Locus of Semantic Interference in Picture Naming: Evidence from Dual-Task Performance

Vitória Piai[1,2], Ardi Roelofs[1], and Herbert Schriefers[1]

[1]Radboud University Nijmegen, The Netherlands

[2]International Max Planck Research School for Language Sciences, Nijmegen, The Netherlands

Author note

This research was supported by a grant (Open Competition MaGW 400-09-138) from the Netherlands Organization for Scientific Research. The authors thank Anna Dieckmann for collecting the data, and Kristoffer Dahlslätt and Harold Pashler for helpful comments.

Correspondence concerning this article should be addressed to Vitória Piai, Radboud University Nijmegen, Donders Institute for Brain, Cognition and Behaviour, Centre for Cognition, Spinoza Building B.01.05, Montessorilaan 3, 6525 HR Nijmegen, The Netherlands. Email: V.Piai@donders.ru.nl. Phone : +31.24.3612635. Fax: +31.24.3612635.

Abstract

Disagreement exists regarding the functional locus of semantic interference of distractor words in picture naming. This effect is a cornerstone of modern psycholinguistic models of word production, which assume that it arises in lexical response-selection. However, recent evidence from studies of dual-task performance suggests a locus in perceptual or conceptual processing, prior to lexical response-selection. In these studies, participants manually responded to a tone and named a picture while ignoring a written distractor word. The stimulus onset asynchrony (SOA) between tone and picture-word stimulus was manipulated. Semantic interference in naming latencies was present at long tone pre-exposure SOAs, but reduced or absent at short SOAs. Under the prevailing structural or strategic response-selection bottleneck and central capacity sharing models of dual-task performance, the underadditivity of the effects of SOA and stimulus type suggests that semantic interference emerges before lexical response-selection. However, in more recent studies, additive effects of SOA and stimulus type were obtained. Here, we examined the discrepancy in results between these studies in six experiments in which we systematically manipulated various dimensions on which these earlier studies differed, including tasks, materials, stimulus types, and SOAs. In all our experiments, additive effects of SOA and stimulus type on naming latencies were obtained. These results strongly suggest that the semantic interference effect arises after perceptual and conceptual processing, during lexical response-selection or later. We discuss several theoretical alternatives with respect to their potential to account for the discrepancy between the present results and other studies showing underadditivity.

Keywords: dual-task performance; picture-word interference; response-selection bottleneck; semantic interference; Stroop

An important question in the psychology of language concerns how speakers select from memory the words that they want to produce. This ability, called lexical selection, is a topic of much research in the field of word production. One way of studying lexical selection consists of presenting participants with pictured objects paired with superimposed distractor words, a paradigm called picture-word interference (PWI) (see Abdel Rahman & Melinger, 2009; Glaser, 1992; Roelofs, 2007, for reviews). Participants are instructed to name the pictures and to ignore the distractors. The relation the distractor word bears with the picture name (e.g., semantic, phonological, etc.) is manipulated and effects obtained are thought to inform researchers about processes involved in word production.

One specific effect has long been assumed to provide evidence about the nature of lexical selection: semantic interference (e.g., Damian & Martin, 1999; Levelt, Roelofs, & Meyer, 1999; Roelofs, 1992; Schriefers, Meyer, & Levelt, 1990; Starreveld & La Heij, 1996). This effect concerns the finding that response times (RTs) are longer for picture naming when the distractor is from the same semantic category as the picture (pictured cat, word *dog*) relative to unrelated distractors (pictured cat, word *pen*). A prominent account of this effect places it at the stage of lexical selection (e.g., Levelt et al., 1999). This account has been computationally implemented in several models, including the WEAVER++ model (Levelt et al., 1999; Roelofs, 1992, 2003, 2007, 2008a, 2008b) and the model of Starreveld and La Heij (1996).

The assumption that the semantic interference effect arises during lexical selection was recently challenged by Dell'Acqua, Job, Peressotti, and Pascali (2007). These authors used PWI as part of a psychological refractory period (PRP) procedure (Pashler, 1984, 1994) to determine at which stage the semantic interference effect emerged. With the PRP procedure, participants have to respond quickly and accurately to two stimuli (S1 and S2) in the right order, that is, the response to S1 has to be given before the response to S2. The stimulus onset asynchrony (SOA) between S1 and S2 is varied. A common finding in PRP experiments is that RTs for the second task increase as the SOA between S1 and S2 decreases, reflecting dual-task interference. The participants of Dell'Acqua et al. performed a manual tone discrimination task (Task 1), followed by a PWI task (Task 2) with distractor words semantically related or unrelated to the picture, using SOAs of 100, 350 or 1000 ms. The authors observed a semantic interference effect and an SOA effect, that is, picture-naming RTs increased as SOA decreased. Moreover, they also observed that the effects of SOA and stimulus type (semantically related or unrelated to the picture) were underadditive, that is, the semantic interference effect was smaller at the 350-ms SOA (23 ms) than at the 1000-ms SOA (68 ms), and absent at the 100-ms SOA (-7 ms). These findings were replicated by Ayora and colleagues (2011) using SOAs of 100 and 1000 ms and by Van Maanen, Van Rijn, and Taagten (2012, Experiment 1), using SOAs of 100, 350, and 800 ms.

This underadditivity of the effects of SOA and stimulus type on mean naming RTs was explained by Dell'Acqua and colleagues following the dominant model of PRP performance in the literature, which assumes that, in the context of overlapping tasks, response selection constitutes a processing bottleneck (Pashler, 1984, 1994). That is, only one response can be selected at a time. Thus selecting a response for Task 2 (PWI) has to wait until a response for Task 1 (tone discrimination) has been selected. This waiting period is known as *slack* (Schweickert, 1980). When there is enough time between the two tasks (i.e., the SOA between S1 and S2 is long), there is no overlap in selecting a response in each task, so an RT effect that is usually observed in single-task performance (e.g., semantic interference) is also observed in dual-task performance. Figure 1A depicts this situation assuming a response-selection bottleneck and a lexical response-selection locus of the semantic interference effect. Models of picture naming assume perceptual and conceptual encoding, lexical selection, word-form encoding, and articulation as the processing stages (e.g., Levelt et al., 1999; Roelofs, 2003). Lexical selection in models of picture naming corresponds to response selection in models of dual-task performance (e.g., Roelofs, 2007, 2008a). This is also assumed by Dell'Acqua and colleagues. In the remaining of the present article, we denote perceptual and conceptual encoding as *pre-selection* stages, lexical selection as *response selection*, and word-form encoding and articulation as *post-selection* stages.

At short SOAs, Task 2 effects that emerge during or after the response-selection bottleneck should be observed in the RTs. If the semantic interference effect in Task 2 arises during response selection, there will be no slack to absorb the effect. Consequently, semantic interference should be of similar magnitude at short and long SOAs. This situation of additivity of effects is depicted in Figure 1B (for a short SOA of 0 ms). In contrast, if semantic interference in Task 2 occurs before response selection (i.e., during stages of perceptual and conceptual encoding), the effect will be "absorbed into slack" (Pashler & Johnston, 1998, p. 170). This situation is depicted in Figure 1C.

The absorption of Task 2 effects into slack corresponds to what was observed by Dell'Acqua et al. (2007) for the semantic interference effect, suggesting a pre-selection locus of the effect (i.e., during perceptual and conceptual stages). In contrast, using the classic colour-word Stroop task as Task 2 (i.e., naming the ink colour of incongruent or congruent colour words), Fagot and Pashler (1992, Experiment 7) found that the Stroop effect (longer RTs in the incongruent condition, e.g., *blue* printed in red ink, relative to the congruent condition, e.g., *red* printed in red ink) was of similar magnitude at short and long SOAs. This confirms earlier evidence that the Stroop effect arises during response selection (see MacLeod, 1991, for a review), which corresponds to the stage of lexical selection in models of word production (e.g., Roelofs, 2003). According to Dell'Acqua et al., the fact that semantic interference is absorbed into slack, whereas the Stroop effect is not, suggests that the semantic interference effect emerges during perceptual or conceptual processing (i.e., pre-selection). This observation challenges the account of Roelofs (2003) implemented in WEAVER++, which assumes that semantic interference and the colour-word Stroop effect both arise in lexical response-selection.

However, in a recent study, Schnur and Martin (2012) failed to replicate the underadditivity of stimulus type and SOA effects on the mean naming RTs. They conducted two experiments with different materials and slightly different experimental parameters than Dell'Acqua and colleagues. In both experiments, equivalent semantic interference effects were obtained at short and long SOAs (31 ms on average) following tone presentation requiring a manual response. Thus, Schnur and Martin obtained additive effects of SOA and stimulus type, compatible with Figure 1B. Furthermore, Piai and Roelofs (2013) also failed to replicate the underadditivity of stimulus type and SOA effects using the SOAs of 0 and 1000 ms[1]. These results suggest a response-selection or post-selection locus of semantic interference under the assumption of a response-selection bottleneck.

To summarise, whereas three experiments obtained underadditive effects of SOA and stimulus type (Ayora et al., 2011; Dell'Acqua et al., 2007; Van Maanen et al., 2012), three other experiments obtained additive effects (Piai & Roelofs, 2013; Schnur & Martin, 2012). The underadditivity of effects suggests a pre-selection locus of the semantic interference effect, whereas the additivity suggests a locus at lexical response-selection or a later stage. Given the prominent role played by the semantic interference effect in informing theories of language production (e.g., Abdel Rahman & Melinger, 2009; Janssen, Schirm, Mahon, & Caramazza, 2008; Levelt et al., 1999), this discrepancy in the literature needs to be resolved.

There are at least two possible explanations for the discrepancy between studies. First, it may be the case that these studies made a Type-I or Type-II error. A Type-II error in the experiments of Dell'Acqua and colleagues (Ayora et al., 2011; Dell'Acqua et al., 2007) and

Van Maanen et al. (2012) would involve a failure to detect a true full-blown semantic interference effect at the short SOA with their samples of participants. A Type-I error in the experiments of Schnur and Martin (2012) and Piai and Roelofs (2013) would involve the detection of a spurious full-blown semantic interference effect at the short SOA with their samples of participants. This explanation is, however, unlikely given that both additivity and underadditivity have been observed three times each.

The second possible explanation for the discrepancy lies in the nature of the processing bottleneck in dual-task performance. In the literature, the assumption of a structural response-selection bottleneck has been challenged (e.g., Hübner & Lehle, 2007; Israel & Cohen, 2011; Karlin & Kestenbaum, 1968; Lehle & Hübner, 2009; Leonhard & Ulrich, 2011; Meyer & Kieras, 1997; Miller, Ulrich, & Rolke, 2009; Navon & Miller, 2002; Pannebakker et al., 2011; Schumacher et al., 1999, 2001; Schvaneveldt, 1969; Tombu & Jolicœur, 2003). According to one alternative account, dual-task interference arises because response-selection processes require central attentional capacity, which may be shared between tasks (Tombu & Jolicœur, 2003). However, this account predicts additive effects of Task 2 response-selection manipulations and SOA (for extensive discussion, see Tombu & Jolicœur, 2003), and therefore cannot explain why some studies obtained additive effects (Piai & Roelofs, 2013; Schnur & Martin, 2012) and other studies observed underadditive effects (Ayora et al., 2011; Dell'Acqua et al., 2007; Van Maanen et al., 2012, Experiment 1). According to another alternative account, the locus of the bottleneck is strategically determined (e.g., Hübner & Lehle, 2007; Israel & Cohen, 2011; Lehle & Hübner, 2009; Leonhard & Ulrich, 2011; Logan & Gordon, 2001; Meyer & Kieras, 1997; but see Ruthruff, Johnston, & Remington, 2009; Ruthruff, Pashler, & Klaassen, 2001) rather than structural and immutable, as argued by Dell'Acqua et al. (2007) and Pashler (1984, 1994). That is, a bottleneck may, in principle, occur at any stage, depending on the amount of overlap between tasks that participants (strategically) allow for. The overlap of response-selection processes for the two tasks may lead to underadditive effects of the Task 2 response-selection manipulation and SOA (e.g., Karlin & Kestenbaum, 1968; Schumacher et al., 1999; Thomson, Watter, & Finkelshtein, 2010). It should be noted, however, that participants usually seem reluctant to select responses for Tasks 1 and 2 in parallel (e.g., often extensive practice with the two tasks is required), so that a response-selection bottleneck typically prevails in dual-task performance.

The strategic bottleneck account assumes that dual-task interference effects may differ between studies, because participants may differ in the strategic determination of the amount of overlap between Task 1 and Task 2 (i.e., the locus of the bottleneck stage), as proposed by Piai, Roelofs, and Schriefers (2011), Roelofs (2007, 2008a), and Roelofs and Piai (2011), following Meyer and Kieras (1997). If the semantic interference effect arises in lexical selection and the participants of Dell'Acqua et al. (2007), Ayora et al. (2011), and Van Maanen et al. (2012, Experiment 1) allowed overlap between response selection in the tone and PWI tasks, then underadditive effects of SOA and stimulus type should be obtained, as empirically observed. In contrast, if the participants of Schnur and Martin (2012), Fagot and Pashler (1992), and Piai and Roelofs (2013) did not allow temporal overlap between the response selection processes, then additive effects of SOA and stimulus type should be obtained, as empirically observed in these studies. Schnur and Martin (p. 306) acknowledged that the strategic bottleneck account of Piai et al. (2011) could provide an explanation for the discrepancy among studies. Moreover, to support such a strategic account, Schnur and Martin reported that participants who made more than 20% errors on Task 1 showed a tendency towards a pattern of underadditivity, possibly indicating differences in strategic scheduling of the tasks.

Recently, Kleinman (2013) proposed that a difference in phonological regularity of the distractor words between Dell'Acqua et al. (2007) and Schnur and Martin (2011), rather than a different locus of the bottleneck, caused the difference in semantic effects at short SOAs between studies. For phonologically regular words, the sequence of phonemes can be derived from the spelling by applying grapheme-phoneme correspondence rules, whereas for phonologically irregular words, this cannot be done. Whereas the spelling-to-sound mapping in Italian, the language used by Dell'Acqua et al., is regular, it is highly irregular for English, the language used by Schnur and Martin. According to Kleinman, at short SOAs, the phonologically regular distractor words of Dell'Acqua et al. could be processed concurrently with selecting a response for the tone, whereas the phonologically irregular distractors of Schnur and Martin could not. As a consequence, assuming a response-selection bottleneck and lexical response-selection locus of semantic interference, the distractor words were already processed before response selection in picture naming at short SOAs in the study of Dell'Acqua et al., eliminating semantic interference, whereas the distractor words were processed during response selection in picture naming in the study of Schnur and Martin, yielding semantic interference.

However, the spelling-to-sound mapping in Dutch, the language that we used (Piai & Roelofs, 2013), is also regular (Booij, 1995; Borgwaldt, Bolger, & Jakab, 2010; Bosman, De Graaff, & Gijsel, 2006; Kerkhoff, Wester, & Boves, 1984; Nunn, 1998; Patel, Snowling, & de Jong, 2004; Seymour, Aro, & Erskine, 2003). In the study of Piai and Roelofs (2013), the distractor words were phonologically regular. Still, the semantic interference effect was clearly present at the short SOA (i.e., 0 ms), in disagreement with the phonological regularity account of Kleinman (2013). Nevertheless, Piai and Roelofs report only one experiment, and it is important to examine whether their findings can be replicated. In the first five experiments in the present article, the distractor words were phonologically regular, allowing for an examination of whether the underadditivity predicted by Kleinman is obtained or whether the additive findings of Piai and Roelofs are replicated.

The present study

Determining whether the semantic interference effect has a pre-selection (i.e., perceptual or conceptual) locus, as maintained by Dell’Acqua et al. (2007), or a locus at lexical-response selection or a later stage, as maintained by Schnur and Martin (2012), is important for our understanding of lexical access. The experiments of Ayora et al. (2011), Dell’Acqua et al. (2007), Fagot and Pashler (1992, Experiment 7), Kleinman (2013), Schnur and Martin (2012), Piai and Roelofs (2013), and Van Maanen et al. (2012) differ in several respects, including tasks, materials, SOAs, and stimulus types. The aim of the experiments reported in the present article was to examine whether any of these factors could have contributed to the difference in results between the earlier studies. Put differently, we investigate under which circumstances the additivity or underaddivity of the effects of SOA and stimulus type can be replicated, or whether additivity prevails regardless of the specific circumstances (suggesting a response-selection bottleneck and a response-selection or post-selection locus of the distractor effects).

Statisticians and investigators have pointed to the importance of replication of results for drawing theoretical conclusions (e.g., Cumming, 2008, 2012; Cumming & Maillardet, 2006; Fisher, 1966; Tukey, 1969). Cumming and Maillardet (2006) stated that “considering whether an effect is replicable is at the heart of drawing inferences from data.” (p. 217). Furthermore, although the additivity of the Stroop effect with SOA observed by Fagot and Pashler (Experiment 7) plays a crucial role in the theoretical argumentation of Dell’Acqua et al., there are no reported replications of this additivity in the literature.

We examined the discrepancy between the earlier studies of Ayora et al. (2011), Dell’Acqua et al. (2007), Fagot and Pashler (1992, Experiment 7), Kleinman (2013, Experiment 1), Schnur and Martin (2012), Piai and Roelofs (2013), and Van Maanen et al. (2012) in six new experiments manipulating various dimensions on which the earlier studies differed, including tasks (PWI, colour-word Stroop), materials (new materials vs. Dutch translations of the original materials used by Ayora et al.), stimulus types (related, unrelated, Stroop-like congruent, neutral), stimulus-set size (3, 32, 35), and SOAs (0, 100, 500, 1000 ms).

In Experiment 1, we directly compared PWI and colour-word Stroop task performance by having a single group of participants perform both tasks. In contrast, Dell'Acqua et al. compared PWI and Stroop task performance between different studies (i.e., Fagot & Pashler and themselves), which differed in several methodological respects. For example, Fagot and Pashler (Experiment 7) only had three colour-word stimuli presented in different conditions, whereas Dell'Acqua et al. had 48 picture stimuli. Moreover, relevant for the strategic bottleneck account (e.g., Meyer & Kieras, 1997; Piai et al., 2011; Roelofs & Piai, 2011; Schumacher et al., 1999), out of order responding (i.e., Task 2 responses occurring before Task 1 responses) was more likely to occur in Fagot and Pashler's study than in Dell' Acqua et al.'s study for two reasons. First, the SOA values used by Fagot and Pashler were shorter than the Task 1 mean RTs. Second, the experiment of Fagot and Pashler included congruent Stroop stimuli, which yield very short RTs. The higher probability of out-of-order responses could have invited the participants of Fagot and Pashler to adopt a more cautious scheduling strategy (i.e., adopting a response-selection rather than post-selection bottleneck), which may have yielded the additive effects in their study. In our Experiment 1, there were three pictures and three colours, presented in incongruent (e.g., pictured leg, word *arm*; colour red, word *green*), congruent (e.g., pictured leg, word *leg*; colour red, word *red*), and neutral conditions (e.g., pictured leg or colour red combined with five Xs). The SOA between tone and PWI or

Stroop stimulus was 0 or 500 ms. The use of the SOAs of 0 and 500 ms is similar to the values used by Fagot and Pashler (1992), whose longest SOA was 450 ms.

In Experiment 2, we omitted the Stroop task, increased the number of PWI stimuli to 32, and included an additional unrelated condition (e.g., pictured leg, word *train*), which allowed for the assessment of Stroop-like effects (incongruent distractor *arm* vs. congruent distractor *leg*) and semantic effects (related distractor *arm* vs. unrelated distractor *train*). In this way, the stimulus-set size and the stimulus types used are similar to Dell'Acqua et al. In Experiment 3, we omitted the congruent condition so that only semantically related and unrelated conditions were included in the experiment, exactly as in the experiment of Dell'Acqua et al. According to Van Maanen et al. (2012), the presence or absence of congruent stimuli in an experiment leads to, respectively, a widening or narrowing of attention to the distractor word, which should yield additive effects in our Experiment 2 and underadditive effects in our Experiment 3. Experiment 4 had the same distractor conditions as Experiment 3, but we replaced the SOA of 500 ms by a longer SOA of 1000 ms, which corresponds to the longest SOA used by Dell'Acqua et al. and Schnur and Martin. Thus, the SOA values used now were longer than the Task 1 mean RTs, presumably decreasing the probability of out of order responses relative to the 500-ms SOA.

In Experiments 1-4, the proportions of trials with short and long SOAs were the same. However, Dell'Acqua et al. used two short SOAs (100 and 350 ms) and one long SOA. This difference could be relevant given the demonstration by Miller et al. (2009) that, as the proportion of short SOAs increases in an experiment, participants tend to shift away from serial processing towards a more parallel mode of processing. Therefore, in Experiment 5, we doubled the number of 0-ms SOA trials, so that the proportion of short and long SOAs corresponded to the study of Dell'Acqua et al. In Experiments 1 to 5, the distractor words were phonologically regular, which should yield underadditive effects of distractor type and SOA, according to Kleinman (2013).

In addition to the design difference among studies that we discussed above, there were several other dimensions on which the previous studies differed. One such difference concerned response-set membership of the distractor words, which is an important variable in Stroop-like interference tasks (e.g., Lamers, Roelofs, & Rabeling-Keus, 2010; Piai, Roelofs, & Schriefers, 2012a). In Fagot and Pashler's (1992) study, the distractor words corresponded to responses in the experiment, whereas that was not the case in the studies of Dell'Acqua et al. (2007), Ayora et al. (2011), Kleinman (2013), Schnur and Martin (2012), and Van Maanen et al. (2012). Moreover, the number of tones used also differed among studies: two tones in Fagot and Pashler (Experiment 7) and Piai and Roelofs (2013) and three tones in the studies of Dell’Acqua et al., Ayora et al., Kleinman, Schnur and Martin, and Van Maanen et al. Therefore, Experiment 6 was a replication of the design of Ayora et al. with the materials translated into Dutch and with SOAs of 100 and 1000 ms (Schnur & Martin used English translations of the materials of Ayora et al.).

In order to allow for an easy comparison of the properties of the present experiments with those published in the literature, Table 1 gives an overview over the communalities and differences of the published experiments and of all experiments of the present article. In all experiments, we assessed whether the effects of SOA were additive or underadditive with the effects of Stroop or PWI stimulus type.

Experiment 1

Although the comparison between PWI and colour-word Stroop task performance played a critical role in the theoretical argumentation of Dell’Acqua et al. (2007), it is somewhat problematic, because their comparison is based on two studies (Dell’Acqua et al. 2007 and Fagot & Pashler, 1992) that differ not only in the task (PWI versus Stroop task), but also in a number of other potentially relevant aspects. For example, the comparison involved different groups of participants performing the Stroop experiment of Fagot and Pashler (1992) and the PWI experiment of Dell'Acqua et al. Moreover, in the Stroop experiment, three colour stimuli were used, requiring only three different responses, whereas there were 48 different responses in the PWI experiment. The distractor words in the PWI experiment were not part of the response set (i.e., they were not actual responses) whereas in the Stroop experiment, all written words corresponded to actual responses. These methodological differences could have affected the outcomes, as explained above, a possibility that is explicitly examined in Experiment 1.

We therefore directly compared PWI and colour-word Stroop task performance by having a single group of participants perform both tasks. Stroop experiments typically have three or four colour stimuli, which are constantly repeated, whereas PWI experiments usually have around 30 pictures, repeated only a few times (if repeated at all). In the present experiment, there were three pictures and three colours. The distractors in PWI were manipulated as to resemble typical Stroop experimental conditions: incongruent (e.g., pictured leg, word *arm*; colour red, word *green*), congruent (e.g., pictured leg, word *leg*; colour red, word *red*), or neutral conditions (e.g., pictured leg or colour red combined with five Xs). The SOA between tone and PWI or Stroop stimulus was 0 or 500 ms. Table 1 presents the experimental parameters of Experiment 1.

Method

Participants. Sixteen young adults (4 male, mean age = 20.1, sd = 2.3) from the participant pool of Radboud University Nijmegen participated in the experiment for course credits or monetary compensation. All participants were right-handed, native speakers of Dutch with normal or corrected-to-normal vision and normal hearing.

Materials and design. The picture stimuli were three black-and-white line drawings of the body parts leg, arm, and finger, taken from the picture gallery of the Max Planck Institute for Psycholinguistics, Nijmegen. In the congruent condition, these three pictures were presented with their Dutch basic-level names as distractors. These Dutch words (i.e., *been*, *arm*, and *vinger*) are phonologically regular (cf. Booij, 1995; Bosman et al. 2006; Kerkhoff et al., 1984; Nunn, 1998). The incongruent condition was formed by pairing the pictured leg with the distractor *finger*, the pictured finger with *arm*, and the pictured arm with *leg*. In the neutral condition, the three pictures were presented along with five Xs. The distractors were presented in white colour in lowercase Arial font, occupying on average 2.8º x 0.9º of visual angle at a viewing distance of approximately 60 cm, and the pictures were on average 5.7º x 5.7º of visual angle. The Stroop stimuli were the Dutch colour names for green, red, and blue (i.e., *groen*, *rood*, and *blauw*, all phonologically regular), printed in the corresponding ink colour, respectively, in the congruent condition, or printed in red, blue and green ink respectively in the incongruent condition. In the neutral condition, a series of five Xs was presented either in green, red or blue ink. The Stroop stimuli were presented in uppercase Arial font (on average 2.8º x 0.9º of visual angle). The pure tones were of 300 Hz (low tone) and of 800 Hz (high tone) and lasted 300 ms. The SOA values used were 0 ms and 500 ms, presented randomly across trials. Participants performed both PWI and Stroop in a blocked manner and the order of presentation of the two was counterbalanced across participants. Each picture-word and Stroop stimulus appeared six times with each tone at each SOA, totalling 432 trials. The two tones were presented randomly across trials. Trials were randomised using Mix (Van Casteren & Davis, 2006) with the constraints that the same tone, stimulus type and SOA did not appear on more than three consecutive trials. One unique list per participant was generated.

Procedure and apparatus. The presentation of stimuli and the recording of responses were controlled by Presentation Software (Neurobehavioral Systems, Albany, CA). The tones were presented via closed headphones and vocal responses were measured with a voice key. The button box was designed using Force Sensitive Resistors in order to make the button presses silent. Participants were instructed to rest the outer side of their left and right hands on the silent button box and to apply slight pressure with their index fingers on the buttons in order to make a response to the tones (left button – low tone; right button – high tone). Moreover, they were instructed to name the pictures and to try to ignore the distractor words, or to name the ink colour of the colour words. We emphasised that they should respond to the tone first, and should try to be fast and accurate in performing both tasks. Next, they were familiarised with the tones. A practice block of six trials of the paradigm they would see next (with different materials from the experimental ones), with the two SOAs presented randomly, preceded each experimental block.

At the 0-ms SOA, a trial began with the visual stimuli and the tone being presented simultaneously. At the 500-ms SOA, the tone was presented first, followed by the visual stimuli. The visual stimuli always remained on the screen for 1250 ms, followed by a black screen for 1750 ms. RTs were measured from stimulus onset (from tone stimuli onset for manual responses and from Stroop/PWI stimuli onset for vocal responses) and lasted until the end of the trial. The whole experimental session lasted approximately 30 minutes.

Analysis. Each trial had a manual response to the tone and a vocal response to the visual stimulus. First, all trials for which a vocal response was given before a manual response were discarded. Trials with manual RTs shorter than 100 ms and trials in which the voice key was triggered by a sound which was not the participant’s response or with vocal RTs shorter than 200 ms were discarded. Trials with incorrect tone classification were coded as errors and subsequently excluded from the RT analyses. Additionally, vocal responses which contained a disfluency, a wrong pronunciation of the word, or a wrong response word were also coded as errors and subsequently excluded. RTs were submitted to by-participant (F_1) repeated measures ANOVAs for each task separately (manual and vocal), with stimulus type (congruent, incongruent and neutral) and SOA (0 and 500 ms) as within-participant and within-item variables, and paradigm (PWI and Stroop) as within-participant and between-item variable (note that with only three items, by-item analyses of the naming RTs do not make sense). Errors were submitted to logistic regression analyses with stimulus type, SOA, and paradigm as predictors. For completeness, 95% confidence intervals (calculated from the variance over participants) and Cohen's *d* (calculated as the difference between two conditions divided by the squared root of their averaged variance, see Cumming, 2012) are provided in addition for the relevant effects (of stimulus type) of the naming responses. We compared both congruent and neutral stimuli to incongruent stimuli and refer to them below as Stroop-interference effects for the Stroop paradigm and as Stroop-like interference effects for the PWI paradigm.

Results

Figure 2 shows the RTs for the manual (Task 1) and vocal (Task 2) responses as a function of SOA and stimulus type for both the colour-word Stroop and the PWI paradigms.

Manual responses. Table 2 presents the error rates for the manual responses as a function of SOA and stimulus type. No predictor was significant in the logistic regression model, all *p*s > .05. For the RTs, there was a marginally significant main effect of stimulus type, $F_1(2,30) = 2.97$, $p = .066$. All remaining comparisons were not significant, all *F*s < 1.

Vocal responses. Table 2 also presents the error rates for the vocal naming responses as a function of SOA and stimulus type. Only stimulus type was a significant predictor in the logistic regression model. The log-odds of an incorrect response in the incongruent condition were 5.38 times higher than in the congruent condition, *ß coefficient* = 1.68, *S.E.* = 0.42, *Wald*

Z = -4.02, $p < .001$; and 6.24 times higher than in the neutral condition, *ß coefficient* = 1.83, *S.E.* = 0.44, *Wald Z* = 4.10, $p < .001$. For the RTs, there was no main effect of paradigm, $F_1(1,15) < 1$, indicating that overall performance was similar in both Stroop and PWI paradigms. SOA and paradigm did not interact, $F_1(1,15) < 1$. There was a main effect of SOA, $F_1(1,15) = 231.6$, $p < .001$; and of stimulus type, $F_1(2,30) = 51.66$, $p < .001$. Crucially, stimulus type and SOA did not interact, $F_1(2,30) < 1$, indicating that the magnitude of the interference effects was similar for both SOAs, that is, they were additive with SOA. Stimulus type and paradigm interacted, $F_1(2,30) = 7.46$, $p = .002$, indicating that the interference effects for the Stroop paradigm were larger than the Stroop-like effects in PWI. Importantly, the Stroop-like effects for the PWI task were significant, incongruent vs. congruent, $t_1(15) = 5.50$, $p < .001$, *95%CI* [55, 128], $d = .17$; incongruent vs. neutral, $t_1(15) = 5.53$, $p < .001$, *95%CI* [55, 126], $d = .15$; and so were the Stroop effects, incongruent vs. congruent, $t_1(15) = 9.76$, $p < .001$, *95%CI* [113, 175], $d = .33$; incongruent vs. neutral, $t_1(15) = 9.42$, $p < .001$, *95%CI* [114, 181], $d = .31$. The three-way interaction between stimulus type, SOA and paradigm was not significant, $F_1(2,30) < 1$.

Discussion

The results of Experiment 1 are clear: Additive effects of SOA and stimulus type were obtained in naming responses for both the Stroop and PWI paradigms. Overall performance was similar in both Stroop and PWI paradigms. The additive effects of SOA and stimulus type in the Stroop task on Task 2 RTs correspond to what Fagot and Pashler (1992, Experiment 7) observed. Furthermore, the additive effects of SOA and stimulus type for Task 2 RTs in the PWI task correspond to what Schnur and Martin (2012) and Piai and Roelofs (2013) observed for the semantic interference effect, but it differs from what Dell'Acqua et al. (2007), Ayora et al. (2011), and Van Maanen et al. (2012, Experiment 1) observed. The additivity of the effects of stimulus type and SOA suggests that the effects occurred at the response-selection stage or later (see Figure 1B), in disagreement with the proposal of a pre-selection locus by Dell'Acqua et al. Moreover, given that all written words were phonologically regular, the findings do not agree with the account of Kleinman (2013).

A somewhat surprising aspect of the present results is that an effect of stimulus type, albeit marginally significant, was obtained in the Task 1 RTs, especially at 500-ms SOA. Such an effect on Task 1 RTs could indicate that participants' performance in the present experiment differed from performance in the studies of Dell'Acqua et al. (2007) and Schnur and Martin (2012). If so, our pattern of additivity would have no bearing on the discussion regarding the locus of interference effects in dual-task performance. Furthermore, a small stimulus set, as in this experiment, is common for colour-word Stroop, but atypical for PWI experiments. Finally, with the stimulus types used (i.e., incongruent, congruent, and neutral), the Stroop-like effect can be examined, but semantic interference cannot be assessed. However, in the theoretical argumentation of Dell'Acqua et al. (2007), semantic interference played a central role. To address these issues, Experiment 2 was conducted.

Experiment 2

Experiment 2 was similar to Experiment 1, except that now only the PWI task was used, with a larger stimulus set and with conditions allowing us to test for semantic interference (semantically related vs. semantically unrelated distractors) and Stroop-like (semantically related vs. congruent distractors) effects in PWI. Table 1 presents the experimental parameters of Experiment 2.

If the additivity of effects of SOA and stimulus type in PWI was obtained in Experiment 1 only because of the small stimulus set and the large number of repetitions, a different pattern should be observed in the present experiment. If semantic interference in PWI arises before lexical response-selection, as argued by Dell'Acqua et al. (2007), the effect should be absent at the short SOA and present at the long SOA. However, if the effect arises in or after the response-selection bottleneck, then the effects of SOA and PWI stimulus type should be additive. Moreover, if Stroop-like effects obtained in PWI are similar to the semantic interference effect, a similar pattern should be observed for both effects.

Method

Participants. Twenty-one young adults (2 male, mean age = 20.9, sd = 2.2) participated from the same participant pool and with the same eligibility requirements as for Experiment 1. None of them had participated in the previous experiment.

Materials and design. The design was very similar to Experiment 1, but now only the PWI paradigm was used. Thirty-two pictures of common objects were selected from the same picture gallery as for Experiment 1. This stimulus set was chosen for having yielded reliable semantic interference effects in previous studies (e.g., Piai et al., 2011, 2012a; Piai & Roelofs, 2013). The objects belonged to eight different semantic categories with four objects per category. Each picture was paired with a semantically related distractor, forming the related condition. The unrelated condition was created by re-pairing the pictures with semantically unrelated distractors. In the congruent condition, the pictures were presented with their Dutch basic-level names as distractors. These Dutch distractor words were phonologically regular (cf. Booij, 1995; Bosman et al. 2006; Kerkhoff et al., 1984; Nunn, 1998). All distractors belonged to the response set. A list of the materials can be found in Appendix 1. Each picture-word stimulus appeared once with each tone at each SOA, totalling 384 trials. The two tones were presented randomly across trials. Trials were randomised using Mix (van Casteren & Davis, 2006) with the same constraints as for Experiment 1, with one unique list per participant.

Procedure, apparatus, and analysis. The procedure and apparatus were the same as in Experiment 1. The same inclusion criteria were used as for Experiment 1. Manual RTs were analysed in the same way as in Experiment 1. Naming RTs were submitted to by-participant (F_1) and by-item (F_2) repeated measures ANOVAs, with stimulus type (congruent, related and unrelated) and SOA (0 and 500 ms) as within-participant and within-item variables. Errors were submitted to logistic regression analyses with stimulus type and SOA as predictors. Cohen's *d* and 95% confidence intervals are reported in addition.

Results

Figure 3 shows the RTs for the manual (Task 1) and vocal (Task 2) responses as a function of SOA and stimulus type.

Manual responses. Table 3 presents the error rates for the manual responses in Experiment 2 as a function of SOA and stimulus type. SOA was a significant predictor in the logistic regression model: The log-odds of an incorrect response at the SOA 0 ms increased by a factor of 1, *β coefficient* = -.002, *S.E.* = .001, *Wald Z* = -2.9, $p = .003$. For the RTs, there was a main effect of SOA, $F_1(1,20) = 7.38$, $p = .013$ but no main effect of stimulus type, $F_1(2,40) = 1.07$, $p = .354$. The interaction between SOA and stimulus type was not significant, $F_1 < 1$. These results indicate that, overall, participants were slower in responding to the tones at the 500-ms than at the 0-ms SOA.

Vocal responses. Table 3 also presents the error rates for the vocal naming responses in Experiment 2 as a function of SOA and stimulus type. For the error percentages, the log-odds of an incorrect response in the related condition were 3.7 times higher than in the congruent condition, *ß coefficient* = -1.30, *S.E.* = .34, *Wald Z* = 3.79, $p < .001$. For the RTs, there was a main effect of SOA, $F_1(1,20) = 265.3$, $p < .001$, $F_2(1,31) = 1185.0$, $p < .001$, and of stimulus type, $F_1(2,40) = 8.29$, $p < .001$, $F_2(2,62) = 29.2$, $p < .001$. Stimulus type and SOA did not interact, *F*s < 1, indicating that the distractor effects were similar at both SOAs, i.e., they were additive with SOA. The Stroop-like effect (congruent vs. related) was significant, $t_1(20) = 4.26$, $p < .001$, *95%CI* [30,88], $d = .17$, $t_2(31) = 7.51$, $p < .001$, and so was the semantic interference effect (related vs. unrelated), $t_1(20) = 6.06$, $p = .002$, *95%CI* [16, 60], $d = .12$, $t_2(31) = 4.26$, $p < .001$.

Discussion

Experiment 2 was more similar to that of Dell'Acqua et al. (2007) regarding the stimulus-set size, although we used the congruent condition in addition to the semantically related and unrelated conditions. As in Experiment 1, we observed that the stimulus type effects were additive with SOA for the naming responses, similar to what Schnur and Martin (2012) obtained, but different from Dell'Acqua et al.'s results. The additivity of the effects of stimulus type and SOA suggests that the semantic and Stroop-like interference occurred at the response-selection stage or later, which challenges the proposal of a pre-selection locus by Dell'Acqua et al. Moreover, given that the distractor words were phonologically regular, the findings are not in agreement with the account of Kleinman (2013).

An unexpected aspect of the data is the finding that Task 1 RTs were shorter at the short compared to the long SOA. One possible explanation for this pattern is that participants grouped their responses for Tasks 1 and 2 (e.g., Sanders, 1964, 1998). That is, the Task 1 response is not executed as soon as it is ready, but it is withheld until the Task 2 response is ready. Grouping is, however, unlikely to account for the SOA effect in Task 1 RTs. If participants group their responses, the difference in RTs between Tasks 1 and 2 should be relatively small (i.e., around 100-200 ms, e.g., Miller & Ulrich, 2008; Sanders, 1964). Contrary to this prediction, differences in RTs at the 0-ms SOA were around 500 ms. It cannot be the case that participants prepared the Task 1 response and waited to group it with the Task 2 response, while still obtaining a difference of around 500 ms between the two tasks. Importantly, response grouping does not seem to affect the predictions of a standard bottleneck model with respect to Task 2 RTs (Ulrich & Miller, 2008). That is, the additivity observed in the present experiment should be obtained even if participants grouped their responses.

In sum, we observed additive effects of SOA and stimulus type for the naming responses, contrary to what Dell'Acqua et al. (2007), Ayora et al. (2011), and Van Maanen et al. (2012, Experiment 1) obtained. However, different from these studies, we had a congruent condition in the experiment. Under a strategic bottleneck model (e.g., Meyer & Kieras, 1997; Roelofs, 2007, 2008a), it is possible that the inclusion of this congruent condition affected participants' strategies. The congruent condition usually elicits shorter RTs than the related and unrelated conditions (e.g., Glaser & Düngelhoff, 1984), increasing the risk for participants to respond to the Task 2 PWI stimulus before responding to the Task 1 tone stimulus, especially at the 0-ms SOA. This could have made participants adopt a more conservative strategy (cf. Meyer & Kieras, 1997), allowing no overlap between response selection processes, causing the additivity we observed. Similarly, as argued by Van Maanen et al. (2012), the inclusion of congruent distractors may influence the amount of attention that participants allocate to the distractors. To see whether the congruent condition may have caused the difference in results between Dell'Acqua et al. (2007), Ayora et al. (2011), and Van Maanen et al. (2012, Experiment 1) and the present study, Experiment 3 was conducted.

Experiment 3

Van Maanen et al. (2012) demonstrated that the presence or absence of congruent distractors may affect whether additive or underadditive effects of distractor type and SOA are obtained. With only semantically related and unrelated distractors in an experiment, they observed that the semantic interference effect was underadditive with the SOA effect (their Experiment 1), replicating Dell'Acqua et al. (2007). However, when congruent distractors were added to the experiment (their Experiment 2), the magnitude of the interference effect was similar at the 100 and 800 ms SOAs. To investigate whether the additivity of Task 2 effects with SOA in our Experiment 2 was due to the inclusion of the congruent condition, this condition was omitted from Experiment 3. The rest of the experiment was identical to Experiment 2. Table 1 presents the experimental parameters of Experiment 3.

According to Van Maanen et al., we should now obtain underadditive effects of stimulus type and SOA because no congruent distractors appear in the PWI task, different from what we obtained in Experiment 2.

Method

Participants. Nineteen young adult participants (4 male, mean age = 20.5, sd = 2.4) from the same participant pool and with the same eligibility requirements as for Experiments 1 and 2 took part in the experiment. None of them had participated in the previous experiments.

Materials and design. The design was very similar to Experiment 2, except that only the related and unrelated conditions were used. Each picture-word stimulus appeared once with each tone at each SOA, totalling 256 trials. The two tones were presented randomly across trials. Trials were randomised using Mix (van Casteren & Davis, 2006) with the same constraints as for Experiment 1, with one unique list per participant.

Procedure, apparatus, and analysis. The procedure and apparatus were the same as for the previous experiments. The same inclusion criteria were used as for Experiments 1 and 2. Errors and manual and vocal RTs were analysed in the same way as in Experiment 2, with stimulus type including only the related and unrelated conditions.

Results

Figure 4 shows the RTs for the manual (Task 1) and vocal (Task 2) responses as a function of SOA and stimulus type.

Manual responses. Table 4 presents the error rates for the manual task as a function of SOA and stimulus type. SOA was a significant predictor in the logistic regression model:

The log-odds of an incorrect response at the SOA 0 ms increased by a factor of 1, *β coefficient* = -.002, *S.E.* = .000, *Wald Z* = -4.1, p < .001. For the RTs, there was a main effect of SOA, $F_1(1,18) = 12.7$, $p = .002$, and a marginally significant main effect of stimulus type, $F_1(1,18) = 3.4$, $p = .080$. The interaction between SOA and stimulus type was not significant, $F_1 < 1$. Thus, overall responses to the tone were longer at the 500-ms SOA than at the 0-ms SOA.

Vocal responses. Table 4 also presents the error rates for the vocal naming responses as a function of SOA and stimulus type. For the errors, the log-odds of an incorrect response in the related condition increases by a factor of 1.64 relative to the unrelated condition, *ß coefficient* = .492, *S.E.* = .211, *Wald Z* = 2.33, $p = .019$. For the RTs, there was a main effect of SOA, $F_1(1,18) = 172.7$, $p < .001$, $F_2(1,31) = 1038.0$, $p < .001$, and of stimulus type, $F_1(1,18) = 81.11$, $p < .001$, *95%CI* [31, 54], $d = .14$, $F_2(1,31) = 23.7$, $p < .001$. Stimulus type and SOA did not interact, *F*s < 1. Altogether, these results indicate that the semantic interference effect was of similar magnitude across SOAs, that is, additive with SOA.

Discussion

In this experiment, we used the semantically related and unrelated conditions only, exactly as Dell'Acqua et al. (2007) did. Yet, we still observed additive effects of SOA and stimulus type in the naming responses, replicating the pattern of results of Schnur and Martin (2012) and Piai and Roelofs (2013). Thus we did not replicate Dell'Acqua et al. (2007), Ayora et al. (2011), and Van Maanen et al. (2012, Experiment 1), who observed underadditive effects of SOA and stimulus type (with semantically related and unrelated distractors only). However, another difference between Dell'Acqua et al.'s (2007) design and the present experiments concerns the SOAs used. Whereas Dell'Acqua et al. used SOAs of 100, 350, and 1000 ms, we used SOAs of 0 and 500 ms. This difference in SOAs could be important for the following reason. The mean RTs for the manual tone-discrimination task (Task 1) were around 600-700 ms, both in the experiment of Dell'Acqua et al. and in our first three experiments. This means that the two SOA values that we used are smaller than the mean RTs of Task 1, whereas this does not hold for Dell'Acqua et al., who had one SOA (1000 ms) larger than the Task 1 mean RTs. Thus in our case, at both SOAs, Task 2 stimuli were presented, on average, before participants had completed Task 1. The likelihood of Task 2 responses preceding Task 1 responses is higher in this case than in Dell'Acqua et al.'s case, which had an SOA longer than Task 1 mean RTs. This property of our design could have influenced participants' strategies to avoid out of order Task 2 responses (i.e., making them more conservative, cf. Meyer & Kieras, 1997), yielding the observed patterns of additivity.

In Experiment 4, we therefore used SOAs of 0 ms and 1000 ms (see also Ayora et al., 2011). If the additivity observed in our first three experiments was caused by the fact that the long SOA was always shorter than the average manual RTs, the effect of SOA and stimulus type should now be underadditive.

Experiment 4

This experiment was very similar to Experiment 3, except that the SOA of 500 ms was replaced by an SOA of 1000 ms. Table 1 presents the experimental parameters of Experiment 4.

Method

Participants. Sixteen young adults (2 male, mean age = 22.5, sd = 3.14) from the same participant pool and with the same eligibility requirements as for the other experiments participated in the experiment.

Materials and design. The design was very similar to Experiment 3, except that now we used the SOAs of 0 ms and 1000 ms between the tone stimulus and the PWI stimulus. Each picture-word stimulus appeared once with each tone at each SOA, totalling 256 trials. The two tones were presented randomly across trials. Trials were randomised in a fashion similar to Experiment 1.

Procedure, apparatus, and analysis. The procedure and apparatus were the same as for the other experiments. The same inclusion criteria were used as for the other experiments. The same analyses were conducted as for Experiment 3.

Results

Figure 5 shows the RTs for the manual (Task 1) and vocal (Task 2) responses as a function of SOA and stimulus type.

Manual responses. Table 4 presents the error rates for the manual responses as a function of SOA and stimulus type for Experiment 4. In the logistic regression model, SOA was a significant predictor: The log-odds of an incorrect response at the SOA 0 ms increased by a factor of 1, *β coefficient* = -.001, *S.E.* = .000, *Wald Z* = -2.4, p = .019. For the RTs, there was a main effect of SOA, $F_1(1,15) = 30.02$, $p < .001$. The effect of stimulus type was not significant, $F_1 < 1$. SOA and stimulus type did not interact, $F_1(1,15) = 3.31$, $p = .089$.

Vocal responses. Table 4 also presents the error rates for the vocal naming responses as a function of SOA and stimulus type for Experiment 4. In the logistic regression model, the log-odds of an incorrect response in the related condition increased by a factor of 1.45 relative to the unrelated condition, *ß coefficient* = .373, *S.E.* = .189, *Wald Z* = 1.97, $p = .049$. For the RTs, there was a main effect of SOA, $F_1(1,15) = 231.6$, $p < .001$, $F_2(1,31) = 2438.00$, $p < .001$, and of stimulus type, $F_1(1,15) = 10.8$, $p = .005$, *95%CI* [3, 55], $d = .07$, $F_2(1,31) = 6.4$, $p = .017$. Stimulus type and SOA did not interact, Fs < 1. Altogether, these results indicate that the magnitude of the semantic interference effect was similar at both SOAs. That is, the semantic interference effect was additive with the SOA effect.

Discussion

In Experiment 4, SOAs of 0 and 1000 ms were used. The difference between the short and long SOAs is similar to the difference between the short and long SOAs of 100 and 1000 ms used by Dell'Acqua et al. (2007). Using a long SOA of 1000 ms in our experiment, however, did not affect the pattern of results. As in the first three experiments, the effects of SOA and stimulus type were additive in the naming latencies. However, different from the earlier experiments, we now obtained an increase of Task 1 RTs at the short SOA. That is, the tone discrimination RTs were longer at the short (0 ms) than at the long (1000 ms) SOA, whereas no such increase was obtained for the short (0 ms) and long (500 ms) SOAs in Experiments 1 to 3. This indicates that the additivity of SOA and stimulus type effects is independent of whether a short-SOA increase is obtained in the Task 1 RTs (which was the case in the present experiment) or not (which was the case in the first three experiments).

Experiments 1 to 4 showed a pattern of additivity of stimulus type and SOA effects in naming latencies, arguing against Dell'Acqua et al.'s (2007) interpretation that the semantic interference effect emerges before lexical response-selection. There is, however, another aspect in the design used by Dell'Acqua et al. (2007) that is different from ours: Dell'Acqua et al. used two relatively short SOAs (100 and 350 ms) and one long SOA, whereas so far we have constantly used the same proportion of short and long SOAs in our experiments. This difference could be important given a demonstration by Miller and colleagues (Miller et al., 2009) that, as the proportion of short SOA increases, participants tend to shift away from serial processing towards a more parallel mode of processing. (However, Miller et al. used two manual tasks rather than manual responding and naming, so their observations need not generalise to our experimental situation.) By encountering twice as many trials with short than long SOAs, the participants of Dell'Acqua et al. could have had the tendency to engage in more parallel processing, allowing response selection in picture naming to temporally overlap with response selection for tone discrimination. Independent evidence that response-selection processes may overlap comes from previous PRP studies showing underadditive effects of Task 2 response-selection manipulations and SOA (e.g., Karlin & Kestenbaum, 1968; Schumacher et al., 1999; Thomson et al., 2010). If the participants of Dell'Acqua et al.

selected the picture name in parallel with the tone-discrimination response on a large number of trials, the underadditivity of semantic interference and SOA effects could be explained by the absorption of the interference effect into slack. Note that this account assumes that the semantic interference effect arises during response selection and that the response-selection bottleneck is strategically imposed rather than structural and immutable.

Experiment 5

In this experiment, which was very similar to Experiment 4, we used the SOAs of 0 and 1000 ms, but now we varied the proportion of SOAs such that the short SOA was presented more than twice as often as the long SOA. Importantly, we increased the number of 0-ms SOA trials rather than adding a different short SOA (e.g., 350 ms) to keep the experiment comparable with the previous ones (which also had only two SOA values). Adding another SOA value would have changed not only the proportion of short- and long-SOA trials, but also the number of SOA values used, making it more difficult to compare Experiment 5 with Experiments 1-4. In our Experiments 1-4 and Piai and Roelofs (2013), we observed additive effects of SOA and distractor type with two SOA values and the same proportion of short- and long-SOA trials. Using more short- than long-SOA trials, Dell'Acqua et al. obtained underadditive effects of SOA and distractor type, whereas Schnur and Martin obtained additive effects (as Kleinman and Van Maanen et al. did in some experiments). Experiment 5 examined whether (with our materials, design, and participant pool) different proportions of short- and long-SOA trials yield additive effects of SOA and distractor type (Schnur and Martin) or underadditive effects (Dell'Acqua et al.). If additive effects are obtained (replicating Schnur and Martin), this would indicate that this pattern of effects occurs regardless of whether the proportions of short- and long-SOA trials are the same (Experiments 1-4) or different (Experiment 5). Instead, if participants shift towards more parallel processing due to the higher probability of short than long SOAs (Miller et al., 2009), we may observe underadditive effects of stimulus type and SOA (as Dell'Acqua et al. did), since the semantic interference would be resolved in parallel with Task 1 processing. Table 1 presents the experimental parameters of Experiment 5.

Method

Participants. Sixteen young adults (all female, mean age = 19.4, sd = 1.6) from the same participant pool and with the same eligibility requirements as for the other experiments participated.

Materials and design. The SOAs of 0 ms and 1000 ms were used. The distractors were either related or unrelated to the picture. Each picture-word stimulus was presented five times in the experiment, totalling 320 trials. The two tones were presented randomly across trials, but equally often with each stimulus type. The 1000-ms SOA was used in 100 trials (50 from the related and 50 from the unrelated conditions) and the 0-ms SOA was used in 220 trials (110 trials from each stimulus type condition). Trials were randomised using Mix (van Casteren & Davis, 2006) with one unique list per participant.

Procedure, apparatus, and analysis. The procedure and apparatus were the same as for the other experiments. The same inclusion criteria were used as for the other experiments. The same analyses were conducted as for Experiment 3.

Results

Figure 6 shows the RTs for the manual (Task 1) and vocal (Task 2) responses as a function of SOA and stimulus type.

Manual responses. Table 5 present the error rates for the manual responses as a function of SOA and stimulus type for Experiment 5. SOA was a significant predictor in the logistic regression model: The log-odds of an incorrect response at the 0-ms SOA increased by a factor of 1, *β coefficient* = -.001, *S.E.* = .000, *Wald Z* = -5.1, $p < .001$. For the RTs, there was a main effect of SOA, $F_1(1,15) = 17.9$, $p < .001$. The effect of stimulus type was not significant, $F_1(1,15) = 1.8$, $p = .240$. SOA and stimulus type did not interact, $F_1 < 1$.

Vocal responses. Table 5 also presents the error rates for the vocal naming responses as a function of SOA and stimulus type for Experiment 5. No predictor was significant in the logistic regression model, all *p*s > .08. For the RTs, there was a main effect of SOA, $F_1(1,15) = 155.0$, $p < .001$, $F_2(1,31) = 2406.0$, $p < .001$, and of stimulus type, $F_1(1,15) = 19.2$, $p < .001$, *95%CI* [14, 55], $d = .09$, $F_2(1,31) = 25.9$, $p < .001$. Stimulus type and SOA did not interact, *F*s < 1. These results demonstrate that a semantic interference effect was obtained, which was of similar magnitude at long and short SOAs.

Discussion

In this experiment, we varied the proportion of short SOAs relative to the long SOAs. Following Miller et al. (2009), we hypothesised that the underadditivity obtained by Dell'Acqua et al. (2007) might be due to their use of two short SOA values (100 and 350 ms), making participants engage in parallel processing, (partly) resolving semantic interference concurrently with Task 1 processing. However, even with the inclusion of twice as many short SOA trials than long SOA trials, we still obtained additivity of semantic interference and SOA effects on the naming responses, in line with our Experiments 1 to 4, Schnur and Martin (2012), and Piai and Roelofs (2013), but different from Dell'Acqua et al. (2007). Thus the greater relative number of trials with short than long SOAs (Experiment 5) did not influence the pattern of additivity of SOA and stimulus type effects in our study, as the results were comparable with the previous experiments, which had the same number of short- and long-SOA trials. Ayora et al. (2011) also had the same number of short- and long-SOA trials, and they obtained underadditivity of semantic interference and SOA effects, similar to Dell'Acqua et al., who had relatively more short SOA trials than long SOA trials.

In Experiments 1 to 5, the distractor words were phonologically regular, which should yield underadditive effects of distractor type and SOA, according to Kleinman (2013). Nevertheless, in all our experiments, we obtained additive effects, in disagreement with the account of Kleinman.

There are, however, yet other differences between our experiments and the experiment of Dell’Acqua et al. (2007). These differences include the number of tones presented to participants (two in our case vs. three in their study), the pitch and duration of the tones, the number of SOAs used and their values (two SOAs in our case vs. three SOAs of 100, 350, and 1000 ms in Dell’Acqua et al.’s study), and the fact that our distractors were members of the response set whereas theirs were not. Perhaps, some of these differences may have affected strategic scheduling of processes, yielding the discrepancy in results. Therefore, Experiment 6 is a final attempt to replicate Dell’Acqua et al., Ayora et al. (2011), and Van Maanen et al. (2012, Experiment 1).

Experiment 6

Experiment 6 is our final attempt to obtain the pattern of underadditivity observed by Dell’Acqua et al. (2007) and later replications. In line with Experiments 1 to 5 and Ayora et al. (2011), we only used two SOAs. To approach the experiments of Dell’Acqua et al. and Ayora et al. as closely as possible, we used an SOA of 100 ms rather than the 0 ms used in Experiments 1 to 5. Since Dell’Acqua et al. did not report their materials, we used the materials reported in Ayora et al. translated into Dutch. This means that our design was as similar as possible to the design of Ayora et al. Table 1 presents the experimental parameters of Experiment 6.

Method

Participants. Sixteen young adult participants (all female, mean age = 18.56, sd = 1.67) from the same participant pool and with the same eligibility requirements as for the other experiments took part in the experiment.

Materials and design. We used the 35 picture names of Ayora et al. (2011), with the corresponding pictures taken from the database of the Max Planck Institute for Psycholinguistics, Nijmegen, or from our own database. Our distractor words (semantically related or unrelated to the pictures) were Dutch translations of the words reported by Ayora et al., so the distractors were not members of the response set. Since their distractor words were matched for frequency and length, we acquired frequency counts for our Dutch distractors from CELEX (Baayen, Piepenbrock, & van Rijn, 1993) and tested for differences in frequency and length between the semantically related and unrelated distractors, ts < 1. Each picture-word stimulus appeared twice at each SOA, totalling 280 trials. The second presentation of the stimuli followed the first presentation of all stimuli. Trials were randomised using Mix (van Casteren & Davis, 2006) using the same constraints as for Experiment 1, with one unique list per participant. As indicated, the SOAs of 100 ms and 1000 ms were used. The tones were pure tones of 300, 600, and 1200 Hz, lasting 50 ms, following Dell'Acqua et al. (2007) and Ayora et al. The three tones were combined randomly with the PWI stimuli and were presented at random across trials, but equally often with each stimulus type and at each SOA.

Procedure, apparatus, and analysis. The same apparatus was used as for the other experiments. The aspects of the procedure that differed with respect to the previous experiments are mentioned here. Participants were instructed to rest two fingers of their choice from one hand and one finger from the other hand on the buttons (left button – low tone; middle button – medium tone; right button – high tone). As in Dell'Acqua et al. (2007) and Ayora et al. (2011), each trial began with the presentation of a fixation cross for 1000 ms, followed by a black screen for 800 ms, followed by one of the three tones. At an SOA of 100 or 1000 ms, the visual stimulus was displayed. The same inclusion criteria were used as for the other experiments. The same analyses were conducted as for Experiment 3.

Results

Figure 7 shows the RTs for the manual (Task 1) and vocal (Task 2) responses as a function of SOA and stimulus type.

Manual responses. Table 5 present the error rates for the manual responses as a function of SOA and stimulus type for Experiment 6. The logistic regression model showed that the log-odds of an incorrect response for unrelated stimuli increased by a factor of 1.43 relative to related stimuli, *β coefficient* = .359, *S.E.* = .164, *Wald Z* = 2.2, p = .029. For the RTs, there was a main effect of SOA, $F_1(1,15) = 32.6$, $p < .001$, and a main effect of stimulus type, $F_1(1,15) = 25.6$, $p < .001$. The interaction between SOA and stimulus type was not significant, $F_1 = 1.9$, $p = .190$.

Vocal responses. Table 5 also present the error rates for the vocal naming responses as a function of SOA and stimulus type for Experiment 6. No predictors were significant in the logistic regression model, all *p*s > .100. For the RTs, there was a main effect of SOA, $F_1(1,15) = 154.0$, $p < .001$, $F_2(1,34) = 1563.7$, $p < .001$, and of stimulus type, $F_1(1,15) = 39.8$, $p < .001$, *95%CI* [32, 72], $d = .13$, $F_2(1,34) = 12.4$, $p = .001$. Stimulus type and SOA did not interact, $F_1(1,15) = 3.03$, $p = .102$, $F_2(1,34) = 2.8$, $p = .104$. These results indicate that a semantic interference effect was present in the data, with similar magnitude across SOAs, that is, the effect was additive with SOA.

Discussion

The results of Experiment 6, which was conducted with the materials of Ayora et al. (2011) translated into Dutch, showed additivity of the effects of stimulus type and SOA in the naming latencies, similar to Schnur and Martin (2012), Piai and Roelofs (2013), and our

Experiments 1 to 5, but different from the results of Dell'Acqua et al. (2007), Ayora et al. (2011), and Van Maanen et al. (2012, Experiment 1). Thus, the differences in SOA values, in the number of tones, their pitch and duration, and in response-set membership do not seem to be factors modulating the patterns of additivity we have obtained with our experiments thus far. Importantly, as Experiment 6 shows, the additivity observed in our experiments using the short SOA of 0 ms (Experiments 1-5) is also observed when the short SOA is 100 ms, which was the SOA used by Dell'Acqua et al., Ayora et al., and Van Maanen et al.

A main effect of stimulus type was found at both SOAs in the manual RTs. The effect of stimulus type on manual RTs at the 1000 ms SOA may seem impossible at first sight since participants responded, on average, within 605 ms. So at an SOA of 1000 ms, they cannot have seen the Task 2 stimulus before responding in Task 1 and, therefore, no effects of a manipulation in Task 2 should be present in Task 1 responses. However, all the analyses reported here were performed on untrimmed data (cf. Miller, 1991; Ulrich & Miller, 1994). This means that, even though mean RTs for the manual task are around 605 ms, there are still many responses included in the analyses that were given after participants had seen Task 2 stimuli, that is, RTs larger than 1000 ms. To test this explanation, we left out of the analyses all manual RTs longer than 1000 ms and tested the effect of stimulus type at the 1000 ms SOA. This test showed that, once we only included the RTs of trials for which we know for sure participants did not see the Task 2 stimulus before responding, there was no longer an effect of stimulus type on Task 1 RTs, $t(15) = 1.53$, $p = .148$.

General Discussion

As outlined previously, the locus of the semantic interference effect in picture naming plays a pivotal role in guiding theories of language production (e.g., Dell'Acqua et al., 2007; Levelt et al., 1999; Miozzo & Caramazza, 2003; Roelofs, 1992). Based on underadditive effects of SOA and stimulus type on picture naming RTs in dual-task performance, Dell'Acqua et al. (2007) and Ayora et al. (2011) argued for a pre-selection locus of the semantic interference effect. However, in three experiments, Schnur and Martin (2012) and Piai and Roelofs (2013) obtained additive effects of SOA and stimulus type, arguing in favour of a locus at lexical response-selection or a later stage. On the basis of the experiments available in the literature, the pattern of results is inconclusive as three experiments show underaddivity of semantic and SOA effects (Ayora et al., Dell'Acqua et al., Van Maanen et al.) and three experiments show additivity of the effects (Schnur & Martin, and Piai & Roelofs). Given the importance of empirical replications to determine the robustness of an experimental finding (e.g., Cumming, 2008, 2012; Cumming & Maillardet, 2006; Fisher, 1966; Tukey, 1969), the present study aimed at manipulating the experimental design in various ways to examine which pattern of SOA and semantic effects in PWI under the PRP procedure is most robustly obtained.

The present results can be summarised as follows. The additivity of Stroop and SOA effects reported by Fagot and Pashler (1992, Experiment 7) was replicable and robust. Furthermore, in all six experiments, the magnitude of stimulus type effects was independent of SOA, and the additivity with SOA was obtained even though the distractors were phonologically regular (cf. Kleinman, 2013). This held regardless of the exact tasks (PWI, colour-word Stroop), materials (new, translations of Ayora et al.), stimulus types (related, unrelated, Stroop-like congruent, neutral), number of tones (two or three), and (proportion of) SOAs (0, 100, 500, 1000 ms) used[2]. Moreover, the additivity of the effects of SOA and stimulus type was obtained regardless of whether there was an SOA or a stimulus-type effect in the Task 1 RTs. Thus, with manual responding and naming, additivity of stimulus type and SOA effects appears to be a persistent pattern, replicable across variations of the experimental procedure. In contrast, the results of Van Maanen et al. (2012) suggest that underadditive effects of stimulus type and SOA are less robustly obtained. They observed that the underadditivity may disappear when congruent PWI stimuli are included in an experiment (although even this does not always happen, see our Experiment 3). The additive effects of distractor type and SOA provide evidence for a locus at response selection or later of the semantic and Stroop-like interference effects and a response-selection bottleneck in dual-task performance (see Figure 1B), whereby the response-selection bottleneck is either structural (Pashler, 1984, 1994) or strategic (Meyer & Kieras, 1997; Piai et al., 2011; Roelofs, 2007, 2008a; Roelofs & Piai, 2011). We further discuss this below.

Regarding the manual RTs, in Experiments 2 and 3, with SOAs of 0 and 500 ms, participants responded more quickly to the tones at the 0-ms SOA than at the 500-ms SOA. However, this pattern reversed in Experiments 4, 5 and 6, where 1000 ms was used for the long SOA: Responses were slower at the short SOAs (0 and 100 ms) than at the long SOA. We argued that this pattern of findings could not be explained by response grouping because the temporal lag between Task 2 and Task 1 responses is too long, i.e., 500 ms on average (cf. Miller & Ulrich, 2008). Importantly, whatever the pattern of results for Task 1 was, in all cases we observed additive effects of stimulus type and SOA for Task 2.

The nature of the processing bottleneck

The apparent malleability of the semantic interference effect (absent at short SOAs in the experiments of Dell’Acqua et al., 2007, Ayora et al., 2011, and Van Maanen et al., 2012, Experiment 1, and present in the experiments of Schnur & Martin, 2012, Piai & Roelofs, 2013, and in the experiments reported here) is difficult to reconcile with an immutable response-selection bottleneck in dual-task performance, as assumed by Dell’Acqua et al. Note that discrepant results emerging from dual-task investigations are not restricted to the present discussion. For example, the effect of practice on the magnitude of the dual-task interference is also different across studies (e.g., Karlin & Kestenbaum, 1968; Ruthruff, Johnston, Van Selst, Whitsell, & Remington, 2003; Schumacher et al., 2001; Van Selst & Jolicoeur, 1997;

Van Selst, Ruthruff, & Johnston, 1999). These and other findings challenge the assumption of a structural response-selection bottleneck in dual-task performance (e.g., Hübner & Lehle, 2007; Israel & Cohen, 2011; Karlin & Kestenbaum, 1968; Lehle & Hübner, 2009; Leonhard & Ulrich, 2011; Meyer & Kieras, 1997; Miller et al., 2009; Navon & Miller, 2002; Pannebakker et al., 2011; Schumacher et al., 1999, 2001; Schvaneveldt, 1969; Szameitat, Schubert, Müller, & Von Cramon, 2002; Szameitat, Lepsien, von Cramon, Sterr, & Schubert, 2006; Tombu & Jolicœur, 2003). Although under the strategic bottleneck account (Meyer & Kieras, 1997; Piai et al., 2011; Roelofs, 2007, 2008a; Roelofs & Piai, 2011), a response-selection bottleneck is optional rather than obligatory (i.e., response selection in Tasks 1 and 2 may, in principle, occur in parallel), the present findings suggest that participants seem to have a very strong preference for not overlapping response-selection processes in dual-task performance.

As mentioned previously, a powerful third alternative account of dual-task performance is that the bottleneck is not structural or strategic but rather arises from central capacity sharing (e.g., Tombu & Jolicoeur, 2003). The capacity sharing account assumes that dual-task interference occurs because response selection requires central capacity in order to proceed. If all capacity is first allocated to response selection in tone discrimination (Task 1) and then to response selection in PWI (Task 2), then the capacity sharing account would mimic the structural response-selection bottleneck account of Dell'Acqua et al. (2007). However, if capacity is divided between Tasks 1 and 2, response selection processes may overlap, just as may occur under the strategic bottleneck account. If capacity is shared between tasks, Task 1 RTs will be longer than when capacity is not shared. Thus, central capacity sharing may explain why sometimes Task 1 RT increases as SOA decreases, as observed in our Experiments 4-6 and in the experiments of Schnur and Martin (2012). This suggests that participant groups may differ in how central capacity is divided between the response selection stages in the two tasks (i.e., we obtained SOA effects on RT1 in some but not all of our experiments). However, the capacity-sharing account cannot explain the opposing data patterns in the literature (i.e., why the semantic interference effect is absent at short SOAs in the experiments of Dell'Acqua et al., 2007, Ayora et al., 2011, and Van Maanen et al., 2012, Experiment 1, and present in the experiments of Schnur & Martin, 2012, Piai & Roelofs, 2013, and in the experiments reported here). Tombu and Jolicœur (2003) demonstrated mathematically that if response selection requires central capacity, additive effects are predicted for experimental manipulations of Task 2 response selection and SOA, regardless of the division of capacity between tasks.

To recapitulate, structural and strategic bottleneck as well as central capacity sharing models can all explain the additive effects of stimulus type and SOA obtained in the present experiments and by Schnur and Martin (2012) and Piai and Roelofs (2013). However, only a strategic bottleneck account can accommodate the opposing patterns in the literature (i.e., the underadditive effects of Dell'Acqua et al., 2007, Ayora et al., 2011, and Van Maanen et al., 2012, Experiment1). The present findings suggest that participants strongly prefer imposing a response-selection bottleneck (yielding the pervasive additive effects) rather than a post-selection bottleneck (yielding the less-pervasive underadditive effects).

The skill of word reading

Participants may not only differ in their preferred bottleneck stage (i.e., response-selection vs. post-selection), but also in reading skill. Ruthruff, Allen, Lien, and Grabbe (2008) observed that reading skill may determine whether additive or underadditive effects are obtained in dual-task performance. Their Task 1 involved auditory or visual discrimination with manual responding and Task 2 involved visual lexical decision concerning high- or low-frequency words as well as nonwords. Ruthruff et al. observed that at short SOAs, a frequency effect was present in the Task 2 RTs for participants with poor reading skill, but the effect was absent for good readers. This suggests that good readers allowed for greater temporal overlap between Tasks 1 and 2 than poor readers.

Reading ability may also affect dual-task performance involving picture-word interference. Following the suggestions of Kleinman (2013) concerning phonological regularity (which were challenged by the results of our experiments), it is possible that distractor word processing (but not lexical response-selection) occurs concurrently with response selection in the tone task for good readers, eliminating semantic interference, whereas distractor word processing is delayed and overlaps with lexical response selection for poor readers, yielding semantic interference. Thus, a difference in reading ability may potentially explain the difference in results between Dell’Acqua et al. (2007), Ayora et al. (2011), and Van Maanen et al. (2012, Experiment 1), on the one hand, and those of Schnur and Martin (2012), Piai and Roelofs (2013), and in the experiments reported here, on the other hand.

However, this reading-skill account meets with a number of difficulties. First, given that we tested a great number of participants (all university students), it is unlikely that most of them were poor readers. Moreover, even if most of our participants were poor readers, it is unlikely that our distractor words were read poorly, because these were all highly familiar high-frequency words that were repeated several times during the experiments. Furthermore, Van Maanen et al. (2012) also used Dutch as the language of their experiments, just like the present study. Yet, the results of Van Maanen et al. and our results do not fully agree, contrary to what would have been predicted by Kleinman’s (2013) hypothesis regarding the phonological regularity of our stimuli. Most importantly, even if differences in reading ability could account for the differences in effects between studies, such an account would assume that the locus of the semantic interference effect is at the stage of lexical response-selection or later, which is the major conclusion we drew from the results of our experiments. Still, it would seem important for future studies to examine whether differences in reading ability can account for the variability of semantic effects at short SOAs.

The locus of the semantic interference effect

The additivity of the effects of SOA and stimulus type suggests that the semantic interference effect arises after the pre-selection stage of perceptual and conceptual processing (cf. Schnur & Martin, 2012), but it leaves open whether the effect occurs at the response-selection stage (e.g., Roelofs, 1992) or at the post-selection stage, close to articulation onset, as held by the response exclusion hypothesis (e.g., Janssen et al., 2008; Miozzo & Caramazza, 2003). However, it seems that the semantic interference effect can be localised to the response-selection stage by taking effects of phonological relatedness in dual-task performance into account. Whereas picture naming RTs are increased by semantic relatedness (i.e., the semantic interference effect), they are reduced by phonological relatedness (e.g., in naming the picture of a cat, RTs are shorter with distractor *cap* than with *arm*). According to the model proposed by Piai et al. (2011, see also Levelt et al., 1999; Roelofs, 1992, 1997, 2003, 2007, 2008a, 2008b), semantic interference arises in lexical response-selection and phonological facilitation arises during the subsequent post-selection stage of word-form encoding. In contrast, according to the response exclusion account (e.g., Janssen et al., 2008; Miozzo & Caramazza, 2003), semantic interference arises at the post-selection stage, during articulatory buffering, when a response to the distractor is excluded from the buffer, whereas effects of phonological relatedness occur also at the post-selection stage, but before rather than during articulatory buffering.

Ayora et al. (2011) examined the effect of semantic and phonological relatedness of distractors on picture naming RTs using the PRP procedure and a single group of participants. They obtained underadditive effects of SOA and semantic relatedness but additive effects of SOA and phonological relatedness. Under the strategic bottleneck account of Piai et al.

(2011), participants may or may not allow overlap between response selection in the tone and picture naming tasks. This implies that phonological effects should always be additive with SOA, as observed by Ayora et al. (2011), whereas semantic effects are additive (Schnur & Martin, 2012; Piai & Roelofs, 2013; present experiments) or underadditive (Dell'Acqua et al., 2007; Ayora et al., 2011; Van Maanen, 2012, Experiment 1) depending on whether overlap of response selection between tasks is allowed or not. In contrast, the response exclusion hypothesis fails to account for these findings.

The additivity of the effects of SOA and phonological relatedness obtained by Ayora et al. (2011) would suggest that the bottleneck is before the onset of phonological encoding (i.e., Ayora et al. assume a lexical response-selection bottleneck). However, according to the response exclusion hypothesis, given that the semantic interference effect arises after phonological encoding, during articulatory buffering, the effects of SOA and semantic relatedness also have to be additive, contrary to what Ayora et al. observed. Similarly, Ferreira and Pashler (2002) presented participants with PWI stimuli (Task 1) followed by tone discrimination (Task 2). They observed that the semantic interference effect from Task 1 propagated into Task 2 RTs whereas the phonological effect did not. The authors interpreted these effects as evidence that lexical response-selection is subject to a central processing bottleneck, whereas phonological encoding is not. According to the response exclusion account, the semantic interference effect arises after phonological encoding, thus the semantic effect should not propagate into Task 2 RTs, contrary to the empirical findings. To conclude, the present findings, taken together with those of Ayora et al. and Ferreira and Pashler, suggest that semantic interference arises in lexical response-selection, in line with modern psycholinguistic models of spoken word production (e.g., Abdel Rahman & Melinger, 2009; Damian & Martin, 1999; Levelt et al., 1999; Roelofs, 1992, 2003, 2007, 2008a, 2008b; Schriefers et al., 1990; Starreveld & La Heij, 1996).

Conclusion

To summarise, we obtained additive effects of SOA and stimulus type on picture naming RTs using the PRP procedure. The additivity was obtained regardless of the exact tasks, SOAs, materials, and distractor conditions used. Under structural or strategic response-selection bottleneck and central capacity sharing accounts of dual-task performance, the additivity of stimulus type and SOA effects in all our experiments argues against a pre-selection locus of semantic interference. However, the literature also reports underadditive effects. We concluded that only a strategic scheduling account can accommodate both the additive and underadditive effects. Moreover, the present results suggest that participants have a strong preference for imposing a strategic response-selection bottleneck. However, we have not been able to change this preference. This in turn implies that, as long as we have no clear means of explicitly manipulating potential strategies, conclusions from PRP performance regarding the locus of semantic interference remain tentative.

References

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*, 713-734.

Ayora, P., Peressotti, F., Alario, F.-X., Mulatti, C., Pluchino, P., Job, R., & Dell'Acqua, R. (2011). What phonological facilitation tells about semantic interference: A dual-task study. *Frontiers in Psychology, 2*(April), 1-10.

Baayen, R. H., Piepenbrock, R., & van Rijn, H. (1993). *The CELEX lexical database (CD-ROM)*. Linguistic Data Consortium, University of Pennsylvania, Philadelphia, PA.

Booij, G. E. (1995). *The phonology of Dutch*. Oxford, UK: Clarendon.

Borgwaldt, S., Bolger, P., & Jakab, E. (2010). Shallow versus deep footprints in pseudo-word grapheme-to-phoneme conversion: Dutch and English. *Journal of Germanic Linguistics, 22*, 425–443.

Bosman, A. M. T., De Graaff, S., & Gijsel, M. A. R. (2006). Double Dutch: The Dutch spelling system and learning to spell in Dutch. In R. M. Joshi and P. G. Aaron (Eds.), *Handbook of orthography and literacy* (pp. 135-150). Mahwah, NJ: Lawrence Erlbaum Associates.

Cumming, G. (2008). Replication and *p* intervals: *p* values predict the future only vaguely, but confidence intervals do much better. *Perspectives on Psychological Science, 3*, 286-300.

Cumming, G. (2012). *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis*. New York: Routledge.

Cumming, G., & Maillardet, R. (2006). Confidence intervals and replication: Where will the next mean fall? *Psychological Methods, 11*, 217-227.

Damian, M. F., & Martin, R. C. (1999). Semantic and phonological codes interact in single word production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*, 345-361.

Dell'Acqua, R., Job, R., Peressotti, F., & Pascali, A. (2007). The picture-word interference effect is not a Stroop effect. *Psychonomic Bulletin & Review*, *14*, 717-722.

Fagot, C., & Pashler, H. (1992). Making two responses to a single object: Implications for the central attentional bottleneck. *Journal of Experimental Psychology: Human Perception and Performance*, *18*, 1058–1079.

Ferreira, V., & Pashler, H. (2002). Central bottleneck influences on the processing stages of word production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*, 1187-1199.

Fisher, R. A. (1966). *The design of experiments* (8th ed.). Edinburgh, UK: Oliver & Boyd.

Glaser, W. R. (1992). Picture naming. *Cognition*, *42*, 61–105.

Glaser, W. R., & Düngelhoff, F. J. (1984). The time course of picture-word interference. *Journal of Experimental Psychology: Human Perception and Performance*, *10*, 640-654.

Hübner, R., & Lehle, C. (2007). Strategies of flanker coprocessing in single and dual tasks. *Journal of Experimental Psychology: Human Perception and Performance*, *33*, 103-123.

Israel, M., & Cohen, A. (2011). Involuntary strategy-dependent dual task performance. *Psychological Research*, *75*, 513-524.

Janssen, N., Schirm, W., Mahon, B. Z., & Caramazza, A. (2008). Semantic interference in a delayed naming task: Evidence for the response exclusion hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 249-256.

Karlin, L., & Kestenbaum, R. (1968). Effects of number of alternatives on the psychological refractory period. *Quarterly Journal of Experimental Psychology*, *20*, 167-178.

Kerkhoff, J., Wester, J., & Boves, L. (1984). A compiler for implementing the linguistic phase of a text-to-speech conversion system. In H. Benes & W.U.S. van Lesse Kloeke (Eds.), *Linguistics in the Netherlands* (pp. 111-117). Amsterdam: North-Holland Publishing Co.

Kleinman, D. (2013). Resolving semantic interference during word production requires attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition.*

Lehle, C., & Hübner, R. (2009). Strategic capacity sharing between two tasks: Evidence from tasks with the same and with different task sets. *Psychological Research*, *73*, 707-726.

Leonhard, T., & Ulrich, R. (2011). Determinants of central processing order in psychological refractory period paradigms: Central arrival times, detection times, or preparation? *Quarterly Journal of Experimental Psychology*, *64*, 2012–2043.

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

Logan, G. D., & Gordon, R. D. (2001). Executive control of visual attention in dual-task situations. *Psychological Review*, *108*, 393-434.

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin*, *109*, 163-203.

Mädebach, A., Oppermann, F., Hantsch, A., Curda, C., & Jescheniak, J. D. (2011). Is there semantic interference in delayed naming? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*, 522–538.

Meyer, D. E., & Kieras, D. E. (1997). A computational theory of executive cognitive processes and multiple-task performance: Part 1. Basic mechanisms. *Psychological Review*, *104*, 3-65.

Miller, J. (1991). Reaction time analysis with outlier exclusion: Bias varies with sample size. *Quarterly Journal of Experimental Psychology*, *43A*, 907-912.

Miller, J., & Ulrich, R. (2008). Bimanual response grouping in dual-task paradigms.

Quarterly Journal of Experimental Psychology, *61*, 888-1019.

Miller, J., Ulrich, R., & Rolke, B. (2009). On the optimality of serial and parallel processing in the psychological refractory period paradigm: Effects of the distribution of stimulus onset asynchronies. *Cognitive Psychology*, *58*, 273-310.

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.

Navon, D., & Miller, J. (2002). Queuing or sharing? A critical evaluation of the single-bottleneck notion. *Cognitive Psychology*, *44*, 193-251.

Nunn, A. M. (1998). *Dutch orthography: A systematic investigation of the spelling of Dutch words*. Den Haag, the Netherlands: Holland Academic Graphics.

Pannebakker, M. M. et al. (2011). Mental rotation impairs attention shifting and short-term memory encoding: Neurophysiological evidence against the response-selection bottleneck model of dual-task performance. *Neuropsychologia*, *49*, 2985-2993.

Pashler, H. (1984). Processing stages in overlapping tasks: Evidence for a central bottleneck. *Journal of Experimental Psychology: Human Perception and Performance*, *10*, 358-377.

Pashler, H. (1994). Dual-task interference in simple tasks: Data and theory. *Psychological Bulletin*, *116*, 220-244.

Pashler, H., & Johnston, J. C. (1998). Attentional limitations in dual-task performance. In H. Pashler (Ed.), *Attention* (pp. 155-189). Hove, UK: Psychology Press.

Patel, T. K., Snowling, M. J., & De Jong, P. F. (2004). A cross-linguistic comparison of children learning to read in English and Dutch. *Journal of Educational Psychology*, *96*, 785–797.

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

Piai, V., Roelofs, A., & Schriefers, H. (2011). Semantic interference in immediate and delayed naming and reading: Attention and task decisions. *Journal of Memory and Language*, *64*, 404-423.

Piai, V., Roelofs, A., & Schriefers, H. (2012a). Distractor strength and selective attention in picture naming performance. *Memory & Cognition, 40*, 614-627.

Piai, V., Roelofs, A., & Schriefers, H. (2012b). *Concurrent processing and semantic interference in picture naming*. Manuscript submitted for publication.

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

Roelofs, A. (1997). The WEAVER model of word-form encoding in speech production. *Cognition*, *64*, 249–284.

Roelofs, A. (2003). Goal-referenced selection of verbal action: Modeling attentional control in the Stroop task. *Psychological Review, 110*, 88-125.

Roelofs, A. (2007). Attention and gaze control in picture naming, word reading, and word categorizing. *Journal of Memory and Language, 57*, 232-251.

Roelofs, A. (2008a). Attention, gaze shifting, and dual-task interference from phonological encoding in spoken word planning. *Journal of Experimental Psychology: Human Perception and Performance, 34*, 1580-1598.

Roelofs, A. (2008b). Dynamics of the attentional control of word retrieval: Analyses of response time distributions. *Journal of Experimental Psychology: General*, *137*, 303-323.

Roelofs, A., & Piai, V. (2011). Attention demands of spoken word planning: A review. *Frontiers in Psychology, 2*(November), 1-14.

Ruthruff, E., Allen, P. A., Lien, M.-C., & Grabbe, J. (2008). Visual word recognition without central attention: Evidence for greater automaticity with greater reading ability. *Psychonomic Bulletin & Review, 15,* 337–343.

Ruthruff, E., Johnston, J. C., & Remington, R. W. (2009). How strategic is the central bottleneck: can it be overcome by trying harder? *Journal of Experimental Psychology: Human Perception and Performance*, *35*, 1368–1384.

Ruthruff, E., Johnston, J. C., Van Selst, M., Whitsell, S., & Remington, R. (2003). Vanishing dual-task interference after practice: Has the bottleneck been eliminated or is it merely latent? *Journal of Experimental Psychology: Human Perception and Performance, 29,* 280–289.

Ruthruff, E., Pashler, H. E., & Klaassen, A. (2001). Processing bottlenecks in dual-task performance: structural limitation or strategic postponement? *Psychonomic Bulletin & Review, 8,* 73–80.

Sanders, A. F. (1964). Selective strategies in the assimilation of successively presented signals. *Quarterly Journal of Experimental Psychology*, *16*, 368-372.

Sanders, A. F. (1998). *Elements of human performance: Reaction processes and attention in human skill*. London: LEA.

Schnur, T.T., & Martin, R. (2012). Semantic picture-word interference is a post-perceptual effect. *Psychonomic Bulletin & Review, 19*, 1-8.

Schumacher, E. H. et al. (1999). Current response-selection processes in dual-task performance: Evidence for adaptive executive control of task scheduling. *Journal of Experimental Psychology: Human Perception and Performance*, *25*, 791-814.

Schumacher, E. H. et al. (2001). Virtually perfect time sharing in dual-task performance: Uncorking the central cognitive bottleneck. *Psychological Science, 12*, 101-108.

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*, 86-102.

Schvaneveldt, R. (1969). Effects of complexity in simultaneous reaction time tasks. *Journal of Experimental Psychology*, *81*, 289-296.

Schweickert, R. (1980). Critical-path scheduling of mental processes in a dual task. *Science, 209*, 704-706.

Seymour, P. H. K., Aro, M., & Erskine, J. M. (2003). Foundation literacy acquisition in European orthographies. *British Journal of Psychology*, *94*, 143–174.

Starreveld, P. A., & La Heij, W. (1996). Time-course analysis of semantic and orthographic context effects in picture naming. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 896-918.

Szameitat, A. J., Schubert, T., Müller, K., & von Cramon, D. Y. (2002). Localization of executive functions in dual-task performance with fMRI. *Journal of Cognitive Neuroscience*, *14,* 1184-1199.

Szameitat, A. J., Lepsien, J., von Cramon, D. Y., Sterr, A., & Schubert, T. (2006). Task-order coordination in dual-task performance and the lateral prefrontal cortex: An event-related fMRI study. *Psychological Research*, *70,* 541-552.

Thomson, S., Watter, S., & Finkelshtein, A. (2010). Parallel response selection in dual-task situations via automatic category-to-response translation. *Attention, Perception, & Psychophysics, 72*, 1791–1802.

Tombu, M., & Jolicœur, P. (2003). A central capacity sharing model of dual-task performance. *Journal of Experimental Psychology: Human Perception and Performance, 29,* 3-18.

Tukey, J. W. (1969). Analyzing data: Sanctification or detective work? *American Psychologist, 24*, 83-91.

Ulrich, R., & Miller, J. (1994). Effects of truncation on reaction time analysis. *Journal of Experimental Psychology: General*, *123*, 34-80.

Ulrich, R., & Miller, J. (2008). Response grouping in the psychological refractory period (PRP) paradigm: Models and contamination effects. *Cognitive Psychology, 57*, 75-121.

van Casteren, M., & Davis, M. H. (2006). Mix, a program for pseudorandomization. *Behavior Research Methods*, *38*, 584-589.

Van Maanen, L., Van Rijn, H., & Taatgen, N. (2012). RACE/A: An architectural account of the interactions between learning, task control, and retrieval dynamics. *Cognitive Science*, *36*, 62-101.

Van Selst, M., & Jolicoeur, P. (1997). Decision and Response in Dual-Task Interference. *Cognitive Psychology, 33*, 266–307.

Van Selst, M., Ruthruff, E., & Johnston, J. C. (1999). Can practice eliminate the psychological refractory period effect? *Journal of Experimental Psychology: Human Perception and Performance*, *25*, 1268–1283.

Footnotes

1. Piai and Roelofs (2013) conducted a main PWI experiment with Stroop-like (related vs. congruent) and semantic (related vs. unrelated) manipulations and a control experiment with a semantic manipulation. They obtained additive effects of SOA and the Stroop-like manipulation. The effect of SOA and the semantic manipulation was overadditive in the main experiment and additive in the control experiment. Thus, overall, the effects of SOA and stimulus type were additive rather than underadditive.

2. The number of participants varied across experiments, but there was no profound reason for this. We planned to test 16 participants for Experiment 1. Experiment 2 addresses a discrepancy in the literature, hence we increased the planned number of participants to 20. Accidentally, our research assistant ran 21 participants instead. We opted for not excluding any participant, and this is why we report the data for $N = 21$. For Experiment 3, again we aimed for 20 participants, but we only managed to collect 19 participants before the beginning of the exams period. After two similar experiments (Experiments 2 and 3) replicating the same results, i.e., showing additivity of effects for nearly all 40 participants analysed, we assumed that the effect was powerful and consistent enough, so we did not need to increase statistical power by having many participants, hence we went back to 16 participants.

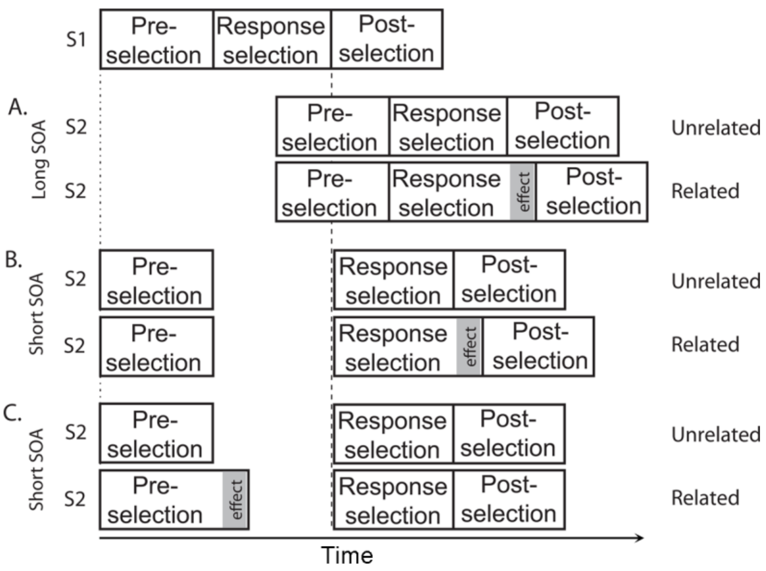

Figure 1. Schematic illustration of a lexical response-selection bottleneck account of the stimulus onset asynchrony (SOA) effect on semantic interference in dual-task performance, as proposed by Dell'Acqua et al. (2007). (A) At long SOAs, the semantic interference of distractor words in Task 2 picture naming is observed in the RTs, regardless of whether the locus of the effect is in response selection, as indicated, or earlier. (B) If the semantic interference effect arises during lexical response-selection, then at short SOAs (here SOA of 0 ms), it is not absorbed into slack and, thus, observed in the RTs. (C) If the semantic interference effect arises during perceptual/conceptual (pre-selection) processing, then at short SOAs (here SOA of 0 ms), it is absorbed into slack and, thus, not observed in the RTs. S1 = Stimulus 1. S2 = Stimulus 2.

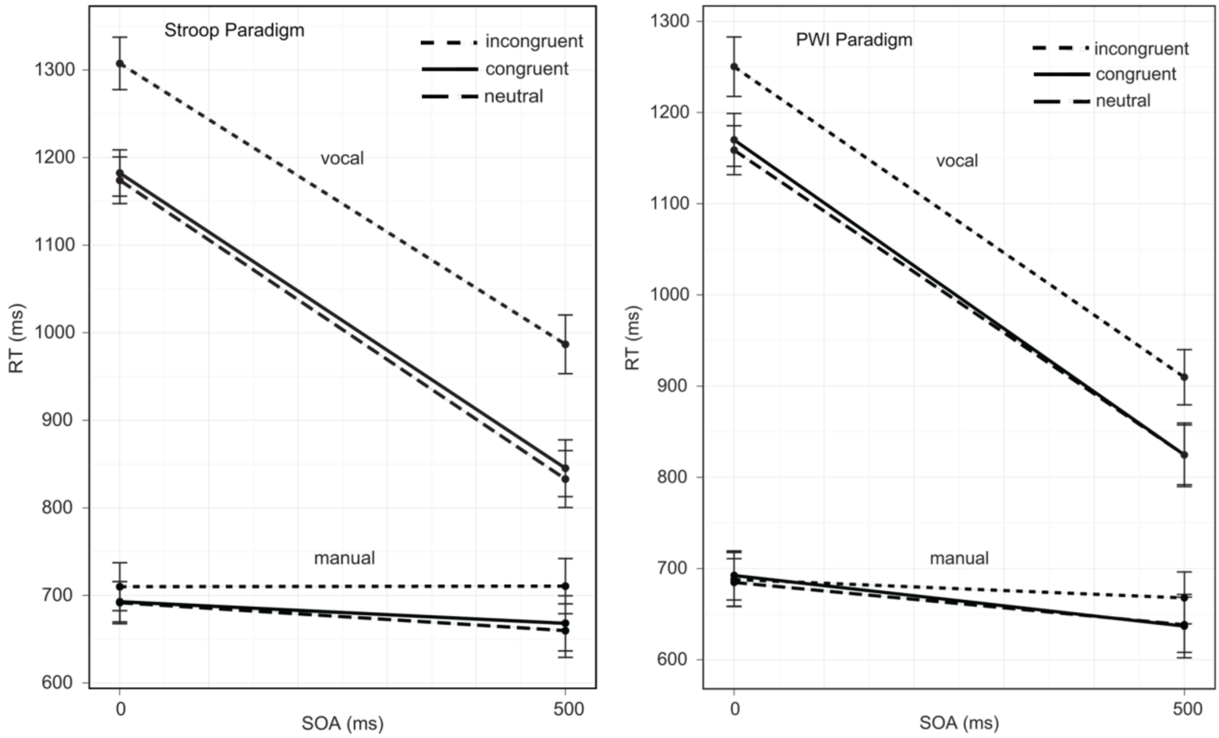

Figure 2. Manual (Task 1) and vocal (Task 2) response times (RTs) as a function of SOA and stimulus type for the Stroop paradigm (left) and the PWI paradigm (right) in Experiment 1. Error bars indicate 95% confidence intervals around the mean, calculated from the variance over participants. SOA = stimulus onset asynchrony, PWI = picture-word interference.

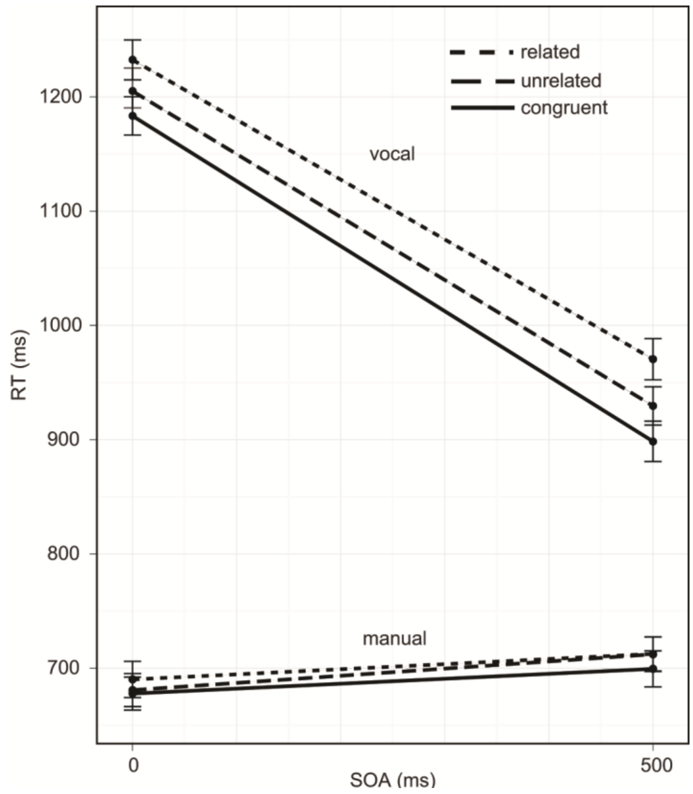

Figure 3. Manual (Task 1) and vocal (Task 2) response times (RTs) as a function of SOA and stimulus type for Experiment 2. Error bars indicate 95% confidence intervals around the mean, calculated from the variance over participants. SOA = stimulus onset asynchrony.

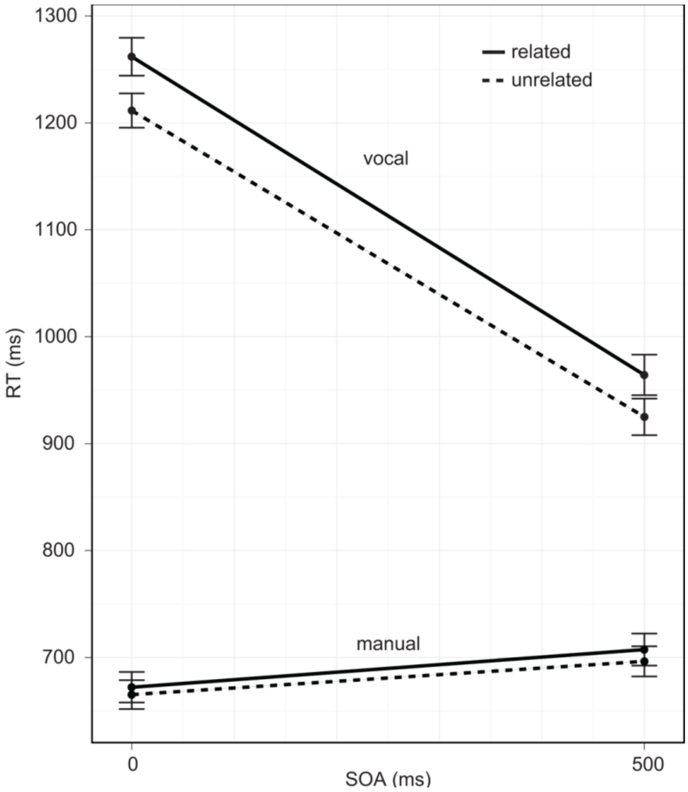

Figure 4. Manual (Task 1) and vocal (Task 2) response times (RTs) as a function of SOA and stimulus type for Experiment 3. Error bars indicate 95% confidence intervals around the mean, calculated from the variance over participants. SOA = stimulus onset asynchrony.

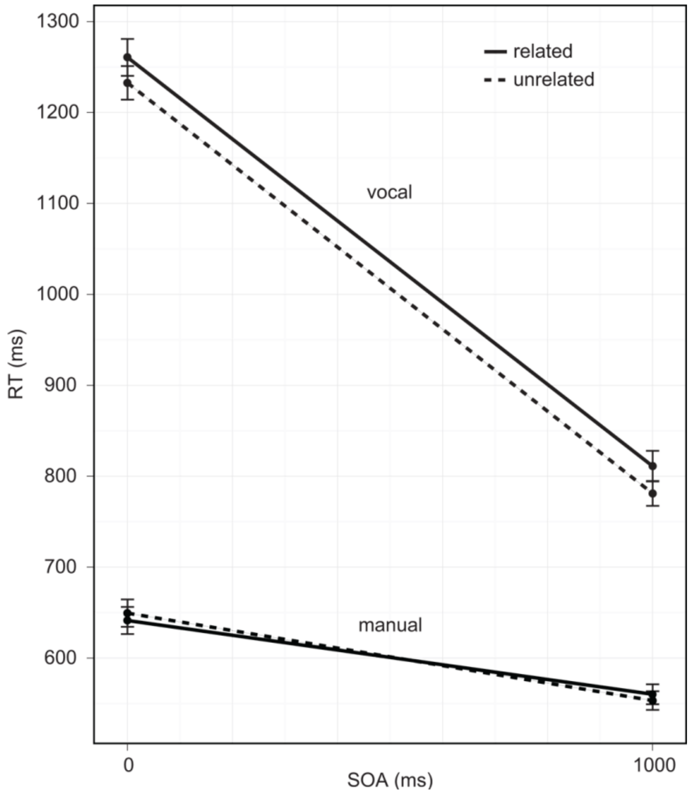

Figure 5. Manual (Task 1) and vocal (Task 2) response times (RTs) as a function of SOA and stimulus type for Experiment 4. Error bars indicate 95% confidence intervals around the mean, calculated from the variance over participants. SOA = stimulus onset asynchrony.

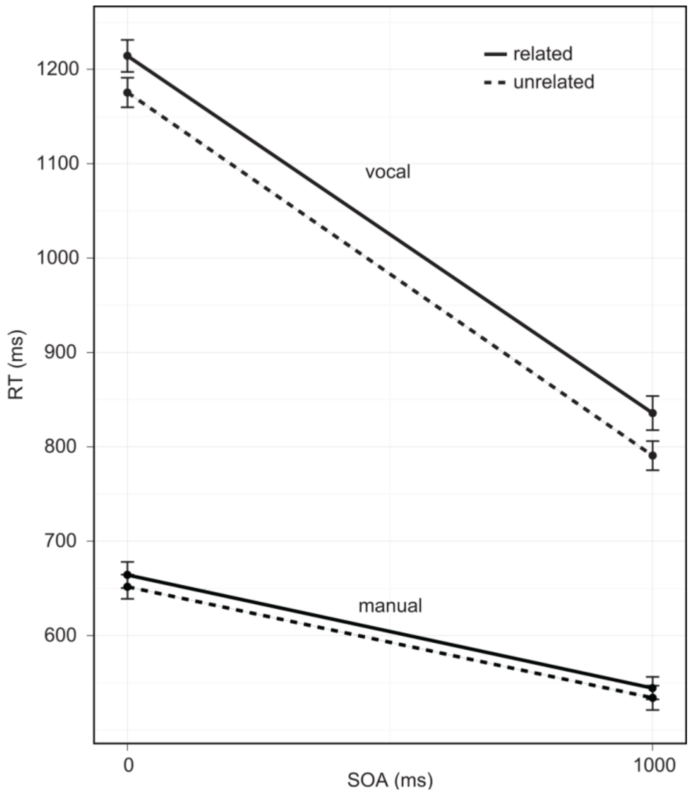

Figure 6. Manual (Task 1) and vocal (Task 2) response times (RTs) as a function of SOA and stimulus type for Experiment 5. Error bars indicate 95% confidence intervals around the mean, calculated from the variance over participants. SOA = stimulus onset asynchrony.

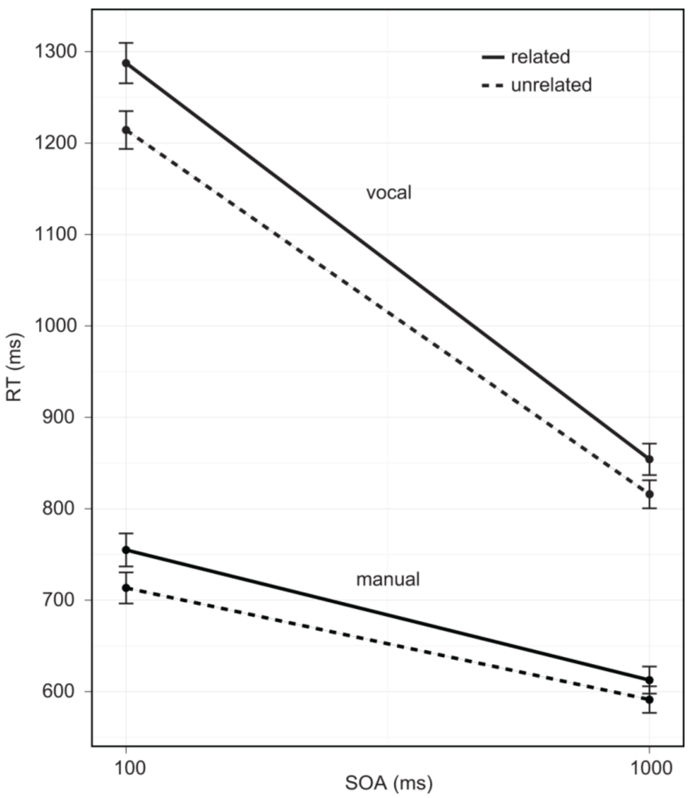

Figure 7. Manual (Task 1) and vocal (Task 2) response times (RTs) as a function of SOA and stimulus type for Experiment 6. Error bars indicate 95% confidence intervals around the mean, calculated from the variance over participants. SOA = stimulus onset asynchrony.

Table 1

Experimental Parameters of Previous Studies and of the Present Experiments. Exp = Experiment; Nr stim = Number of Stimuli; Picts = Pictures; Resp = Distractors in the Response Set; SOA = Stimulus Onset Asynchrony.

	Nr stim	SOA values	Resp	Stimulus types	Tones
Ayora et al. (2011)	35 picts	100, 1000	No	Related, unrelated, phonologically related	300, 600, 1200 Hz, 50 ms
Dell'Acqua et al. (2007)	48 picts	100, 350, 1000	No	Related, unrelated	300, 600, 1200 Hz, 50 ms
Fagot & Pashler (1992)	3 colours	-50, 50, 150, 450	Yes	Congruent, incongruent	300, 800 Hz, 300 ms
Kleinman (2013)	27 picts	100, 350, 1000	No	Related, unrelated	300, 600, 1200 Hz, 50 ms
Piai & Roelofs (2013)	32 picts	0, 1000	Yes	Related, unrelated, congruent	300, 800 Hz, 300 ms
Schnur & Martin (2012)	48 picts	100, 350, 1000 (Exp 1a), 1500 (Exp 1b)	No	Related, unrelated	300, 600, 1200 Hz, 50 ms
Van Maanen et al. (2012)	49 picts	100, 350, 800	No	Related, unrelated, congruent (Exp 2)	300, 600, 1200 Hz, 150 ms
			Present study		

Exp 1	3 picts, 3 colours	0, 500	Yes	Congruent, incongruent, neutral	300, 800 Hz, 300 ms
Exp 2	32 picts	0, 500	Yes	Related, unrelated, congruent	300, 800 Hz, 300 ms
Exp 3	32 picts	0, 500	Yes	Related, unrelated	300, 800 Hz, 300 ms
Exp 4	32 picts	0, 1000	Yes	Related, unrelated	300, 800 Hz, 300 ms
Exp 5	32 picts	0 (larger %), 1000	Yes	Related, unrelated	300, 800 Hz, 300 ms
Exp 6	35 picts	100, 1000	No	Related, unrelated	300, 600, 1200 Hz, 50 ms

1

Table 2

Error Rates (%) for the Manual (Task 1) and Vocal (Task 2) Responses as a Function of SOA, Stimulus Type, and Paradigm in Experiment 1. SOA = Stimulus Onset Asynchrony. PWI = Picture-Word Interference.

	Manual				Vocal			
Paradigm	Stroop	PWI	Stroop	PWI	Stroop	PWI	Stroop	PWI
SOA (ms)	0		500		0		500	
Stimulus type								
Congruent	4.7	4.6	2.7	3.1	0.6	1.3	1.1	0.6
Incongruent	5.4	6.0	2.4	1.1	6.3	6.5	5.7	4.8
Neutral	7.2	5.9	3.2	3.7	0.9	1.1	0.9	0.6

Table 3

Error Rates (%) for the Manual (Task 1) and Vocal (Task 2) Responses as a Function of SOA and Stimulus Type in Experiment 2. SOA = Stimulus Onset Asynchrony.

	Manual		Vocal	
SOA (ms)	0	500	0	500
Stimulus type				
Congruent	5.4	1.4	0.8	0.6
Related	3.9	1.6	2.9	4.1
Unrelated	4.0	2.0	3.1	1.6

Table 4

Error Rates (%) for the Manual (Task 1) and Vocal (Task 2) Responses as a Function of SOA and Stimulus Type in Experiments 3 and 4. SOA = stimulus Onset Asynchrony.

	Experiment 3				Experiment 4			
	Manual		Vocal		Manual		Vocal	
SOA (ms)	0	500	0	500	0	1000	0	1000
Stimulus type								
Related	4.1	1.8	2.6	3.2	2.6	1.8	3.9	4.2
Unrelated	4.4	2.0	1.4	2.2	3.7	1.8	2.3	3.6

Table 5

Error Rates (%) for the Manual (Task 1) and Vocal (Task 2) Responses as a Function of SOA and Stimulus Type in Experiments 5 and 6. SOA = Stimulus Onset Asynchrony.

	Experiment 5				Experiment 6			
	Manual		Vocal		Manual		Vocal	
SOA (ms)	0	1000	0	1000	100	1000	100	1000
Stimulus type								
Related	5.2	1.4	3.2	1.3	3.4	4.2	2.4	3.0
Unrelated	5.3	1.9	2.1	2.2	5.9	4.4	1.7	2.8

Appendix 1. Materials from Experiments 2, 3, 4, and 5 (English translations between parentheses).

	Picture Name	Related Distractor	Unrelated Distractor
Animals	hert (deer)	konijn	bureau
	konijn (rabbit)	hert	arm
	zwaan (swan)	geit	rok
	geit (goat)	zwaan	beker
Clothing	jas (jacket)	hemd	kasteel
	hemd (singlet)	jas	oor
	rok (skirt)	trui	zwaan
	trui (sweater)	rok	dolk
Transportation	auto (car)	bus	konijn
	bus (bus)	auto	glas
	trein (train)	fiets	kerk
	fiets (bicycle)	trein	kast
Buildings	kerk (church)	fabriek	been
	fabriek (factory)	kerk	neus
	molen (mill)	kasteel	kan
	kasteel (castle)	molen	jas
Weapons	dolk (dagger)	zwaard	trui
	zwaard (sword)	dolk	tafel
	kanon (cannon)	pistool	bord
	pistool (gun)	kanon	bed
Kitchen Wear	kan (pitcher)	beker	molen
	beker (cup)	kan	geit
	bord (plate)	glas	kanon
	glas (glass)	bord	bus
Furniture	bed (bed)	tafel	pistool
	tafel (table)	bed	zwaard
	bureau (desk)	kast	hert
	kast (wardrobe)	bureau	fiets
Body Parts	neus (nose)	arm	fabriek
	arm (arm)	neus	trein
	been (leg)	oor	auto
	oor (ear)	been	hemd