Bilingualism: Language and Cognition

cambridge.org/bil

Reverse cueing effects in L2 production: Another dissociation from conflict tasks

Giacomo Spinelli[1] and Simone Sulpizio[2]

[1]University of Zurich, Switzerland and [2]University of Milano-Bicocca, Italy

Research Article

Cite this article: Spinelli, G. and Sulpizio, S. (2026). Reverse cueing effects in L2 production: Another dissociation from conflict tasks. *Bilingualism: Language and Cognition* 1–13. https://doi.org/10.1017/S1366728926101278

Received: 4 May 2025
Revised: 29 January 2026
Accepted: 12 March 2026

Keywords:
adaptive control; cognitive control; bilingualism; bilingual advantage; cueing effect; cognate effect; L2 production

Corresponding author:
Giacomo Spinelli;
Email: giacomo.spinelli@uzh.ch

This research article was awarded Open Data and Open Materials badges for transparent practices. See the Data Availability Statement for details.

Abstract

Does L2 production involve adaptive control? Previous research drawing on a parallel between Stroop effects in L1 and cognate effects in L2 produced no support for this idea when inducing adaptive control implicitly (i.e., involuntarily). Reasoning that adaptive control might be hard to implement implicitly in L2 production, here, we induced adaptive control explicitly by presenting informative cues revealing whether the upcoming stimulus would be congruent/incongruent (in L1 Stroop) or cognate/noncognate (in L2 picture naming). Adaptive control was successfully induced in L1 Stroop, with informative cues, relative to uninformative ones, having a facilitatory effect. Such was not the case for L2 picture naming, in which informative cues had an inhibitory effect. While there might be several reasons for this reverse cueing effect, this finding represents another dissociation between L2 production and conflict tasks, which likely has implications for theories assuming a close connection between domain-general and bilingualism-specific control.

Highlights

- We induced adaptive control in L2 production explicitly by cueing cognate status.
- Relative to uninformative cues, informative cues slowed down L2 picture naming.
- Cueing congruency in L1 Stroop produced the regular (facilitatory) cueing effect.
- Results suggest a dissociation between L2 production and conflict tasks.
- The results are informative for theories assuming a close connection between the two.

For a few years now, the relationship between language and domain-general control has been at the forefront of the bilingualism research agenda, both for characterizing the processes involved in controlling multiple languages (e.g., Branzi et al., 2016; Calabria et al., 2012; Declerck et al., 2017, 2021) and for exploring the consequences of their routine use on cognition (e.g., in the form of a bilingual advantage in domain-general control: Bialystok & Craik, 2022). Prominent in this research have been switching paradigms, in which participants switch between tasks (in the task-switching version) or languages (in the language-switching version) typically based on a cue (Festman & Schwieter, 2015). Recently, however, it has been noted that some effects in language-switching paradigms, such as the switch cost and the N – 2 repetition cost, are modulated by language dominance in ways that are still difficult to predict (Gade et al., 2021; Koch et al., 2024). Further, those paradigms might not capture the typical bilingual experience (Blanco-Elorrieta & Pylkkänen, 2018).

To address these limitations, we have recently proposed that the focus of this line of research should be shifted to conflict paradigms, such as the Stroop (1935) task, which appear to better capture the natural dynamics of bilingual language use. In the Stroop task, participants are slower and, often, less accurate in naming the ink color of a word when that word is an incongruent color name (e.g., the word RED in the color blue) than a congruent one (e.g., RED in red) – the Stroop effect. The reason is that, although the word is nominally irrelevant (i.e., a distractor), it is inevitably processed, thus facilitating processing of the color (i.e., the target) when congruent and interfering with it when incongruent. Similarly, when bilinguals intend to produce the name of a concept in one of their languages (i.e., the target language), the name of that concept in the other language (i.e., the nontarget, or distractor, language) is highly active nonetheless (Green, 1998), especially when the target language is bilinguals' L2 and the nontarget language is their L1 (e.g., English and Italian, respectively, in Italian–English unbalanced bilinguals). Presumably, the result, similar to the Stroop effect, is facilitation when the two names have similar pronunciations (i.e., they are cognates, e.g., *elephant* and *elefante*) and interference when they have different pronunciations (i.e., they are noncognates; e.g., *dog* and *cane*) – the cognate effect (Costa et al., 2000). Although, with there being no clear option for a neutral baseline to separate facilitation from interference (unlike for the Stroop effect, for which a few options exist), the exact amount of each cannot be determined in the cognate effect, both facilitation and interference reflect a relative inability to select a target when a more salient distractor is inevitably processed, a selection difficulty that would be intrinsic to the normal bilingual experience for any model

© The Author(s), 2026. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

assuming between-language competition during production (e.g., Green, 1998; see also Green & Abutalebi, 2013). Because, further, the Stroop effect is also thought to index such a selection difficulty (MacLeod, 1991), that effect and the cognate effect would appear to provide a useful and representative test bed for examining how similar language and domain-general control are, and what implications a potential similarity would have for consequences of bilingualism on cognition.

Building on this rationale, we recently conducted two studies examining how the Stroop effect in an L1 Stroop color-naming task (henceforth, "L1 Stroop") and the cognate effect in an L2 picture-naming task (henceforth, "L2 picture naming") are affected by parallel manipulations aimed at demonstrating ADAPTIVE control – a form of control that is thought to play a key role in bilingualism and its broader consequences (Bialystok & Craik, 2022; Green & Abutalebi, 2013). Adaptive control refers to the ability to modulate aspects of control such as selection in order to accommodate the current context (Braem et al., 2019). Adaptive control, that is, is a dynamic aspect of control, to be distinguished from static ones such as the selection process itself. Manipulations aimed at demonstrating adaptive control typically contrast two contexts: one supposed to induce tighter control (typically in an anticipatory or proactive fashion) and one supposed to induce more relaxed control (with control typically being used in an as-needed or reactive fashion; Braver, 2012). Using two different manipulations of that class, we found both a dissociation and an association between Stroop and cognate effects. In the former case (Spinelli & Sulpizio, 2024), we used a proportion manipulation supposed to induce tighter selection when stimuli associated with interference (i.e., incongruent and noncognate stimuli in L1 Stroop and L2 picture naming, respectively) outweigh those associated with facilitation (i.e., congruent and cognate stimuli) compared with when the proportions are swapped. However, the expected interaction pattern (i.e., a smaller effect in the mostly interfering context) only emerged in L1 Stroop (similar to other interference tasks such as the flanker task, Bugg & Gonthier, 2020, and the Simon task, Wühr et al., 2015). In L2 picture naming, responses were overall slower in the mostly interfering context, but the cognate effect itself was unaffected – an additive pattern typical of nonconflict tasks (Spinelli et al., 2019). In the other case (Spinelli & Sulpizio, 2025), we had participants perform L1 Stroop and L2 picture naming concurrently with an *n*-back task, which was supposed to impair participants' ability to proactively tighten selection (Kalanthroff et al., 2015). Here, both L1 Stroop and L2 picture naming showed the same pattern, although not the expected one: Rather than increasing the Stroop or the cognate effect, load had exclusively an additive effect. We interpreted such a pattern as reflecting the fact that, at baseline, selection is handled reactively just as it has to be handled with a concurrent load.

Overall, despite the structural similarity between the Stroop and the cognate effect, our studies only provided mixed evidence that both effects are under adaptive control, with, in fact, no such evidence for the cognate effect. Importantly, however, both manipulations on which this conclusion is based are of the implicit kind (i.e., involuntary). For example, participants were not informed of the stimuli proportions in Spinelli and Sulpizio (2024), or whether they could proactively tighten selection in Spinelli and Sulpizio (2025). Although participants can become aware of these facts/possibilities during the experiment, whether they actually do so appears to have no implication for the resulting patterns (e.g., Blais et al., 2012), suggesting that explicit awareness is not necessarily involved in the underlying processes.

The same cannot be said of manipulations involving cues to the nature of the upcoming stimulus. Bugg and Smallwood (2016) introduced this type of manipulation in a Stroop task in which each stimulus was preceded by a cue that revealed either nothing (an uninformative cue) or the stimulus's congruent/incongruent status (an informative, 100% valid cue). This study and others (e.g., Jiménez et al., 2020, 2021) produced a cueing effect (faster latencies for informative than uninformative cues) for incongruent trials, but only if participants were specifically instructed to use the cues and these were salient, 100% valid and presented for a relatively long (2000 ms) duration. For congruent trials, a cueing effect emerged even when some (but not all) of these conditions were relaxed (e.g., with shorter cue durations) and was generally larger than for incongruent trials (for a similar pattern for repeat and switch trials in task switching, see, e.g., Koch, 2008), presumably because, while a very efficient word-reading strategy was possible for congruent trials, such a strategy was not possible for incongruent ones. In any case, these cueing effects, like the effects produced by proportion manipulations, also demonstrate that control can be adapted to optimize task performance; however, this type of adaptation must be explicit (i.e., voluntary), or else it would not be subject to such stringent boundary conditions, especially for incongruent trials.

In the present study, we contrasted Stroop and cognate effects focusing on this explicit kind of adaptive control. A possible explanation for the mixed pattern of our previous studies, and particularly for the failure to observe adaptive control with the cognate effect, is that, for that effect, no distractor is explicitly presented (i.e., participants simply name pictures in their L2 without being presented with their L1 names or knowing that the L1 would be relevant), unlike for the Stroop effect (which is based on an explicitly presented word; but see Kiyonaga and Egner, 2014). Although adaptation to distractors not physically present or of which there is no awareness seems to be possible in principle (Kiyonaga & Egner, 2014; Van Gaal et al., 2010), it might be somewhat difficult to implement, for example, because in such a situation interference may not be felt, and so the need to adapt to it may not arise (Desender et al., 2014). Such would not be the case, however, when individuals are instructed to recognize the role the L1 plays in generating facilitation and interference in L2 production and are given reliable information to anticipate them, as they would be with a cue manipulation. More generally, explicit manipulations might, on the one hand, boost adaptive-control processes that are available to bilinguals, but hard to implement with implicit manipulations. On the other hand, they might create the necessary conditions for such processes to be used (e.g., awareness of interference), without which adaptive control would not be possible.

Based on these considerations, we had Italian–English bilinguals perform an L1 Stroop task with a cue manipulation of proven efficacy and an L2 picture-naming task with a parallel manipulation. In both tasks, participants were first trained to recognize the difference between the two types of stimuli involved in the task (i.e., congruent versus incongruent in L1 Stroop and cognate versus noncognate in L2 picture naming). Then, they performed the task with either uninformative or informative cues to the upcoming stimulus's nature, with the instructions emphasizing the need to use those cues when informative. We analyzed the results with a focus on the pattern of cueing effects in the two tasks, but associations between tasks and with various indices of bilingual experience were also explored.

In sum, evidence thus far points to bilinguals being unable to optimize control over L1 facilitation/interference when they would

https://doi.org/10.1017/S1366728926101278 Published online by Cambridge University Press

have to do so implicitly; our study provides the first opportunity to determine whether they could do so explicitly. We tested this idea hypothesizing a beneficial effect of explicit cues in L2 picture naming, in addition to L1 Stroop, in parallel manipulations in the two tasks.

Note, however, that there are reasons to believe that these parallel manipulations might interact differently with the nature of the two tasks. For example, it might be argued that by directing participants' attention to the distractor, informative cues, compared with uninformative ones, might have the side effect of increasing the saliency of the distractor and, therefore, its potential to create additional facilitation and/or interference (in addition to any facilitation produced by the use of the information conveyed by the cues). The specific pattern of this additional facilitation/interference will depend on the type of stimulus and processing level involved. In the Stroop literature, for example, a distinction is typically made between the task, semantic, and response levels (i.e., the levels at which one must determine which task – color naming and reading – to perform, which concept to select, and which response to produce, respectively: Parris et al., 2022; but for a critical discussion of the task level, see Parris et al., 2023). For incongruent stimuli, interference can occur at any level. For congruent stimuli, facilitation can occur at the semantic and the response level, but interference can also occur at the task level (because with congruent stimuli, as with incongruent stimuli, both color naming and reading are possible and compete with one another). Similarly, in bilingual production, one can distinguish between language selection (i.e., which language to use, akin to the task level in the Stroop literature) and item selection (i.e., which translation equivalent to use, akin to the semantic and/or response levels in the Stroop literature; Van Assche et al., 2013). Also similar to the Stroop literature, noncognate stimuli likely suffer interference at both levels, whereas cognate stimuli likely enjoy facilitation at the item level while suffering interference at the language level (because, even with a cognate stimulus, one must resolve the competition between the two languages). Therefore, even if informative cues increase the distractor's saliency, based on this discussion, the impact on performance should be similar in L1 Stroop and L2 picture naming.

Importantly, however, the distractor's saliency at baseline (i.e., with uninformative cues) is likely lower in L2 picture naming than in L1 Stroop because in the former, but not in the latter, the distractor is not explicitly presented. As a result, relative to uninformative cues, informative cues might increase the distractor's saliency to a greater extent in L2 picture naming than in L1 Stroop, potentially resulting in larger additional facilitation/interference in that task. This difference and others to be discussed later on in the manuscript make the contrast between the two tasks perhaps less clear, but informative nonetheless because the low saliency of distractors at baseline in L2 picture naming represents the most typical situation in L2 production (i.e., the one in which a single language is relevant, with little or no need to monitor another). Further, even in L2 picture naming, while informative cues might strongly increase the distractors' saliency, they would still convey information that can help performance. Therefore, the present study allows us to address the nontrivial question of whether those cues' net effect is positive as it tends to be in L1 Stroop.

Method

Participants

The determination of the sample size was based on a power analysis conducted with G*Power 3.1 (Faul et al., 2009) using the effect size of the main effect of cue type reported by Bugg and Smallwood (2016) for the 2-s cue duration condition in their Experiment 1 ($\eta_p^2 = .24$), the Stroop task we reproduced in the present study. Although the minimum sample size suggested by the analysis for a .80 power to detect that effect size was 28, we aimed to reach a sample size comparable to that used in our previous studies (Spinelli & Sulpizio, 2024, 2025), that is, 48 participants. Participants were recruited from the same pool (i.e., the University of Milano-Bicocca community) and in the same fashion described for those studies (participants who already participated in those studies were prevented from participating again). Participants received course credits as compensation for their participation.

To participate, volunteers were required to consider Italian to be their native (or one of their native) language(s), to have normal or corrected-to-normal vision and hearing, to be between 18 and 45 years old and to pass an English prescreening test (see below). One hundred and twenty-one participants completed the prescreening test. Of these, 78 passed it, 55 came to the lab to complete the study and 48 remained after exclusions (see below). Of the final sample, 38 identified themselves as female, 8 male and 2 nonbinary; they were 22.33 years old on average (*SD* = 2.50, range = 18–28); 46 reported knowing a third language beside Italian and English and 18 a fourth language, although their proficiency, immersion and dominance in those languages (as calculated using the Language History Questionnaire 3.0 (LHQ3), Li et al., 2019) were lower than those reported for Italian or English on average; finally, all were born and resided in Italy. We report additional information on Italian (L1) and English (L2), the two languages involved in the study, in Table 1.

Materials and procedure

Prescreening session

As in Spinelli and Sulpizio (2024, 2025), participants were prescreened using Cambridge's online test for adult learners of English (https://www.cambridgeenglish.org/test-your-english/general-english/). The test provides an English proficiency estimate within the Common European Framework of Reference for Languages (CEFRL). To pass the test, participants were required to perform at an estimated B2 CEFRL level. To participate in the prescreening test and the subsequent lab session, participants expressed their informed consent. The study was evaluated by the local commission for minimal-risk studies of the Psychology Department at the University of Milano-Bicocca (protocol RM-2021-445).

Table 1. Characteristics of Italian (L1) and English (L2) for our participants

	Italian (L1)			English (L2)		
Characteristic	Mean	SD	Range	Mean	SD	Range
Proficiency	.97	.06	.75–1	.80	.09	.46–1
Immersion	.92	.03	.83–.98	.74	.06	.46–.82
Dominance	.63	.06	.47–.73	.47	.07	.26–.62
Prescreening score				21.33	2.07	18–25
Lexical fluency				65.67	9.54	45–85

Note: Proficiency, immersion and dominance are aggregated scores ranging from 0 to 1 calculated using the formulas in the LHQ3 (applying the corrections explained in Spinelli and Sulpizio, 2024; note that because of those corrections, dominance for the final sample could not be calculated for Italian in 14 cases and for English in one case). The prescreening score is the sum of correct responses to the 25 questions included in Cambridge's online test for adult learners of English, and lexical fluency is the number of correct L1-to-L2 translations provided for 90 words (see "Materials and procedure").

https://doi.org/10.1017/S1366728926101278 Published online by Cambridge University Press

Lab session

Participants who passed the prescreening test were invited to participate in the lab session, which comprised a language background questionnaire, followed by an L1 Stroop task and an L2 picture-naming task, followed by an L1-to-L2 translation task. The order of the L2 picture-naming task and the L1 Stroop task was counterbalanced across participants to control for potential practice effects for the cue manipulation, which was common to both tasks. All instructions were given in Italian. The experimenter was in the room with the participant during all instructions and practice sessions and ensured that the participant understood each task, providing additional explanations when necessary. The whole session took about 2.5 hours to complete.

Language background questionnaire. To assess participants' language background, we used the LHQ3 (Li et al., 2019), a validated tool to measure, by self-report, several aspects of the bilingual experience such as age of acquisition (AoA), proficiency and patterns of language use. We used the same questionnaire as in Spinelli and Sulpizio (2024, 2025), an Italian translation of the English version presented using the Jotform (https://www.jotform.com/) survey services.

L1 stroop task. The task was a close replication of Bugg and Smallwood's (2016) Experiment 1, exclusively for the condition in which the cue was presented for 2000 ms. We chose that condition in that experiment because it was the only one among the tested conditions that showed a cueing effect for both congruent and incongruent stimuli in a situation in which, for the incongruent stimuli, the cue was only allowed to predict the stimulus' incongruent status (as opposed to both that status and, given the identity of the word, the identity of the color – a form of compound-cue contingency learning: Schmidt & Lemercier, 2019). Further, that pattern has been replicated twice (in Bugg and Smallwood's, 2016, Experiment 4, and in Jiménez et al.'s, 2021, Experiment 7). The main deviations from the original experiment were the following: 1) All materials were adapted to Italian, participants' L1; 2) we included an additional practice session to ensure participants would understand the cues' meaning; 3) we doubled the number of trials in the experimental sessions (from 96 to 192); and 4) response accuracy was not coded online (in fact, the experimenter was not in the room during the experimental sessions).

Apart from these changes, the experiment was as close as possible to Bugg and Smallwood's (2016). Specifically, we used four colors (red, green, blue and yellow) and the corresponding Italian names (ROSSO, VERDE, BLU and GIALLO) to create, for each of the two lists we used (the uninformative list and the informative list, see below), 96 congruent and 96 incongruent stimuli, with the latter formed by combining each color with all three of the incongruent words for that color. The frequency of each color–word combination in each list in our experiment is presented in Table 2. Note that this design is such that the Stroop effect is likely confounded with a contingency-learning effect (Schmidt & Besner, 2008) because each word (presented 48 times in total) is more likely to predict its congruent color (24 presentations) than any of the other (incongruent) colors (8 presentations each). However, that confound is irrelevant for the key contrast – the contrast between lists – because both of those lists were based on the same design. What did change between lists was the nature of the cues (for a representation of the composition of the two lists, see Figure 1A and B): In the informative list (Figure 1B), the cues were the word CORRISPONDENTI ("corresponding") for congruent stimuli and the word DIFFERENTI ("different") for incongruent stimuli (i.e., 100% valid cues); in the uninformative list (Figure 1A), the cue was the string XXXXXXXXXXXX (i.e., 12 Xs, the average length of the two word cues) for both congruent and incongruent stimuli in all trials (i.e., a completely uninformative cue).

Table 2. Frequency of color–word combinations in each of the lists in the L1 Stroop task

Colors	Words			
	ROSSO	VERDE	BLU	GIALLO
Red	24	8	8	8
Green	8	24	8	8
Blue	8	8	24	8
Yellow	8	8	8	24

In both lists, each trial involved the cue presented for 2000 ms, a blank screen presented for 100 ms and the stimulus presented for 2000 ms or until response. Both the cues and the stimuli were presented in Courier New in the center of the screen against a medium-gray background. The stimuli were presented in one of the four colors in pt. 14 regular format, whereas the cues were presented in black, pt. 21 bold format for increased saliency. There was no performance feedback. Each list was divided into two blocks of 96 trials each, with a self-paced pause in the middle.

The order of presentation of the stimuli within a block was randomized. The order of the informative and uninformative lists was counterbalanced across participants. This order, however, was always consistent with the order used for the L2 picture-naming task (e.g., participants who were presented with the uninformative list first in the L1 Stroop task were presented with the uninformative list first in the L2 picture-naming task as well). Regardless of list order, the task always began with a general practice session. In this session, participants were first given the general task instructions – that is, naming the color in Italian, their L1, as quickly and as accurately as possible – and recommendations for responding vocally – that is, speaking clearly without hesitations. Then, they were given definitions and examples of "corresponding" stimuli (defined as those for which there was correspondence between the color and the word, i.e., congruent) and "different" stimuli (defined as those for which there was no correspondence between the color and the word, i.e., incongruent). To ensure participants would acquire this distinction, they were then presented with eight trials with a special procedure. Each trial involved a plus symbol presented for 500 ms, a blank screen presented for 100 ms and a stimulus presented for 4000 ms (twice as long as in the experimental session) or until response; a question reading "CORRISPONDENTI o DIFFERENTI?" ("corresponding or different?") presented until response; and a feedback display presented until participants pressed Enter on the numpad. All text was presented in Courier New, black (except for the stimuli), pt. 14, regular format in the center of the screen against a medium-gray background. Participants were instructed to name the color of the stimulus when it was presented (with speed being de-emphasized for this session) and to respond to the question by pressing either 1 or 2 on the numpad to categorize the stimulus as "corresponding" or "different," respectively. The feedback display told participants whether their categorization response was correct or incorrect and what the correct response, the color and the word for the stimulus in that trial were. There were one congruent stimulus and one incongruent stimulus for each of the four colors and words used. At the end of the session, participants were told that they would need not categorize stimuli anymore but should

https://doi.org/10.1017/S1366728926101278 Published online by Cambridge University Press

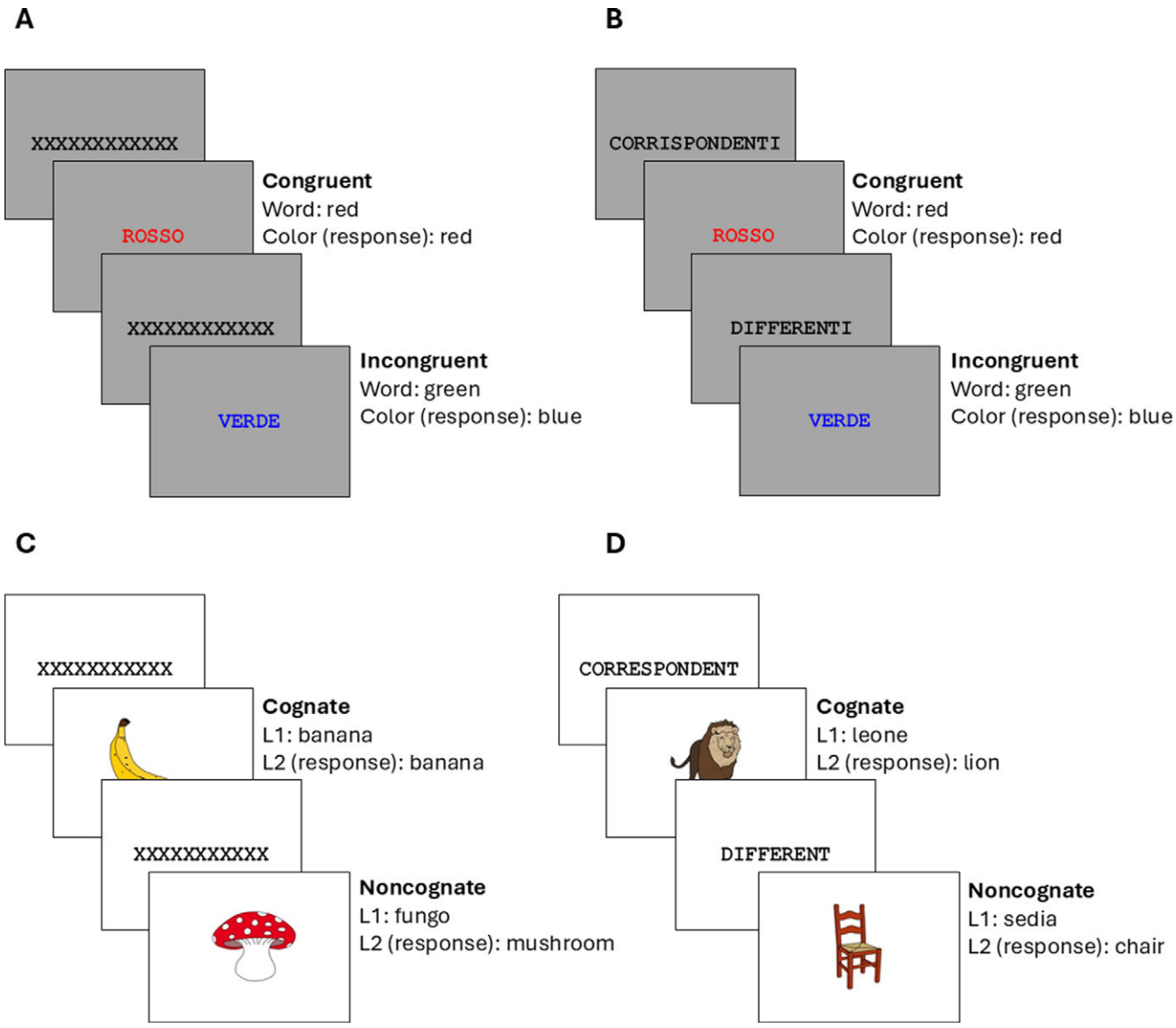


Figure 1. Representation of the lists used in the L1 Stroop task with uninformative cues (A) and with informative cues (B) and in the L2 picture-naming task with uninformative cues (C) and with informative cues (D).

remember the distinction because they would receive cues in either the first or second part of the task, allowing them to anticipate whether the upcoming stimulus was of the "corresponding" or "different" type. Their task, in any case, was naming the color in their L1.

Additional instructions were presented at the beginning of each of the two lists as well. In both lists, participants were reminded of the general task instructions and recommendations for responding vocally (for the list presented immediately following the general practice session, they were also told that they would now have 2000 ms instead of 4000 ms to respond). In the uninformative list, they were also told that they would not be able to anticipate the type of the upcoming stimulus in that list, but the stimulus would still be of the two learned types. In the informative list, they were also told that they would be able to anticipate the type of the upcoming stimulus in that list and that it was very important for them to do their best to use that information to prepare for that stimulus. They were then given the example of a "different" cue, with which they should prepare to ignore the word, which, in that case, would surely lead them astray. Finally, they were told that the cues were always valid. Each of the two lists then included a list-specific practice session in which participants performed eight trials with the list's procedure described above (again, with one congruent and one incongruent stimulus for each of the four colors and words used). DMDX (Forster & Forster, 2003) was used to program the experiment.

L2 picture-naming experiment. We used the same picture stimuli used by Spinelli and Sulpizio (2024, 2025). They were 192 colored drawings, 96 with cognate and 96 with noncognate English and Italian names, selected from the MultiPic dataset (Duñabeitia et al., 2018) based on the results of a pilot study described in full, along with the selection process, in Spinelli and Sulpizio (2024). The most relevant characteristics are presented in Table 3. Note that cognate and noncognate stimuli differed widely on phonological similarity with Italian while being matched on picture visual complexity and word frequency. It was impossible to match cognate and noncognate words on length in syllables as well, because English words in an English–Italian cognate pair (typically words of Latin origin, e.g., *elephant* from Latin *elephantus*) tend to be longer than English words in an English–Italian noncognate pair (typically words of

https://doi.org/10.1017/S1366728926101278 Published online by Cambridge University Press

Table 3. Characteristics of the cognate and noncognate stimuli used in the L2 picture-naming task

	Cognate			Noncognate			*t*-test	
Characteristic	Mean	SD	Range	Mean	SD	Range	*t*	*p*
Visual complexity (picture)	2.46	0.46	1.19–3.45	2.36	0.32	1.42–3.39	1.53	.127
Number of syllables (word)	2.18	0.79	1–3	1.50	0.66	1–2	6.40	< .001
Zipf frequency (word)	4.08	0.42	3.12–5.02	4.13	0.55	0.70–5.09	−0.81	.419
Phonological similarity with Italian (word)	78.41	9.86	60.48–99.64	3.84	3.11	0.88–19.72	70.67	< .001

Note: Visual complexity was extracted from the MultiPic norms (Duñabeitia et al., 2018) and is expressed on a 1–5 scale. Number of syllables was extracted from N-Watch (Davis, 2005). Zipf frequency was extracted from SUBTLEX-UK (van Heuven et al., 2014). Phonological similarity with Italian was extracted from Spinelli and Sulpizio's (2024) pilot study and is expressed on a 0–100 scale. A *t*-test for independent samples was conducted to compare the mean values for each characteristic for cognate and noncognate stimuli.

Germanic origin, e.g., *goat* from Old English *gat*; see, e.g., Bar-Ilan & Berman, 2007). However, the only impact such a mismatch could have is to reduce the cognate effect (for a detailed discussion, see Spinelli & Sulpizio, 2024). Because the aim of the present study is not to produce a completely confound-free cognate effect but to examine whether this effect would interact with the cue manipulation, this particular mismatch does not pose a problem (note also that our counterbalancing scheme, as explained below, was such that the same cognate and noncognate pictures appeared in informative and uninformative conditions across participants).

Both the cognate set and the noncognate set were split into two subsets of 48 stimuli – the same used in Spinelli and Sulpizio (2025) – roughly matched on the most relevant characteristics reported in Table 3. One subset was assigned to the uninformative list and the other to the informative list in a counterbalanced fashion (for a representation of the composition of the two lists, see Figure 1C and D). All pictures were 300 pixels wide and 300 pixels high. The procedure was identical to that used for the L1 Stroop task, with the following exceptions: 1) In the informative list (Figure 1D), the words "CORRESPONDING" and "DIFFERENT" (i.e., in English, the language of the task) were used as cues (and as category labels in the general practice session), whereas in the uninformative list (Figure 1C), the string XXXXXXXXXXXX (i.e., 11 Xs, close to the average length of the two word cues) was used as cue; 2) the naming response timeout was 3000 ms in the experimental session (6000 ms in the general practice session); 3) the background was white; and 4) each list involved a single block of 96 trials.

Participants also received instructions structured and phrased in a fashion similar to those of the L1 Stroop task. For general task instructions and recommendations, they were told to name the picture in English, their L2, as quickly as possible with the name that they thought was the most appropriate, and to speak clearly, without hesitations and without excessive worry about their Italian accent; "corresponding" stimuli were defined as those for which there was correspondence between their English and Italian names and "different" stimuli as those for which there was no correspondence; and in the informative list, we emphasized the importance for participants to do their best to use the cue to prepare for the upcoming stimulus, with the example of a "different" cue with which they should prepare to ignore the Italian name of the picture in case it comes to mind, because, in that case, it would surely lead them astray. They were also given the same number of stimuli as in the L1 Stroop task for practice (eight in the general practice session and eight in each of the two list-specific practice sessions, of which half were cognate and half were noncognate and none were included in the experiment proper). Although in the general

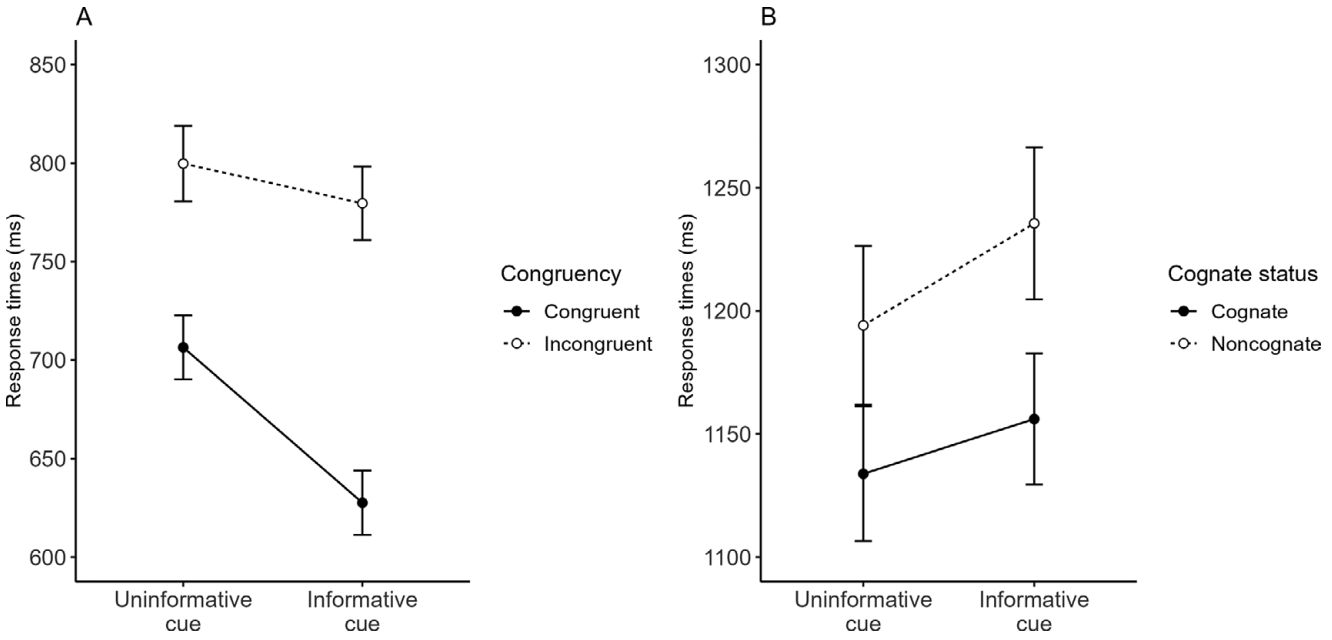


Figure 2. Mean participant-based response times (and corresponding 95% confidence intervals calculated using Cousineau's (2019) method) in the L1 Stroop task (A) and the L2 picture-naming task (B).

https://doi.org/10.1017/S1366728926101278 Published online by Cambridge University Press

practice session, in addition to naming the pictures in their L2, they also categorized them as "corresponding" (i.e., cognate) or "different" (i.e., noncognate), at the end of the session, they were told that they would need not categorize stimuli anymore but should remember the distinction for when informative cues would be provided. Their task, in any case, was naming the picture in their L2. DMDX (Forster & Forster, 2003) was used to program the experiment.

L2-to-L1 translation task. To assess participants' L2 lexical fluency, we used the same L1-to-L2 translation task used by Spinelli and Sulpizio (2024, 2025). The task comprised 30 high-frequency, 30 medium-frequency and 30 low-frequency Italian words, all of which had one (in the case of one of the words, two) acceptable English translation(s) according to Word Reference (https://www.wordreference.com/) and none of which had been involved in the previous tasks or were Italian–English cognates. Participants completed this task with no time limit in an Excel spreadsheet in which the words appeared one above the other in a fixed order of descending frequency.

Data analysis

Here, we report the appropriate confirmatory analyses to examine the impact of a cue manipulation in adaptive control and bilingual language production. Thus, these analyses focus on the group-level results for L1 Stroop and L2 picture naming. Exploratory analyses examining individual-level associations between linguistic variables and performance on L1 Stroop or L2 picture naming, and between performance across the two tasks, are reported in the Supplementary Materials.

For both experiments, the waveforms of vocal responses to the colored (L1 Stroop) and picture stimuli (L2 picture naming) were manually inspected with CheckVocal (Protopapas, 2007) to determine the accuracy of the response and the correct placement of timing marks. In line with our previous studies (Spinelli & Sulpizio, 2024, 2025), we scored participants' responses with some leniency on pronunciation in the L2 picture-naming experiment (e.g., with *mountain* pronounced [ˈmɔntain] instead of [ˈmaʊntɪn] being considered acceptable) and on spelling in the L1-to-L2 translation task (e.g., with *rackoon* instead of *raccoon* being considered acceptable), but a response was considered correct only if it matched the expected response. Prior to the analyses, invalid trials due to technical failures, responses faster than 300 ms and null responses (257 observations for L1 Stroop and 975 observations for L2 picture naming) were discarded. Prior to the latency analyses, incorrect responses (254 observations for L1 Stroop and 1016 observations for L2 picture naming) were also discarded. After discarding invalid and incorrect responses, seven participants contributed fewer than 70% of their original observations in the uninformative list (i.e., the baseline condition) in the L2 picture-naming experiment. Following a criterion determined a priori in line with previous work (Spinelli et al., 2020; Spinelli & Lupker, 2023; Spinelli & Sulpizio, 2024, 2025), those participants (whose original observations were 2688 for L1 Stroop and 1344 for L2 picture naming) were removed from the analyses. As a result, as noted, 48 participants remained in the final sample. Analyses with the full sample, reported in the Supplementary Materials, produced the same pattern of results.

All analyses were conducted in R version 4.2.2 (R Core Team, 2022). To analyze performance in the L1 Stroop and L2 picture-naming tasks, R-default treatment contrasts were changed to sum-to-zero contrasts (i.e., contr.sum) to help interpret lower-order effects in the presence of higher-order interactions. Separate analyses were conducted for L1 Stroop and L2 picture naming. For both experiments, linear mixed-effects models were used to fit trial-level response times (RTs) and generalized linear mixed-effects models were used to fit trial-level accuracy specifying a binomial distribution with a logit link between fixed effects and the dependent variable.

Concerning random effects, the models for both experiments included random intercepts for participants and target stimuli. Analyses with the maximal random structure allowed by the data (Bates et al., 2015a), reported in the Supplementary Materials, produced a similar pattern of results. For L1 Stroop, the fixed effects were Congruency (congruent versus incongruent) and Cue Type (uninformative versus informative); for L2 picture naming, they were Cognate Status (cognate versus noncognate) and Cue Type (uninformative versus informative).[1]

For the analyses reported here, in addition to the regression models described, we also obtained through backward selection the best-fitting model for RTs, as it is in that dependent variable that cueing effects typically emerge (Bugg & Smallwood, 2016; Jiménez et al., 2021). Because the two key effects are the main effect of Cue Type and the interaction between Cue Type and Congruency (for L1 Stroop)/Cognate Status (for L2 picture naming), we also conducted a Bayes factor analysis to quantify the evidence for/against those effects in the RT data. Specifically, for the main effect of Cue Type, we contrasted a model with both main effects, interpreted as the alternative hypothesis H_1, and a model without the Cue Type main effect, interpreted as the null hypothesis H_0. Similarly, for the interaction effect, we contrasted a model with the interaction, interpreted as the alternative hypothesis H_1, and a model without that interaction, interpreted as the null hypothesis H_0. Each of these contrasts yielded BF_{10}, with values above 1 representing evidence for the presence of the effect and values below 1 representing evidence for the absence of the effect (values around 1 would represent no real evidence for either hypothesis).

The lmer function in the *lmerTest* package, version 3.1–3 (Kuznetsova et al., 2017), and the glmer function in the *lme4* package, version 1.1–30 (Bates et al., 2015b), were used to fit the linear and generalized linear mixed-effects regression models, respectively, and obtain probability estimates. The emmeans function in the *emmeans* package, version 1.7.5 (Lenth, 2022), was used to conduct follow-up analyses. The lmBF function in the *BayesFactor* package, version 0.9.12.4.4 (Morey & Rouder, 2022), was used to fit RT models for the Bayes factor analyses.

Results

L1 Stroop task

The mean participant-based RTs are presented in Figure 2A and in Table 4 along with mean error rates. Full results from the RT and accuracy models are reported in Table 5. As Table 5 shows, in the accuracy model, the main effect of Congruency was significant (reflecting, as expected, more accurate responses to congruent than incongruent stimuli), whereas the main effect of Cue Type was marginal (reflecting a numerical tendency for more accurate responses with informative cues overall). The interaction between

[1] Analyses with Cue Type Order (i.e., the order in which the lists with the two different cue types were administered within an experiment) and Experiment Order (i.e., the order in which the two experiments were administered) as additional fixed effects, reported in the Supplementary Materials, revealed that neither order predictor modulated the key interactions between Congruency and Cue Type (for L1 Stroop) and Cognate Status and Cue Type (for L2 picture naming).

https://doi.org/10.1017/S1366728926101278 Published online by Cambridge University Press

Table 4. Mean participant-based response times and percentage error rates (and corresponding 95% confidence intervals calculated using Cousineau's (2019) method) in the L1 Stroop task

	Response times		Error rates	
Congruency	Uninformative cue	Informative cue	Uninformative cue	Informative cue
Congruent	706 [690, 723]	628 [611, 644]	0.39 [0.16, 0.63]	0.15 [0.02, 0.29]
Incongruent	800 [781, 819]	780 [761, 798]	1.98 [1.15, 2.81]	2.31 [0.94, 3.68]
Congruency effect	94	152	1.59	2.16

Table 5. Variances and standard deviations for the random effects and coefficients, and standard errors, statistics and probability values for the fixed effects used in the models of response times and accuracy in the L1 Stroop task

	Response times				Accuracy			
Random effect	*Variance*		*SD*		*Variance*		*SD*	
Participant (intercept)	8687.83		93.21		0.727		0.853	
Color (intercept)	108.68		10.43		0.045		0.211	
Fixed effect	β	*SE*	*t*	*p*	β	*SE*	*z*	*p*
Intercept	728.20	14.47	30.31	< .001	5.308	0.209	25.45	< .001
Congruency	−61.25	1.14	−53.65	< .001	1.100	0.114	9.64	< .001
Cue Type	24.95	1.14	21.86	< .001	−0.195	0.114	−1.71	.088
Congruency × Cue Type	14.64	1.14	12.83	< .001	−0.276	0.114	−2.42	.015

Note: The accuracy coefficients are in the logit scale, not in the response scale.

the two was also significant. Follow-up analyses revealed that this effect reflected more accurate responses with informative than noninformative cues for congruent stimuli (a small, but significant 0.24% difference), $\beta = -0.942$, $SE = 0.433$, $z = -2.17$, $p = .030$, whereas no cueing effect emerged for incongruent stimuli (a numerically larger, but nonsignificant 0.33% difference in the other direction), $\beta = 0.163$, $SE = 0.143$, $z = 1.14$, $p = .255$. More importantly, in the RTs, all of the effects were significant and in the expected direction: the main effect of Congruency (reflecting faster responses to congruent than incongruent stimuli), the main effect of Cue Type (reflecting faster responses with informative than noninformative cues) and the interaction between the two. The latter reflected that, while both congruent and incongruent stimuli showed a significant cueing effect in the expected direction ($\beta = 79.19$, $SE = 3.21$, $z = 24.64$, $p < .001$, and $\beta = 20.63$, $SE = 3.24$, $z = 6.36$, $p < .001$, respectively), that effect was larger for congruent stimuli (78 ms) than for incongruent ones (20 ms). Overall, this pattern is the typical one for cueing manipulations in conflict tasks. The backward selection procedure confirmed that the interactive model was the best-fitting model. The Bayes factor analyses favored the inclusion of both the main effect of Cue Type, $BF_{10} = 5.94 \times 10^{100} \pm 3.35\%$, and the interaction, $BF_{10} = 7.53 \times 10^{33} \pm 4.35\%$.

L2 picture-naming task

The mean participant-based RTs are presented in Figure 2B and in Table 6 along with mean error rates. Full results from the RT and accuracy models are reported in Table 7. As Table 7 shows, the only effect reaching significance in the accuracy data was the main effect of Cognate Status (reflecting, as expected, more accurate responses to cognate than noncognate stimuli). In the RTs, there was also a significant main effect of Cognate Status reflecting, as expected, faster responses to cognate than noncognate stimuli. However, there was also a significant main effect of Cue Type reflecting a REVERSE cueing effect, with slower responses with informative than noninformative cues overall. Further, the interaction was not significant. Indeed, using backward selection, the best-fitting model was the additive one (see the Supplementary Materials for full results). Consistent with this, the Bayes factor analyses strongly favored the inclusion of the (reverse) main effect of Cue Type, $BF_{10} = 104.21 \pm 3.04\%$, but not the interaction, $BF_{10} = 0.09 \pm 5.19\%$.

Table 6. Mean participant-based response times and percentage error rates (and corresponding 95% confidence intervals calculated using Cousineau's (2019) method) in the L2 picture-naming task

	Response times		Error rates	
Cognate status	Uninformative cue	Informative cue	Uninformative cue	Informative cue
Cognate	1134 [1106, 1161]	1156 [1129, 1183]	7.84 [6.35, 9.34]	8.21 [6.52, 9.91]
Noncognate	1194 [1162, 1226]	1236 [1205, 1266]	11.21 [9.37, 13.06]	10.96 [8.75, 13.18]
Cognate effect	60	80	3.37	2.75

https://doi.org/10.1017/S1366728926101278 Published online by Cambridge University Press

Table 7. Variances and standard deviations for the random effects and coefficients, and standard errors, statistics and probability values for the fixed effects used in the models of response times and accuracy in the L2 picture-naming task

	Response times				Accuracy			
Random effect	Variance		SD		Variance		SD	
Participant (intercept)	16454.61		128.28		0.213		0.462	
Picture (intercept)	43925.18		209.58		1.829		1.352	
Fixed effect	*β*	*SE*	*t*	*p*	*β*	*SE*	*z*	*p*
Intercept	1203.75	24.24	49.66	< .001	2.941	0.134	21.93	< .001
Cognate Status	−30.98	15.64	−1.98	.049	0.249	0.110	2.25	.024
Cue Type	−16.26	3.94	−4.12	< .001	< 0.001	0.039	0.01	.996
Cognate Status × Cue Type	5.56	3.94	1.41	.159	0.027	0.039	0.68	.496

Note: The accuracy coefficients are in the logit scale, not in the response scale.

Discussion

When speaking in their L2, can bilinguals optimize control over their L1 when they are explicitly told whether facilitation or interference should be expected? For a positive answer to this question, our L2 picture-naming task should have shown a pattern parallel to that produced by L1 Stroop tasks: better performance for both stimuli for which the distractor produces facilitation (congruent in L1 Stroop, cognate in L2 picture naming) and for stimuli for which the distractor produces interference (incongruent in L1 Stroop, noncognate in L2 picture naming) with cues that reveal the nature of the upcoming stimulus (informative cues) relative to cues that reveal no information at all (uninformative cues; Bugg & Smallwood, 2016; Jiménez et al., 2021). We replicated that type of pattern in our L1 Stroop task, with faster responses to both incongruent and (especially) congruent stimuli with informative versus uninformative cues. With the design we used, this pattern can be interpreted as reflecting that selection is relaxed when a congruent stimulus is cued (in which case, participants can base their response exclusively on the distractor rather than the target), whereas selection is tightened when an incongruent stimulus is cued (Bugg & Smallwood, 2016). More generally, the fact that we obtained regular cueing effects in L1 Stroop confirms that the present cue manipulation, an explicit manipulation, CAN be effective at eliciting performance optimization through adaptive control.

The same manipulation, however, was not effective at eliciting that sort of control in our L2 picture-naming task. Indeed, informative (versus uninformative) cues did not facilitate responses to cognate or noncognate pictures in that task – they actually slowed down those responses, a REVERSE cueing effect. In other words, when applied to L2 picture naming, our manipulation seemed to induce a form of MALADAPTIVE, rather than adaptive, control. In order to understand whether similar or different processes underlie these forms of control in the two tasks, it is useful to consider the results of one of the exploratory analyses reported in the Supplementary Materials, an analysis in which the participant's size of the cueing effect in L2 picture naming was included as an additional predictor of L1 Stroop RTs (similar to a correlational analysis). As represented in Figure 3, larger cueing effects in L2 picture naming were associated with a larger difference between informative and uninformative lists in L1 Stroop, $\beta = 2.36$, $SE = 1.16$, $t = 2.04$, $p = .042$, a positive association (with cueing effects in L2 picture naming having no overall impact on L1 Stroop RTs, suggesting that the association between cueing effects does not reflect merely processing speed). Thus, although the cueing effect in L2 picture naming was overall negative (versus positive in L1 Stroop), the processes involved in using the cues in that task were potentially similar, to some extent, to those involved in using the cues in L1 Stroop (although their impact clearly changed based on the characteristics of the task). As additional support to the idea that reverse cueing effects in L2 picture naming were mainly due to domain-general processes of this type, those effects were not modulated by participants' L2 lexical fluency. This suggests that the impairment caused by informative cues in L2 picture naming likely does not involve processes that are sensitive to how well one handles a language (but do note that L2 lexical fluency did modulate the regular cueing effect in L1 Stroop and had a strong impact on both overall latencies and the cognate effect in L2 picture naming, in line with previous research: Spinelli & Sulpizio, 2024, 2025; see also Costa et al., 2000; more importantly, the results of these exploratory analyses must be taken with caution).

The fact that the cueing effect was reverse in L2 picture naming, in any case, requires an explanation. One possibility is that, while the L2 picture-naming process typically involves parallel activation of the L2 target and the L1 distractor, informative cues, as discussed in the Introduction, might have the side effect of increasing the

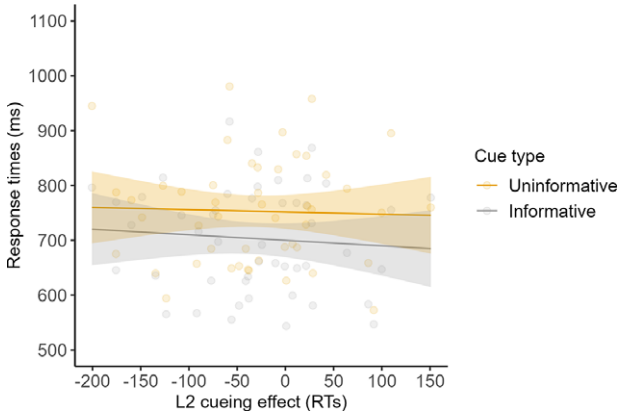


Figure 3. The impact of the L2 picture-naming cueing effect on RTs for uninformative and informative conditions in the L1 Stroop task.
Note: The gold and gray circles represent mean participant RTs, and the lines (with 95% confidence bands) represent the model-estimated trends for the uninformative and informative conditions (for further details, see Supplementary Materials). See the online version of this article for colors.

https://doi.org/10.1017/S1366728926101278 Published online by Cambridge University Press

distractor's saliency. While this increased saliency can produce facilitation at the item selection level for cognate stimuli (because the L1 name is similar to the target L2 name), it will produce interference for both cognate and noncognate stimuli at the language selection level (because, with both languages highly active, it will be harder to determine which one must be selected). As a result, even for cognate stimuli, any advantage given by retrieving the L2 name based on the similar-sounding L1 name will be offset by the between-language competition that must be resolved for those stimuli as well as for noncognate ones.

As also discussed, informative cues can also increase distractor saliency in L1 Stroop. If this increased distractor saliency translates into additional task conflict, it might counteract and potentially offset any facilitation due to the use of the information conveyed by the cues for both congruent and incongruent stimuli (stimuli that both involve that type of conflict). At the same time, the fact that a positive net effect of informative cues was observed in this task (versus a negative effect in L2 picture naming) suggests that this additional task conflict is somewhat limited (and even eliminated for congruent stimuli if participants switch to reading with those stimuli, as we discuss below) and/or that other differences exist between L1 Stroop and L2 picture naming, despite their structural similarity, that can explain the different patterns they produced.

One such difference is the fact, noted in the Introduction, that the distractor is internally generated in L2 picture naming versus explicitly presented in L1 Stroop. As we discussed, this fact might make the saliency of the distractor lower in L2 picture naming than in L1 Stroop at baseline (i.e., with uninformative cues). As a result, informative cues would increase the distractor's saliency to a greater extent in L2 picture naming than in L1 Stroop, causing so much additional between-language competition that is hard to overcome in L2 picture naming but not so much additional task conflict in L1 Stroop that it cannot be overcome by using the information conveyed by the cues.

It could be argued, at this point, that, considering the large benefit associated with CONGRUENT cues in L1 Stroop, using the information conveyed by COGNATE cues in L2 picture naming should produce a benefit large enough to counteract most, if not all, of the additional between-language competition caused by informative cues. Yet, the reverse cueing effect for cognate stimuli was statistically indistinguishable from that for noncognate stimuli (although it was numerically smaller). A potential reason for this fact is our use of very similar-sounding (e.g., elephant/elefante) but typically not IDENTICAL cognates (e.g., avocado/avocado). As a result, with a cognate cue, participants could use the distractor (i.e., the L1 name of the picture) to get most of the correct response (i.e., the L2 name), but, typically, not all of it (for the potential relevance of identical versus non-identical cognates, see Arana et al., 2022). In other words, with a cognate cue, participants could, in most cases, RELY on the L1 name but could not directly USE that name to respond, that is, they could not switch from L2 naming to L1 naming for that trial. Such was not the case for congruent stimuli in L1 Stroop, for which the distractor (i.e., the word) always WAS the correct response (e.g., the word ROSSO in the color red required *rosso* as a response). Therefore, with a congruent cue, participants were free to switch from color naming to word reading for that trial. Indeed, as noted, such a fact is likely responsible for the large cueing effect typically observed with congruent stimuli (Bugg & Smallwood, 2016). More to the point, for congruent stimuli in L1 Stroop, participants did not need to engage in any of the monitoring processes that were likely needed in L2 picture naming (e.g., selecting the exact L1 name rather than the similar-sounding L2 name), processes that likely involve some between-language competition as for noncognate stimuli (Broersma et al., 2016) and that likely contributed to slowing down responses in that situation.

As also discussed, adaptive control might be harder (albeit not impossible: Kiyonaga & Egner, 2014; Van Gaal et al., 2010) to implement with internally generated distractors because the interference/facilitation produced by those distractors may not be immediately felt in that situation. By explicitly directing attention to distractors, our manipulation certainly reduced this type of gap between the two tasks. However, it likely did not eliminate it completely. Specifically, even with attention directed at distractors, distinguishing their representations from target representations might be easier when distractors are physically present (in L1 Stroop) versus only present in working memory (in L2 picture naming). As a result, controlling those representations would be easier in L1 Stroop, leaving little potential for the distractor to be used maladaptively as it seems to occur in L2 picture naming. For example, in L1 Stroop, participants might be able to use the cue in order to adjust word versus color processing, whereas no similar processing adjustment would help in L2 picture naming.

Another potentially important methodological difference between the two tasks lies in the size of the stimulus set. With only four potential responses (i.e., the four color names) being used in L1 Stroop (versus many unique pictures in L2 picture naming, each with its own L2 name), participants might not need to rely on the distractor to guide retrieval of the target's name from long-term memory, simply because, in such a situation, potential responses are readily available and may not need to be retrieved from long-term memory at every trial. As a result, again, there would be less potential for maladaptive use of the distractor in L1 Stroop.

Closely examining these potential reasons for the dissociation we observed is a key item in the future research agenda. For example, the present study could be replicated using a language pair that involves a large number of identical cognates (e.g., Spanish–Catalan), such that, different from the present study, participants would be able to switch between languages when presented with a cognate cue.[2] Set size could also be manipulated, such that a smaller set of pictures is presented repeatedly, more similar to the Stroop situation.[3]

In conducting such follow-ups, however, it will be important to consider whether and how the modification applied might affect the paradigm's ability to capture the normal bilingual experience. For example, another way to make our L2 picture naming more similar to a Stroop task would be to present the picture's L1 name

[2]Our study did include a small number of identical or near-identical cognates (e.g., banana, pizza). To determine whether the degree of L1–L2 similarity of cognate pictures might have modulated the cueing effect they produced, we conducted an additional exploratory analysis on the RT and error data for those pictures with L1–L2 similarity from the norms we previously collected (Spinelli & Sulpizio, 2024) as a predictor in addition to Cue Type. Inconsistent with the tested hypothesis, there was neither a main effect of L1–L2 similarity nor, most crucially, an interaction with Cue Type in either analysis (all $ps \geq .159$).

[3]Another potential follow-up would involve conducting the picture-naming task in participants' L1. In theory, the requirement to monitor the L2 with informative cues might create a situation of reverse dominance (i.e., one in which using the nominally dominant L1 is actually slower than using the L2; Declerck & Koch, 2023). In such a situation, the reverse cueing effect might even be larger than that observed in the present L2 picture-naming task. In practice, however, participants might stop using cues when doing so turns out to be so damaging (and considering the picture's L1 name is typically available earlier than its L2 name), resulting in a small, if any, cueing effect. The impact of a cue manipulation in L1 picture naming, in sum, is largely an empirical question.

https://doi.org/10.1017/S1366728926101278 Published online by Cambridge University Press

superimposed on it, that is, making it a picture–word interference paradigm (Lupker, 1979). Doing so would certainly address the concern, discussed above, that adaptive control might be difficult to apply to distractors that are not explicitly presented. On the other hand, in the normal bilingual experience, naming a concept in one language typically does NOT involve some explicit presentation of the translation equivalent of that concept in the nontarget language. Instead, the distraction produced by the translation equivalent is typically internally generated, a situation that seems to be better captured by regular picture naming. Note that this point is important also because, while, as discussed, there might be several reasons why our L2 picture-naming task produced no evidence of adaptive control, that task is still quite representative of how L2 production normally unfolds. Therefore, while there might be situations in which L2 production does involve adaptive control, the present results, along with the previous ones in this line of research (Spinelli & Sulpizio, 2024, 2025), suggest that those situations might not be particularly common (see also Blanco-Elorrieta & Pylkkänen, 2018).

Indeed, to avoid these problems with ecological validity for L2 production, one might make the L1 Stroop task more similar to an L2 picture-naming task (e.g., by increasing the stimulus set size: Spinelli et al., 2019) rather than the other way around as discussed thus far. This point is important because it emphasizes the fact that a dissociation between bilingual and domain-general control might depend on the tool used to measure EITHER of those constructs, not just the bilingual-control one.

Overall, following our previous findings of a dissociation (Spinelli & Sulpizio, 2024) and an association, albeit complicated by an unexpected pattern (Spinelli & Sulpizio, 2025), between Stroop and cognate effects using implicit manipulations of adaptive control, the present study, involving an explicit manipulation of adaptive control, produced another dissociation between the two effects. Although dissociations of this kind, as noted, might depend on the tools used, they are nonetheless important for understanding how bilingual and domain-general control compare to one another.

A broader implication concerns potential long-term consequences of bilingualism on broader cognition (Bialystok & Craik, 2022). For there to be potential for those consequences to arise, there has to be some commonality between domain-general processes and the processes involved in bilingual language use (De Bruin et al., 2021). The type of qualitative differences between the two types of processes evidenced by our research, however, suggests that these commonalities are likely limited, which in turn might help explain the difficulty in consistently observing bilingual advantages in domain-general abilities (Paap, 2022).

More generally, qualitative differences between domain-general and bilingualism-specific processes help reach a more complete understanding of how bilingualism adapts basic cognitive tools to address challenges that are more typical of the bilingual situation. The present study is a further step in that direction.

Supplementary material. The supplementary material for this article can be found at http://doi.org/10.1017/S1366728926101278.

Data availability statement. The materials, datasets and scripts used to perform the analyses are available on the Open Science Framework at https://osf.io/6arq2/.

Acknowledgements. This research was supported by the University of Milano-Bicocca's "assegno di ricerca" (research grant) 21A2/24 to Giacomo Spinelli. Giacomo Spinelli is now affiliated with the University of Zurich.

Competing interests. The authors declare none.

Publishing ethics. The authors assert that all procedures contributing to this work comply with the ethical standards of the relevant national and institutional guides on the care and use of laboratory animals.

Disclosure of use of AI tools. No AI tools were used.

References

Arana, S. L., **Oliveira, H. M.**, **Fernandes, A. I.**, **Soares, A. P.**, & **Comesaña, M.** (2022). The cognate facilitation effect depends on the presence of identical cognates. *Bilingualism: Language and Cognition*, **25**(4), 660–678. https://doi.org/10.1017/S1366728922000062.

Bar-Ilan, L., & **Berman, R. A.** (2007). Developing register differentiation: The Latinate-Germanic divide in English. *Linguistics*, **45**(1), 1–35. https://doi.org/10.1515/ling.2007.001.

Bates, D., **Kliegl, R.**, **Vasishth, S.**, & **Baayen, H.** (2015a). Parsimonious mixed models. *ArXiv*, 1506.04967. https://doi.org/10.48550/arXiv.1506.04967.

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

Bialystok, E., & **Craik, F. I. M.** (2022). How does bilingualism modify cognitive function? Attention to the mechanism. *Psychonomic Bulletin & Review*, **29**(4), 1246–1269. https://doi.org/10.3758/s13423-022-02057-5.

Blais, C., **Harris, M. B.**, **Guerrero, J. V.**, & **Bunge, S. A.** (2012). Rethinking the role of automaticity in cognitive control. *Quarterly Journal of Experimental Psychology*, **65**(2), 268–276. https://doi.org/10.1080/17470211003775234.

Blanco-Elorrieta, E., & **Pylkkänen, L.** (2018). Ecological validity in bilingualism research and the bilingual advantage. *Trends in Cognitive Sciences*, **22**(12), 1117–1126. https://doi.org/10.1016/j.tics.2018.10.001.

Branzi, F. M., **Calabria, M.**, **Boscarino, M. L.**, & **Costa, A.** (2016). On the overlap between bilingual language control and domain-general executive control. *Acta Psychologica*, **166**, 21–30. https://doi.org/10.1016/j.actpsy.2016.03.001.

Braver, T. S. (2012). The variable nature of cognitive control: A dual mechanisms framework. *Trends in Cognitive Sciences*, **16**(2), 106–113. https://doi.org/10.1016/j.tics.2011.12.010.

Braem, S., **Bugg, J. M.**, **Schmidt, J. R.**, **Crump, M. J.**, **Weissman, D. H.**, **Notebaert, W.**, & **Egner, T.** (2019). Measuring adaptive control in conflict tasks. *Trends in Cognitive Sciences*, **23**(9), 769–783. https://doi.org/10.1016/j.tics.2019.07.002.

Broersma, M., **Carter, D.**, & **Acheson, D. J.** (2016). Cognate costs in bilingual speech production: Evidence from language switching. *Frontiers in Psychology*, **7**, 1461. https://doi.org/10.3389/fpsyg.2016.01461.

Bugg, J. M., & **Gonthier, C.** (2020). List-level control in the flanker task. *Quarterly Journal of Experimental Psychology*, **73**(9), 1444–1459. https://doi.org/10.1177/1747021820912477.

Bugg, J. M., & **Smallwood, A.** (2016). The next trial will be conflicting! Effects of explicit congruency pre-cues on cognitive control. *Psychological Research*, **80**, 16–33. https://doi.org/10.1007/s00426-014-0638-5.

Calabria, M., **Hernández, M.**, **Branzi, F. M.**, & **Costa, A.** (2012). Qualitative differences between bilingual language control and executive control: Evidence from task-switching. *Frontiers in Psychology*, **2**, 399. https://doi.org/10.3389/fpsyg.2011.00399.

Costa, A., **Caramazza, A.**, & **Sebastian-Galles, N.** (2000). The cognate facilitation effect: Implications for models of lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **26**(5), 1283–1296. https://doi.org/10.1037/0278-7393.26.5.1283.

Cousineau, D. (2019). Correlation-adjusted standard errors and confidence intervals for within-subject designs: A simple multiplicative approach. *The Quantitative Methods for Psychology*, **15**(3), 226–241. https://doi.org/10.20982/tqmp.15.3.p226.

Davis, C. J. (2005). N-Watch: A program for deriving neighborhood size and other psycholinguistic statistics. *Behavior Research Methods*, **37**(1), 65–70. https://doi.org/10.3758/bf03206399.

De Bruin, A., **Dick, A. S.**, & **Carreiras, M.** (2021). Clear theories are needed to interpret differences: Perspectives on the bilingual advantage debate. *Neurobiology of Language*, **2**(4), 433–451. https://doi.org/10.1162/nol_a_00038.

Declerck, M., **Grainger, J.**, **Koch, I.**, & **Philipp, A. M.** (2017). Is language control just a form of executive control? Evidence for overlapping processes in language switching and task switching. *Journal of Memory and Language*, **95**, 138–145. https://doi.org/10.1016/j.jml.2017.03.005.

Desender, K., **Van Opstal, F.**, & **Van den Bussche, E.** (2014). Feeling the conflict: The crucial role of conflict experience in adaptation. *Psychological Science*, **25**(3), 675–683. https://doi.org/10.1177/0956797613511468.

Declerck, M., & **Koch, I.** (2023). The concept of inhibition in bilingual control. *Psychological Review*, **130**(4), 953–976. https://doi.org/10.1037/rev0000367.

Declerck, M., **Meade, G.**, **Midgley, K. J.**, **Holcomb, P. J.**, **Roelofs, A.**, & **Emmorey, K.** (2021). On the connection between language control and executive control—An ERP study. *Neurobiology of Language*, **2**(4), 628–646. https://doi.org/10.1162/nol_a_00032.

Duñabeitia, J. A., **Crepaldi, D.**, **Meyer, A. S.**, **New, B.**, **Pliatsikas, C.**, **Smolka, E.**, & **Brysbaert, M.** (2018). MultiPic: A standardized set of 750 drawings with norms for six European languages. *Quarterly Journal of Experimental Psychology*, **71**(4), 808–816. https://doi.org/10.1080/17470218.2017.1310261.

Faul, F., **Erdfelder, E.**, **Buchner, A.**, & **Lang, A.-G.** (2009). Statistical power analyses using G*power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, **41**(4), 1149–1160. https://doi.org/10.3758/brm.41.4.1149.

Festman, J., & **Schwieter, J. W.** (2015). Behavioral measures of language control. In J. W. Schwieter (Ed.), *The Cambridge handbook of bilingual processing* (pp. 527–547). Cambridge University Press. https://doi.org/10.1017/cbo9781107447257.023.

Forster, K. I., & **Forster, J. C.** (2003). DMDX: A windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, & Computers*, **35**(1), 116–124. https://doi.org/10.3758/bf03195503.

Gade, M., **Declerck, M.**, **Philipp, A. M.**, **Rey-Mermet, A.**, & **Koch, I.** (2021). Assessing the evidence for asymmetrical switch costs and reversed language dominance effects–a meta-analysis. *Journal of Cognition*, **4**(1), 55. https://doi.org/10.5334/joc.186.

Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition*, **1**(2), 67–81. https://doi.org/10.1017/s1366728998000133.

Green, D. W., & **Abutalebi, J.** (2013). Language control in bilinguals: The adaptive control hypothesis. *Journal of Cognitive Psychology*, **25**(5), 515–530. https://doi.org/10.1080/20445911.2013.796377.

Jiménez, L., **Abrahamse, E.**, **Méndez, C.**, & **Braem, S.** (2020). Does incidental sequence learning allow us to better manage upcoming conflicting events? *Psychological Research*, **84**(8), 2079–2089. https://doi.org/10.1007/s00426-019-01201-6.

Jiménez, L., **Méndez, C.**, **Abrahamse, E.**, & **Braem, S.** (2021). It is harder than you think: On the boundary conditions of exploiting congruency cues. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **47**(10), 1686–1704. https://doi.org/10.1037/xlm0000844.

Kalanthroff, E., **Avnit, A.**, **Henik, A.**, **Davelaar, E. J.**, & **Usher, M.** (2015). Stroop proactive control and task conflict are modulated by concurrent working memory load. *Psychonomic Bulletin & Review*, **22**, 869–875. https://doi.org/10.3758/s13423-014-0735-x.

Kiyonaga, A., & **Egner, T.** (2014). The working memory Stroop effect: When internal representations clash with external stimuli. *Psychological Science*, **25**(8), 1619–1629. https://doi.org/10.1177/0956797614536739.

Koch, I. (2008). Instruction effects in task switching. *Psychonomic Bulletin & Review*, **15**(2), 448–452. https://doi.org/10.3758/PBR.15.2.448.

Koch, I., **Declerck, M.**, **Petersen, G.**, **Rister, D.**, **Scharke, W.**, & **Philipp, A. M.** (2024). Reassessing the role of language dominance in *n*—2 language repetition costs as a marker of inhibition in multilingual language switching. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **50**(9), 1516–1528. https://doi.org/10.1037/xlm0001333.

Kuznetsova, A., **Brockhoff, P. B.**, & **Christensen, R. H. B.** (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, **82**(13), 1–26. https://doi.org/10.18637/jss.v082.i13.

Lenth, R. (2022). Emmeans: Estimated Marginal Means, aka Least-Squares Means. R package version 1.7.5. https://CRAN.R-project.org/package=emmeans

Li, P., **Zhang, F.**, **Yu, A.**, & **Zhao, X.** (2019). Language history questionnaire (LHQ3): An enhanced tool for assessing multilingual experience. *Bilingualism: Language and Cognition*, **23**(5), 938–944. https://doi.org/10.1017/s1366728918001153.

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

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin*, **109**(2), 163–203. https://doi.org/10.1037/0033-2909.109.2.163.

Morey, R., & **Rouder, J.** (2022). *BayesFactor: Computation of bayes factors for common designs*. R package version 0.9.12–4.4. https://CRAN.R-project.org/package=BayesFactor

Paap, K. (2022). *The bilingual advantage in executive functioning hypothesis: How the debate provides insight into psychology's replication crisis*. Routledge. https://doi.org/10.4324/9781003308027.

Parris, B. A., **Hasshim, N.**, **Ferrand, L.**, & **Augustinova, M.** (2023). Do task sets compete in the Stroop task and other selective attention paradigms? *Journal of Cognition*, **6**(1), 23. https://doi.org/10.5334/joc.272.

Parris, B. A., **Hasshim, N.**, **Wadsley, M.**, **Augustinova, M.**, & **Ferrand, L.** (2022). The loci of Stroop effects: A critical review of methods and evidence for levels of processing contributing to color-word Stroop effects and the implications for the loci of attentional selection. *Psychological Research*, **86**(4), 1029–1053. https://doi.org/10.1007/s00426-021-01554-x.

Protopapas, A. (2007). Check vocal: A program to facilitate checking the accuracy and response time of vocal responses from DMDX. *Behavior Research Methods*, **39**(4), 859–862. https://doi.org/10.3758/bf03192979.

R Core Team. (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Schmidt, J. R., & **Besner, D.** (2008). The Stroop effect: Why proportion congruent has nothing to do with congruency and everything to do with contingency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **34**(3), 514–523. https://doi.org/10.1037/0278-7393.34.3.514.

Schmidt, J. R., & **Lemercier, C.** (2019). Context-specific proportion congruent effects: Compound-cue contingency learning in disguise. *Quarterly Journal of Experimental Psychology*, **72**(5), 1119–1130. https://doi.org/10.1177/1747021818787155.

Spinelli, G., **Krishna, K.**, **Perry, J. R.**, & **Lupker, S. J.** (2020). Working memory load dissociates contingency learning and item-specific proportion-congruent effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **46**(11), 2007–2033. https://doi.org/10.1037/xlm0000934.

Spinelli, G., & **Lupker, S. J.** (2023). Target-distractor correlation does not imply causation of the Stroop effect. *Quarterly Journal of Experimental Psychology*. Advance online publication. https://doi.org/10.1177/17470218231182854.

Spinelli, G., **Perry, J. R.**, & **Lupker, S. J.** (2019). Adaptation to conflict frequency without contingency and temporal learning: Evidence from the picture–word interference task. *Journal of Experimental Psychology: Human Perception and Performance*, **45**(8), 995–1014. https://doi.org/10.1037/xhp0000656.

Spinelli, G., & **Sulpizio, S.** (2024). Is adaptation involved in bilingual language production? A fresh look at the assumptions motivating potential bilingual-monolingual differences in adaptive control. *Psychonomic Bulletin & Review*, **31**, 2681–2691. https://doi.org/10.3758/s13423-024-02503-6.

Spinelli, G., & **Sulpizio, S.** (2025). No evidence that selection is resource-demanding in conflict and bilingual language production tasks: Implications for theories of adaptive control and language-control associations. *Psychonomic Bulletin & Review*. Advance online publication. https://doi.org/10.3758/s13423-025-02672-y.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, **18**(6), 643–662. https://doi.org/10.1037/h0054651.

Van Assche, E., **Duyck, W.**, & **Gollan, T. H.** (2013). Whole-language and item-specific control in bilingual language production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **39**(6), 1781–1792. https://doi.org/10.1037/a0032859.

Van Gaal, S., **Lamme, V. A.**, & **Ridderinkhof, K. R.** (2010). Unconsciously triggered conflict adaptation. *PLoS One*, **5**(7), e11508. https://doi.org/10.1371/journal.pone.0011508.

https://doi.org/10.1017/S1366728926101278 Published online by Cambridge University Press

van Heuven, W. J. B., **Mandera, P.**, **Keuleers, E.**, & **Brysbaert, M.** (2014). Subtlex-UK: A new and improved word frequency database for British English. *Quarterly Journal of Experimental Psychology*, **67**(6), 1176–1190. https://doi.org/10.1080/17470218.2013.850521

Wühr, P., **Duthoo, W.**, & **Notebaert, W.** (2015). Generalizing attentional control across dimensions and tasks: Evidence from transfer of proportion-congruent effects. *Quarterly Journal of Experimental Psychology*, **68**(4), 779–801. https://doi.org/10.1080/17470218.2014.966729.

https://doi.org/10.1017/S1366728926101278 Published online by Cambridge University Press