of Memory and Language, 66*(1), 163–176. https://doi.org/10.1016/j.jml.2011.08.004

Estes, Z., Golonka, S., & Jones, L. L. (2011). Thematic thinking. The apprehension and consequences of thematic relations. *Psychology of Learning and Motivation - Advances in Research and Theory, 54*, 249–294. https://doi.org/10.1016/B978-0-12-385527-5.00008-5

Finkbeiner, M., & Caramazza, A. (2006). Lexical selection is not a competitive process: A reply to La Heij et al (2006). *Cortex, 42*(7), 1032–1035. https://

doi.org/10.1016/S0010-9452(08)70210-7

Gauthier, I., James, T. W., Curby, K. M., & Tarr, M. J. (2003). The influence of conceptual knowledge on visual discrimination. *Cognitive Neuropsychology, 20*(3–6), 507–523. https://doi.org/10.1080/02643290244000275

Gehring, W. J., Goss, B., Coles, M. G. H., Meyer, D. E., & Donchin, E. (1993). A neural system for error detection and compensation. *Psychological Science, 4*(6), 385–390. https://doi.org/10.1111/j.1467-9280.1993.tb00586.x

Geng, J., Kirchgessner, M., & Schnur, T. (2013). The mechanism underlying lexical selection: Evidence from the picture–picture interference paradigm. *The Quarterly Journal of Experimental Psychology, 66*(2), 261–276. https://doi.org/10.1080/17470218.2012.705861

Geng, J., Schnur, T. T., & Janssen, N. (2014). Relative speed of processing affects interference in Stroop and picture–word interference paradigms: Evidence from the distractor frequency effect. *Language, Cognition and Neuroscience, 29*(9), 1100–1114. https://doi.org/10.1080/01690965.2013.846473

Glaser, W. R., & Düngelhoff, F.-J. (1984). The time course of picture-word interference. *Journal of Experimental Psychology: Human Perception and Performance, 10*(5), 640–654. https://doi.org/10.1037/0096-1523.10.5.640

Glaser, W. R., & Glaser, M. O. (1989). Context effects in stroop-like word and picture processing. *Journal of Experimental Psychology: General, 118*(1), 13. https://doi.org/10.1037//0096-3445.118.1.13

Hansen, S. J., McMahon, K. L., Burt, J. S., & de Zubicaray, G. I. (2017). The locus of taboo context effects in picture naming. *The Quarterly Journal of Experimental Psychology, 70*(1), 75–91. https://doi.org/10.1080/17470218.2015.1124895

Hantsch, A., Jescheniak, J. D., & Mädebach, A. (2012). Naming and categorizing objects: Task differences modulate the polarity of semantic effects in the picture–word interference paradigm. *Memory & Cognition, 40*(5), 760-768. https://doi.org/10.3758/s13421-012-0184-6

Hantsch, A., Jescheniak, J. D., & Schriefers, H. (2005). Semantic competition between hierarchically related words during speech planning. *Memory and Cognition, 33*(6), 984–1000. https://doi.org/10.3758/BF03193207

Hantsch, A., Jescheniak, J. D., & Schriefers, H. (2009). Distractor modality can turn semantic interference into semantic facilitation in the picture–word interference task: Implications for theories of lexical access in speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*(6), 1443–1453. https://doi.org/10.1037/a0017020

Higgins, J. P. T., & Green, S. E. (2011). The Cochrane Collaboration. www.cochrane-handbook. org

Humphreys, G. W., Lloyd-Jones, T. J., & Fias, W. (1995). Semantic interference effects on naming using a postcue procedure: Tapping the links between semantics and phonology with pictures and words. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(4), 961–980. https://doi.

org/10.1037/0278-7393.21.4.961

Hutson, J., & Damian, M. F. (2014). Semantic gradients in picture-word interference tasks: Is the size of interference effects affected by the degree of semantic overlap? *Frontiers in Psychology, 5*, 872. https://doi.org/10.3389/fpsyg.2014.00872

Hutson, J., Damian, M. F., & Spalek, K. (2013). Distractor frequency effects in picture–word interference tasks with vocal and manual responses. *Language and Cognitive Processes, 28*(5), 615–632. https://doi.org/10.1080/01690965.2011.605599

Jescheniak, J. D., Matushanskaya, A., Mädebach, A., & Müller, M. M. (2014). Semantic interference from distractor pictures in single-picture naming: Evidence for competitive lexical selection. *Psychonomic Bulletin & Review, 21*(5), 1294–1300. https://doi.org/10.3758/s13423-014-0606-5

Jolicoeur, P., Gluck, M. A., & Kosslyn, S. M. (1984). Pictures and names: Making the connection. *Cognitive Psychology, 16*(2), 243–275. https://doi.org/10.1016/0010-0285(84)90009-4

Klein, G. S. (1964). Semantic power measured through the interference of words with color-naming. *American Journal of Psychology, 77*, 576–588. https://doi.org/10.2307/1420768

Kuipers, J. R., La Heij, W., & Costa, A. (2006). A further look at semantic context effects in language production: The role of response congruency. *Language and Cognitive Processes, 21*(7–8), 892–919. https://doi.org/10.1080/016909600824211

La Heij, W., Heikoop, K. W., Akerboom, S., & Bloem, I. (2003). Picture naming in picture context: Semantic interference or semantic facilitation? *Psychology Science, 45*(1), 49–62.

La Heij, W., Dirkx, J., & Kramer, P. (1990). Categorical interference and associative priming in picture naming. *British Journal of Psychology, 81*(4), 511–525. https://doi.org/10.1111/j.2044-8295.1990.tb02376.x

Landauer, T. K., Foltz, P. W., & Laham, D. (1998). An introduction to latent semantic analysis. *Discourse Processes, 25*(2–3), 259–284. https://doi.org/10.1080/01638539809545028

Levelt. W. J. M. (1989). *Speaking: From infmtion to arriculation.* MIT Press.

Levelt, W. J. M. (1999). Models of word production. *Trends in Cognitive Sciences, 3*(6), 223–232. https://doi.org/10.1016/S1364-6613(99)01319-4

Levelt, W. J. M. (1992). Accessing words in speech production: Stages, processes and representations. *Cognition, 42*(1–3), 1–22. https://doi.org/10.1016/0010-0277(92)90038-J

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22*(1), 1–75. https://doi.org/10.1017/S0140525X99001776

Lin, E. L., Murphy, G. L., & Shoben, E. J. (1997). The effects of prior processing episodes on basic-level superiority. *Quarterly Journal of Experimental Psy-*

chology Section A: Human Experimental Psychology, 50(1), 25–48. https://doi.org/10.1080/713755686

Lupker, S. J. (1979). The semantic nature of response competition in the picture-word interference task. *Memory & Cognition, 7*(6), 485–495. https://doi.org/10.3758/BF03198265

Lupker, S. J., & Katz, A. N. (1981). Input, decision, and response factors in picture-word interference. *Journal of Experimental Psychology: Human Learning and Memory, 7*(4), 269–282. https://doi.org/10.1037/0278-7393.7.4.269

Mädebach, A., Markuske, A.-M., & Jescheniak, J. D. (2018). When does reading dirty words impede picture processing? Taboo interference with verbal and manual responses. *Psychonomic Bulletin & Review, 25*(6), 2301–2308. https://doi.org/10.3758/s13423-018-1468-z

Mädebach, A., Wöhner, S., Kieseler, M.-L., & Jescheniak, J. D. (2017). Neighing, barking, and drumming horses—object related sounds help and hinder picture naming. *Journal of Experimental Psychology: Human Perception and Performance, 43*(9), 1629–1646. https://doi.org/10.1037/xhp0000415

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), 503–535. https://doi.org/10.1037/0278-7393.33.3.503

Matushanskaya, A., Mädebach, A., Müller, M. M., & Jescheniak, J. D. (2016). When sufficiently processed, semantically related distractor pictures hamper picture naming: Implications for models of lexical access in speech production. *Experimental Psychology, 63*(6), 307–317. https://doi.org/10.1027/1618-3169/a000340

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: I. An account of basic findings. *Psychological Review, 88*(5), 375–407. https://doi.org/10.1037/0033-295X.88.5.375

McRae, K., Cree, G. S., Seidenberg, M. S., & McNorgan, C. (2005). Semantic feature production norms for a large set of living and nonliving things. *Behavior Research Methods, 37*(4), 547–559. https://doi.org/10.3758/BF03192726

Miozzo, M., & Caramazza, A. (2003). When more is less: A counterintuitive effect of distractor frequency in the picture-word interference paradigm. *Journal of Experimental Psychology: General, 132*(2), 228–252. https://doi.org/10.1037/0096-3445.132.2.228

Muehlhaus, J., Heim, S., Sachs, O., Schneider, F., Habel, U., & Sass, K. (2013). Is the motor or the garage more important to the car? The difference between semantic associations in single word and sentence production. *Journal of Psycholinguistic Research, 42*(1), 37–49. https://doi.org/10.1007/s10936-012-9209-3

Mulatti, C., & Coltheart, M. (2012). Picture–word interference and the response–exclusion hypothesis. *Cortex, 48*(3), 363–372. https://doi.org/10.1016/j.cortex.2011.04.025

Müller, H. J., Geyer, T., Zehetleitner, M., & Krummenacher, J. (2009). Attentional capture by salient color singleton distractors is modulated by top-down dimensional set. *Journal of Experimental Psychology: Human Perception and Performance, 35*(1), 1–16. https://doi.org/10.1037/0096-1523.35.1.1

Navarrete, E., & Costa, A. (2005). Phonological activation of ignored pictures: Further evidence for a cascade model of lexical access. *Journal of Memory and Language, 53*(3), 359–377. https://doi.org/10.1016/J.JML.2005.05.001

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

Oppenheim, G. M., & Dell, G. S. (2008). Inner speech slips exhibit lexical bias, but not the phonemic similarity effect. *Cognition, 106*(1), 528–537. https://doi.org/10.1016/j.cognition.2007.02.006

Piai, V., Roelofs, A., & Schriefers, H. (2012). Distractor strength and selective attention in picture-naming performance. *Memory & Cognition, 40*(4), 614–627. https://doi.org/10.3758/s13421-011-0171-3

Piai, V., Roelofs, A., & van der Meij, R. (2012). Event-related potentials and oscillatory brain responses associated with semantic and Stroop-like interference effects in overt naming. *Brain Research, 1450*, 87–101. https://doi.org/10.1016/j.brainres.2012.02.050

Protopapas, A., Archonti, A., & Skaloumbakas, C. (2007). Reading ability is negatively related to Stroop interference. *Cognitive Psychology, 54*(3), 251–282. https://doi.org/10.1016/j.cogpsych.2006.07.003

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

Roelofs, A. (2003). Goal-referenced selection of verbal action: Modeling attentional control in the stroop task. *Psychological Review, 110*(1), 88–125. https://doi.org/10.1037/0033-295X.110.1.88

Roelofs, A., Piai, V., & Schriefers, H. (2013). Context effects and selective attention in picture naming and word reading: Competition versus response exclusion. *Language and Cognitive Processes, 28*(5), 655–671. https://doi.org/10.1080/01690965.2011.615663

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology, 8*(3), 382–439. https://doi.org/10.1016/0010-0285(76)90013-X

Rose, S. B., Aristei, S., Melinger, A., & Abdel Rahman, R. (2019). The closer they are, the more they interfere: Semantic similarity of word distractors increases competition in language production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 45*(4), 753–763. https://doi.

org/10.1037/xlm0000592

Rosinski, R. R. (1977). Picture-word interference is semantically based. *Child Development, 48*(2), 643-647. https://doi.org/10.2307/1128667

Rosinski, R. R., Golinkoff, R. M., & Kukish, K. S. (1975). Automatic semantic processing in a picture-word interference task. *Child Development, 46*(1), 247-253. https://doi.org/10.2307/1128859

Sailor, K., & Brooks, P. J. (2014). Do part–whole relations produce facilitation in the picture–word interference task? *The Quarterly Journal of Experimental Psychology, 67*(9), 1768–1785. https://doi.org/10.1080/17470218.2013.870589

Sailor, K., Brooks, P. J., Bruening, P. R., Seiger-Gardner, L., & Guterman, M. (2009). Exploring the time course of semantic interference and associative priming in the picture–word interference task. *The Quarterly Journal of Experimental Psychology, 62*(4), 789–801. https://doi.org/10.1080/17470210802254383

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

Severens, E., Kühn, S., Hartsuiker, R. J., & Brass, M. (2012). Functional mechanisms involved in the internal inhibition of taboo words. *Social Cognitive and Affective Neuroscience, 7*(4), 431–435. https://doi.org/10.1093/scan/nsr030

Smith, M. C., & Magee, L. E. (1980). Tracing the time course of picture–word processing. *Journal of Experimental Psychology: General, 109*(4), 373–392. https://doi.org/10.1037/0096-3445.109.4.373

Spalek, K., Damian, M. F., & Bölte, J. (2013). Is lexical selection in spoken word production competitive? Introduction to the special issue on lexical competition in language production. *Language and Cognitive Processes, 28*(5), 597–614. https://doi.org/10.1080/01690965.2012.718088

Spence, D. P., & Owens, K. C. (1990). Lexical co-occurrence and association strength. *Journal of Psycholinguistic Research, 19*(5), 317–330. https://doi.org/10.1007/BF01074363

Starreveld, P. A., & La Heij, W. (1995). Semantic interference, orthographic facilitation, and their interaction in naming tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(3), 686–698. https://doi.org/10.1037/0278-7393.21.3.686

Starreveld, P. A., La Heij, W., & Verdonschot, R. (2013). Time course analysis of the effects of distractor frequency and categorical relatedness in picture naming: An evaluation of the response exclusion account. *Language and Cognitive Processes, 28*(5), 633–654. https://doi.org/10.1080/01690965.2011.608026

Steyvers, M., & Tenenbaum, J. B. (2005). The large-scale structure of semantic networks: Statistical analyses and a model of semantic growth. *Cognitive*

Science, 29(1), 41–78. https://doi.org/10.1207/s15516709cog2901_3

Vieth, Harrison E, McMahon, K. L., & de Zubicaray, G. I. (2014a). The roles of shared vs distinctive conceptual features in lexical access. *Frontiers in Psychology, 5*. https://doi.org/10.3389/fpsyg.2014.01014

Vieth, H. E., McMahon, K. L., & de Zubicaray, G. I. (2014b). Feature overlap slows lexical selection: Evidence from the picture–word interference paradigm. *Quarterly Journal of Experimental Psychology, 67*(12), 2325–2339. https://doi.org/10.1080/17470218.2014.923922

Vigliocco, G., Lauer, M., Damian, M. F., & Levelt, W. J. M. (2002). Semantic and syntactic forces in noun phrase production. *Journal of Experimental Psychology: Learning Memory and Cognition, 28*(1), 46–58. https://doi.org/10.1037/0278-7393.28.1.46

Vigliocco, G., Vinson, D. P., Lewis, W., & Garrett, M. F. (2004). Representing the meanings of object and action words: The featural and unitary semantic space hypothesis. *Cognitive Psychology, 48*(4), 422–488. https://doi.org/10.1016/J.COGPSYCH.2003.09.001

Vitkovitch, M., & Tyrrell, L. (1999). The effects of distractor words on naming pictures at the subordinate level. *The Quarterly Journal of Experimental Psychology Section A, 52*(4), 905–926. https://doi.org/10.1080/713755854

Wheeldon, L. R., & Levelt, W. J. M. (1995). Monitoring the time course of phonological encoding. *Journal of Memory and Language, 34*(3), 311–334. https://doi.org/10.1006/jmla.1995.1014

White, K. K., Abrams, L., Hsi, L. R., & Watkins, E. C. (2018). Are precues effective in proactively controlling taboo interference during speech production? *Cognition and Emotion, 32*(8), 1625–1636. https://doi.org/10.1080/02699931.2018.1433637

White, K. K., Abrams, L., Koehler, S. M., & Collins, R. J. (2017). Lions, tigers, and bears, oh sh!t: Semantics versus tabooness in speech production. *Psychonomic Bulletin & Review, 24*(2), 489–495. https://doi.org/10.3758/s13423-016-1084-8

White, K. K., Abrams, L., LaBat, L. R., & Rhynes, A. M. (2016). Competing influences of emotion and phonology during picture-word interference. *Language, Cognition and Neuroscience, 31*(2), 265–283. https://doi.org/10.1080/23273798.2015.1101144