Words can interfere with perception at their associated locations: The role of orthography in spatial interference

Elisa Scerrati[1,2] & Zachary Estes[3]

In press (Aug 2025), *Language and Cognition*

Abstract: The *spatial interference effect*, whereby words with implicit spatial associations (e.g., "bird") hinder identification of unrelated visual targets (e.g., a square) at the associated locations (i.e., at the top of a display), has been demonstrated many times in English, though it has failed to replicate several times in Italian. The current study tested whether the replication failures in Italian may be due to insufficient semantic processing of the words. Indeed, while languages with highly inconsistent pronunciations such as English are more likely to involve semantic processing during word reading, languages with highly consistent pronunciations such as Italian tend to evoke weaker semantic processing during reading. In two experiments, semantic processing in Italian was induced by including a high proportion of irregularly-stressed words. Spatial interference occurred in both experiments. It is concluded that relatively deep semantic processing is necessary for spatial interference to occur.

Keywords: Spatial interference; semantic processing; irregular stress; orthographic transparency; identification task.

Open Science: Experiment 2 was preregistered (available at http://aspredicted.org/blind.php?x=4bb6zh) and all data and code are available at: https://osf.io/fbm7d/?view_only=46603bd6ea574f6eb987b96fe4cce18a.

[1] Department of Psychology, Catholic University of the Sacred Heart, Milan, Italy
[2] Department of Psychology and Health Science, Pegaso Telematic University, Naples, Italy
[3] Bayes Business School, City University of London

In his classic studies on spatial cueing, Posner (1980) demonstrated that symbolic cues can orient spatial attention, thereby affecting visual perception. For instance, an upward-pointing arrow presented in the center of a display facilitates detection of a square at the top of the display and hinders detection at the bottom. Like those symbolic cues (e.g., arrows), some words also have explicit spatial meanings, and those words also affect visual perception at their associated locations (i.e., *linguistic cueing*). For instance, the explicitly spatial word "up" facilitates detection of a square at the top of a display, and impairs detection at the bottom (e.g., Gibson & Kingstone, 2006; Gibson & Sztybel, 2014; Hommel, Pratt, Colzato, & Godijn, 2001; Logan, 1995; Ostarek & Vigliocco, 2017; Pauszek & Gibson, 2018; Shaki & Fischer, 2023a, 2023b). In fact, words with merely implicit spatial associations, such as religious words (e.g., "god" = high, "satan" = low; Chasteen, Burdzy, & Pratt, 2010) and temporal words (e.g., "before" = left, "after" = right; Ouellet, Santiago, Funes, & Lupiánez, 2010), can also facilitate perception at their associated locations.

In some circumstances, however, the opposite may occur: Counter-intuitively, words with implicit spatial associations sometimes *hinder* identification of a visual target at their associated location. In the earliest demonstration of this *spatial interference effect*, Richardson, Spivey, Barsalou, and McRae (2003) presented brief sentences with either a vertical association (e.g., "The eagle flies to the river") or a horizontal association (e.g., "The miner pushes the cart"), followed by a visual target (■ or ●) on either the vertical axis (top or bottom of screen) or the horizontal axis (left or right). They found that vertically associated sentence cues slowed identification of targets along either end of the vertical axis. In a more fine-grained demonstration, Bergen, Lindsay, Matlock, and Narayanan (2007) similarly embedded spatial cue words within brief sentences, and they showed that high-associated cues (e.g., "The mule climbed") slowed identification of visual targets specifically at the top location, whereas low-associated cues (e.g., "The chair toppled") slowed identification at the bottom location. Estes, Verges, and Barsalou (2008) presented implicitly spatial cue words in isolation (e.g., "hat"), again demonstrating location-specific interference even without any semantic reference frame. Here two experiments are reported that investigate the conditions under which this spatial interference effect may or may not replicate.

Why Replicate Spatial Interference?

The spatial interference effect warrants replication for several reasons. (1) The effect is surprising, and surprising effects are relatively likely to be false-positives (Forstmeier, Wagenmakers, & Parker, 2017). (2) The spatial interference effect has been interpreted as an important source of evidence for *grounded cognition*, as explained below. (3) Consequently, those early demonstrations of spatial interference have had some impact on the field. Richardson et al. (2003), Bergen et al. (2007), and Estes et al. (2008) collectively have accrued 1265 citations on Google Scholar, and 618 citations in Scopus (both retrieved 31 October 2024). (4) A series of replication failures in Italian has been reported (Petrova et al., 2018). In summary, because the spatial interference effect is counter-intuitive, it has had some theoretical impact, but it may be a false-positive effect, and at this point, the true state of the effect is unknown.

Defining the Spatial Interference Effect

The original demonstrations of spatial interference, and most of the replication failures, shared several methodological commonalities. Those commonalities, which will be delineated below, will be taken as a working definition of the spatial interference effect.

Multiple Cue Categories. Prior demonstrations of spatial interference have used cue words from various semantic categories (e.g., animals, clothing, vehicles, etc.). Studies in which the cue words are from a single category (e.g., house-related words such as "attic" and "cellar") do not elicit spatial interference (Gozli, Chasteen, & Pratt, 2013).

Short SOA. Stimulus onset asynchrony (*SOA*) is the delay between cue and target onsets. Estes et al. (2008) presented cue words for 100 ms, followed by a blank delay of 50 ms, and finally the visual target (i.e., SOA = 150 ms). It has been shown that with SOAs longer than about 400 ms, facilitation may occur instead of interference (Goodhew, McGraw, & Kidd, 2014; Gozli et al., 2013; Zhang, Luo, Zhang, Wang, Zhong, & Li, 2013).

Nonsemantic Targets. The critical factor that makes the spatial interference effect so counter-intuitive is the use of nonsemantic targets. If semantically related targets are used instead, the result is rather intuitive: Cue words with spatial associations (e.g., "bird") *facilitate* recognition of the denoted object (i.e., an image of a bird) at its associated location (Estes et al., 2015). Tests of the spatial interference hypothesis, in contrast, use nonsemantic targets such as geometric shapes (e.g., ■ or ●) or alphanumeric characters (e.g., p or q).

Identification Task. Spatial interference occurs in the *identification task*, in which the visual target must be identified. For instance, Richardson et al. (2003) and Bergen et al. (2007) had participants press one or another button to identify whether the target was a square or a circle. A *detection task*, in which participants merely indicate the presence of a stimulus rather than identifying it, does not produce spatial interference (e.g., Dudschig, Lachmair, de la Vega, De Filippis, & Kaup, 2012; Gozli et al., 2013).

In sum, tests of spatial interference use cue words from multiple semantic categories followed shortly by nonsemantic targets in an identification task. In the General Discussion we consider why each of these factors affects spatial interference.

Theoretical Explanation of Spatial Interference

The spatial interference effect is thought to arise from two separable and counteracting components: (i) facilitation of attentional orienting and (ii) interference of object recognition. Those two components, in turn, are thought to arise from distinct processes of (i) linguistically mediated visual search and (ii) perceptual simulation of the denoted object.

Linguistic Orienting. Spatial interference may be partially understood in terms of linguistically mediated visual search (Estes et al., 2015; Gozli, Pratt, Martin, & Chasteen, 2016). A wealth of evidence from the "visual world paradigm," in which people hear spoken language while viewing object arrays, indicates that words elicit a visual search for semantically related objects (for review see Huettig, Rommers, & Meyer, 2011). For example, hearing the word "cake" leads people to fixate on an image of a cake in a visual scene (e.g., Altmann & Kamide, 1999). Moreover, the visual search for target objects is not random; rather, people systematically search for objects in the locations where they occur most often (i.e., contextual cueing; Chun & Jiang, 1998). The word "bird," for instance, elicits a search for a bird-related image toward the top of a display. Thus, for words with implicit spatial associations, the visual search is biased toward the associated location (Estes et al., 2008). This linguistic orienting is most simply shown via eye-tracking, where saccade launches are faster toward the word's associated location (Dudschig, Souman, Lachmair, de la Vega, & Kaup, 2013; Dunn, Kamide, & Scheepers, 2014). For instance, after hearing or reading "bird," saccades are initiated faster upward than downward. This effect is also evident in ERP studies, where targets appearing in the cue's associated location evoke a larger N1 response, which is linked to attentional shifts (Zhang et al., 2013).

Perceptual Simulation. Spatial interference may also be partially understood in terms of grounded cognition (Barsalou, 2008). Essentially, words evoke a perceptual simulation of the denoted object or event, which entails a reactivation of the neural patterns involved in prior experiences of that object or event (Barsalou, 1999, 2008, 2016). For example, the word "bird" may partially reactivate the neural pattern involved in actual perception of a real bird, including its appearance, sound, and so on. It may also reactivate a typical situation in which we experience birds, including typically co-occurring objects such as trees and contexts such as hiking in a forest. Thus, perceptual simulation is one mechanism by which situation models (van Dijk & Kintsch, 1983; Zacks & Tversky, 2001; Zwaan, 2016) are constructed and updated during language comprehension.

Presumably due to perceptual simulation, words facilitate recognition of the denoted object. That is, "bird" speeds recognition of bird-related images by pre-activating the perceptual representation of a bird. And conversely, objects are recognized more slowly when preceded by a semantically unrelated word, compared to a semantically related word or no word (e.g., Lupyan & Ward, 2013). For instance, the word "bird" hinders recognition of an apple. Proponents of grounded cognition attribute this interference effect to neural or perceptual competition (e.g., Bergen et al., 2007; Estes et al., 2008): Perceptual simulation of a word neurally competes with or perceptually masks the unrelated target object. That is, "bird" pre-activates the perceptual representation of a bird, which interferes with the perceptual identification of an apple.

Location-Specific Perceptual Simulation. Neither linguistic orienting nor perceptual simulation alone can explain the spatial interference effect. To begin with, the speeded orienting toward a cue word's associated location theoretically should *facilitate* perception at that location, not hinder it. Indeed, when the denoted object appears in its associated location, its recognition is facilitated (Estes et al., 2015; Gozli et al., 2016). For instance, "bird" speeds recognition of bird-related images at the top of a display. Critically, however, spatial interference occurs with semantically *unrelated* targets, such as when "bird" precedes a square target. Nor can perceptual simulation of the cue word fully explain the spatial interference, because the unrelated target is the same across varying locations (e.g., top, bottom), yet "bird" differentially hinders recognition of squares at those different locations.

The spatial interference effect thus appears to rely on the particular combination of linguistic orienting *and* perceptual simulation. That is, spatial interference appears to result from (i) an attention shift to the cue word's associated location and (ii) a perceptual simulation of the denoted object in that specific location. The cue word "bird" shifts attention to the top of the display and activates the perceptual representation of a bird. Thus, when a bird image appears at the top of the display, recognition is facilitated. When that bird image instead appears at the bottom of the display, recognition is slightly delayed (Estes et al., 2015), because attention must shift down from the top to the bottom location.

Less intuitive is the case when an unrelated object follows the cue word (e.g., "bird"). Regardless of the target's location, unrelated targets (e.g., a square) are recognized substantially more slowly than related targets (i.e., a bird; Estes et al., 2015). That is, the perceptual simulation of the cue word substantially delays recognition of the unrelated target (Lupyan & Ward, 2013). If recognition of the unrelated target were simply a matter of overcoming the perceptual simulation of the cue word (e.g., awaiting its de-activation), then presumably that target should be recognized more quickly in the cue's associated location, because recognizing that target in the opposite location would additionally require an attention shift down from the top to the bottom location. But in fact the opposite occurs: The unrelated target is recognized more *slowly* in the cue's associated location (i.e., spatial interference). Why?

It appears that the perceptual simulation of the cue word is location-specific. So when a square appears at the top of the display, the pre-activated perceptual representation of "bird" neurally competes with or perceptually masks recognition of that target at that location. Only after that bird representation dissipates can the square be identified. When the square instead appears at the bottom of the display, however, it requires an attention shift down to that bottom location. And critically, that attention shift appears to disengage the visual system from the bird representation at the top of the visual field, allowing faster recognition of the square at that bottom location. That is, because the perceptual representation of "bird" occurs at the top location, it creates stronger neuro-perceptual competition at the associated location than at other, non-cued locations. And thus, recognition of a square is faster at the bottom than at the top location because, evidently, shifting attention away from the perceptual representation of "bird" is faster than waiting for that perceptual representation to dissipate.

In sum, spatial interference appears to arise from a location-specific perceptual simulation of the cue word, which competes with or masks the unrelated visual target at the cue's associated location. This explanation, however, assumes that the spatial interference effect is indeed real and reliable. And the evidence of that is mixed.

Prior Evidence of Spatial Interference

All known tests of the spatial interference hypothesis are summarized in Table 1.

Successes. Bergen et al. (2007, Experiments 1 and 2) twice demonstrated spatial interference with brief sentences (e.g., "The mule climbed"). Estes et al. (2008) demonstrated the effect twice with word pairs (e.g., "cowboy hat;" Experiments 1 and 2) and once with single-word cues (e.g., "hat;" Experiment 3). Verges and Duffy (2009) replicated that effect twice with noun cues (e.g., "bird") and once with verb cues (e.g., "rise"). Gozli et al. (2013, Experiments 3, 4, and 6) replicated it a further three times with noun cues, and Estes et al. (2015, Experiments 3 and 4) replicated it once more with concrete nouns (e.g., "bird") and once with abstract nouns (e.g., "truth"). Finally, Petrova et al. (2018, Experiment 7) replicated the effect when they explicitly directed participants' attention to the cue words' spatial associations. Thus, spatial interference has been demonstrated fourteen times.

Failures. Bergen et al. (2007, Experiments 3 and 4) twice failed to obtain spatial interference with metaphorical cues (e.g., "The market sank"). Petrova, Sulpizio, Navarrete, Job, Suitner, and Peressotti (2013) failed to replicate the spatial interference effect in the absence of semantic context. In the largest test of spatial interference to date, Estes (2016, personal communication) also failed to obtain spatial interference. As part of the Reproducibility Project (Nosek et al., 2015), Renkewitz and Müller (2015) failed to replicate the effect. Most recently, Petrova et al. (2018) reported a series of ten replication attempts, nine of which failed to replicate the spatial interference effect. Thus, fourteen failures to replicate the spatial interference effect have been reported in the literature.

Weighing the Evidence. In total, there have been 28 known tests of the spatial interference hypothesis by five independent research groups (see Table 1). Yet, the evidential status of spatial interference remains equivocal: The effect has been successfully obtained fourteen times, and fourteen failures to replicate the effect have also been reported. So, is the effect real or not? On one hand, fourteen successful demonstrations of spatial interference seem too many for them all to be false-positive Type I errors. Moreover, because four independent research groups have found the effect, it is also unlikely to be attributable to methodological idiosyncrasies. On the other hand, fourteen known tests of spatial interference have failed to replicate the effect, and due to publication bias, there may well be more. Although nearly all of those replication attempts were substantially underpowered, it seems unlikely that they are all false-negative Type II errors. In order to achieve 80% power to reject a small effect, a replication study must have a sample that is about 2.5 times larger than the original sample (Simonsohn, 2015). In fact the majority of the replication attempts actually had *smaller* samples than the original (see Table 1).

Thus, there appears to be valid evidence both for and against the veracity of a spatial interference effect. How can this apparent discrepancy be reconciled?

Table 1. Prior tests of the spatial interference hypothesis.

Source, Study (Condition)	N	Orthography	Outcome
Bergen et al. (2007)			
1	63	opaque	success
2	59	opaque	success
3	59	opaque	failure
4	64	opaque	failure
Estes et al. (2008)			
1	18	opaque	success
2 (unmasked)	26	opaque	success
3	27	opaque	success
Verges & Duffy (2009)			
1 (words)	25	opaque	success
2 (nouns)	48	opaque	success
2 (verbs)	48	opaque	success
Gozli et al. (2013)			
3	26	opaque	success
4	40	opaque	success
6	25	opaque	success
Petrova et al. (2013)			
1	24	transparent	failure
Estes et al. (2015)			
3	52	opaque	success
4	39	opaque	success
Renkewitz & Muller (2015)			
1	22	transparent	failure
Estes (2016)			
1	116	transparent	failure
Petrova et al. (2018)			
1	39	transparent	failure
2	39	transparent	failure
3	20	transparent	failure
4	18	transparent	failure
5	20	transparent	failure
6	24	transparent	failure
7	25	transparent	success
8 (biased)	20	transparent	failure
8 (neutral)	20	transparent	failure
9	40	opaque	failure

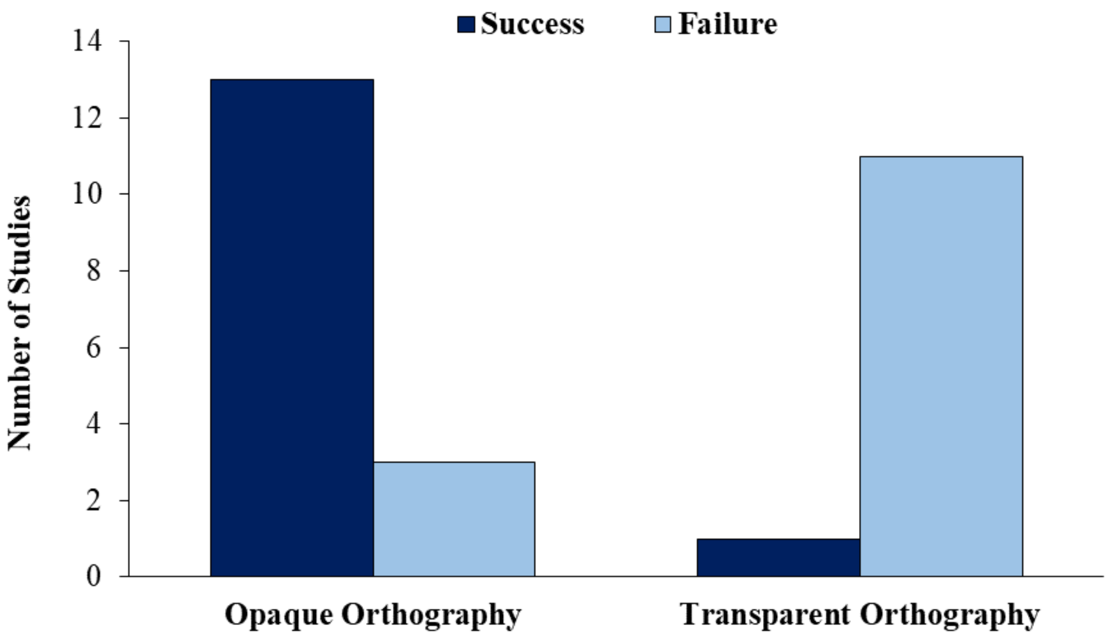
Orthographic Transparency

Estes and Barsalou (2018) noted that prior tests of spatial interference conducted in English tended to produce a significant effect, whereas tests in other languages (i.e., Italian or German) tended to produce no effect. In attempt to understand this pattern, they searched for relevant language properties on which (a) Italian and German are similar to one another and

(b) both Italian and German differ from English. One salient factor that fits this description is *orthographic transparency*, which refers to the consistency of print-to-sound correspondences within a language. In orthographically "transparent" languages, a given letter (or string of letters) tends to be pronounced the same across different words. In orthographically "opaque" languages, in contrast, a given letter (or string of letters) may be pronounced in different ways across different words. In English, for instance, the letter "o" has a soft pronunciation in "on" but a hard pronunciation in "no." Orthographic transparency is a matter of degree, and as it turns out, Italian and German both have more transparent orthography than English. More specifically, Italian orthography may be considered transparent, whereas German is semi-transparent, and English has opaque orthography (Schmalz, Marinus, Coltheart, & Castles, 2015).

Figure 1 illustrates the reliability of spatial interference as a function of orthographic transparency. In English, an opaque orthography, there have been sixteen known tests of spatial interference. Thirteen of those (81%) produced significant effects. In more transparent orthographies such as Italian and German, there have been twelve known tests of spatial interference, eleven of which (92%) failed to replicate the effect. In a meta-analysis of these 28 tests of spatial interference, Estes and Barsalou (2018) found that orthographic transparency significantly moderated the effect. Specifically, in English the overall effect was significant and of moderate size (19 ms, $p < .001$). In more transparent languages, however, there was no spatial interference effect (1 ms, $p = .44$). Thus, they identified orthographic transparency as a "hidden moderator" of spatial interference. But two limitations of that observation are important to note here. First, this presumed moderator was identified post hoc, and has not been directly tested. Second, this presumed moderation is merely descriptive. Spatial interference does indeed appear substantially more reliable in opaque languages, but *why*?

Figure 1. Prior studies in an orthographically opaque language (English) more often successfully demonstrated spatial interference, whereas studies in more orthographically transparent languages (Italian, German) more often failed to replicate the effect.



Semantic Processing

Spatial congruence effects are sensitive to semantic processing (Lebois, Wilson-Mendenhall, & Barsalou, 2015; Santiago, Oullet, Roman, & Valenzuela, 2012). For instance, Meier and Robinson (2004) showed that positive words (e.g., "love") are evaluated more quickly when presented at the top of a display, whereas negative words (e.g., "hate") are evaluated more quickly at the bottom. Subsequently, however, this valence-space congruence effect was shown to be affected by attention to the words' meanings. Brookshire, Ivry, and Casasanto (2010) showed that this effect occurred when distractor trials required a semantic judgment, but not when they required a perceptual judgment. Santiago et al. (2012) replicated the effect only when they oriented participants' attention to either the meaning of the word or the word's spatial location on the display. Lebois et al. (2015) further showed that the effect occurred only when participants judged the words' spatial associations. Thus, semantic processing appears to influence spatial congruence effects.

Languages vary in the extent to which they involve semantic processing during reading (Katz & Frost, 1992; Schmalz, Marinus, Coltheart, & Castles, 2015). Word reading entails converting graphemes (letters) to phonemes (sounds), and in transparent orthographies such as Italian, the highly consistent mapping of letters to sounds allows words to be read directly, with relatively little activation of lexical-semantic representations (Burani, Arduino, & Barca, 2007; Kwok, Cuetos, Avdyli, & Ellis, 2017; Peressotti & Job, 2003; Schmalz , Marinus, Coltheart, & Castles, 2015). That is, words can be read with relatively little activation of their meanings (i.e., *nonsemantic reading*). In fact, computational models that entirely lack a semantic system nonetheless can correctly read Italian words with up to 98% accuracy (Pagliuca & Monaghan, 2010). In contrast, in opaque orthographies such as English, due to the highly inconsistent grapheme-phoneme mappings, many words cannot be read correctly via phonological rules. Such "exception words" with irregular pronunciation can be read correctly only by accessing the lexical-semantic system (i.e., *semantic reading*), and moreover, the high prevalence of exception words induces semantic processing in general, even when reading words with regular pronunciation.

Several lines of evidence confirm that semantic processing is more robust when reading in opaque orthographies (e.g., English) than in transparent orthographies (e.g., Italian). First, brain areas involved in phoneme processing are more strongly activated when reading in Italian, but brain areas involved in lexical-semantic processing are more strongly activated when reading in English (Paulesu et al., 2000). Second, semantic factors such as imageability and age-of-acquisition have more robust effects on reading in English (Balota, Cortese, Sergent-Marshall, Spieler, & Yap, 2004) than in Italian or other transparent orthographies (Barca, Burani, & Arduino, 2002; Bates, Burani, D'Amico, & Barca, 2001; Buchanan & Besner, 1993; Burani et al., 2007; see also Bakhtiar & Weekes, 2015). Finally, semantic priming is more robust in English and other opaque orthographies (Hutchison et al., 2013) than in transparent orthographies such as Italian (Frost, Katz, & Bentin, 1987; Peressotti & Job, 2003; Tabossi & Laghi, 1992). Thus, a great deal of theoretical and empirical research indicates that semantic processing is stronger when reading in English than in Italian.

This is not to say that semantic processing never occurs when reading in Italian, of course. Rather, tasks that typically induce semantic processing, and manipulations that experimentally induce semantic processing, also elicit semantic processing in Italian. For instance, the lexical decision and picture naming tasks elicit deeper semantic processing than the reading aloud task. Accordingly, semantic effects occur in Italian with lexical decisions and picture naming to a greater extent than with reading aloud (e.g., Bates et al., 2001; Burani et al., 2007). Moreover, when semantic processing is experimentally induced, such as by requiring semantic judgments (Peressotti & Job, 2003) or by including irregularly-pronounced words (Tabossi & Laghi, 1992), then semantic priming emerges also in Italian.

Although semantic processing can be observed in Italian – either by experimentally inducing it or by using tasks that naturally entail it – simple reading in Italian does not naturally elicit deep semantic processing (Burani et al., 2007; Kwok et al., 2017; Peressotti & Job, 2003; Schmalz et al., 2015). And critically, the linguistic cueing paradigm that is used to test the spatial interference effect does not require semantic processing. Participants do not respond to the cue words in any way, the cue words do not predict the location of the subsequent target, and indeed, the task can be completed successfully without even reading the cue words. Thus, we suggest that prior tests of spatial interference in Italian may not have induced sufficiently deep semantic processing of the cue words.

Given that (i) semantic processing is necessary for spatial congruence effects, and (ii) semantic processing is more likely when reading in English than in Italian, it follows that the spatial interference effect should be more likely in English than in Italian. In other words, lack of spatial interference in Italian may be attributable to insufficient semantic processing of the cue words.

The Present Research

Spatial interference has been demonstrated thirteen times in English, and has failed to replicate ten times in Italian. Little would be learned by attempting to either replicate the effect again in English, or fail to replicate the effect again in Italian. In the present research we hypothesized that spatial interference could be obtained in Italian by bolstering semantic processing during the task. To this end, a subtle and natural method for increasing semantic processing during reading was used, that is, words with irregular stress were included.

Lexical stress refers to the prominence given to a certain syllable when pronouncing a polysyllabic word. Stress consists of a wide range of phonetic properties, such as loudness, vowel length, and pitch. Within some languages, the same syllable may be stressed in most polysyllabic words. In Italian, for instance, about 70% of three-syllable words have stress on the penultimate syllable (Colombo & Zevin, 2009; Spinelli, Sulpizio, & Burani, 2017). Such words, which have stress on the typical syllable within the language, are said to have *regular stress* (henceforth “regular words”). Others, in which a different syllable is stressed, have *irregular stress* (henceforth “irregular words”). Regular words can be read via sublexical processing, with relatively little semantic activation, based on the statistical-distributional knowledge that readers acquire about their language (Colombo & Zevin, 2009; Sulpizio, Burani, & Colombo, 2015). In contrast, irregular words more strongly activate the lexical-semantic representation in order to retrieve the correct pronunciation (Colombo, 1991; Colombo & Zevin, 2009; Sulpizio et al., 2015). That is, irregular stress can induce deeper semantic processing even in transparent orthographies. For example, age-of-acquisition (a semantic factor) affects reading of irregular words but not of regular words (Wilson, Ellis, & Burani, 2012).

Interestingly, when regular words occur in the context of many irregular words, then the regular words are also processed more semantically. For instance, semantic priming typically does not occur in reading aloud Italian words. When irregular words are added to the experimental list, however, semantic priming emerges for both the regular and irregular words (Tabossi & Laghi, 1992; see also Colombo & Tabossi, 1992). Thus, in the present research, we additionally included some irregular words among the cues. If the prior failures to obtain spatial interference in Italian (Estes, 2016; Petrova et al., 2013, 2018) were due to insufficient semantic processing of the cue words, then the inclusion of irregular cues should evoke spatial interference in Italian.

Experiments 1 and 2

Procedurally, Experiments 1 and 2 were prototypical tests of the spatial interference hypothesis. Single cue words with high (e.g., "hat") or low (e.g., "boot") spatial associations were presented centrally on a computer display, followed shortly (SOA = 150 ms) by an unrelated visual target (■ or ●) appearing at either the top or bottom of the display. Participants' task was simply to identify whether the target was a square or a circle. The experimental cue words, all of which had spatial associations and regular stress, were taken from Petrova et al. (2018). As in prior tests of spatial interference, there were *congruent* trials (i.e., those in which the target appeared in the cue's associated location, that is, high cue with top target and low cue with bottom target) and *incongruent* trials (i.e., those in which the target appeared in the opposite location, that is, high cue with bottom target and low cue with top target). The proportion of spatially congruent trials was 50% in both experiments, so that the cue's spatial association did not predict the target location. Thus, the experiments were close conceptual replications of Estes et al. (2008, Experiment 3).

Experiments 1 and 2 were identical, except that they included different filler cues. The irregular fillers in Experiment 1 had high or low spatial associations. That experiment provided strong conditions for obtaining spatial interference, because (1) the presence of 50% irregular cues should induce semantic processing of all cues, and (2) the presence of spatial associations in 100% of cues should ensure that those spatial associations are activated during that semantic processing. The irregular fillers in Experiment 2 instead had no spatial associations, thus providing a more conservative test of spatial interference, because only 50% of the cues had spatial associations. Thus, if the inclusion of irregular cues is sufficient for obtaining spatial interference in Italian, then a spatial interference effect of similar magnitudes should occur in Experiments 1 and 2. Alternatively, if a high proportion of spatially-associated cues is necessary for obtaining spatial interference in Italian, then the spatial interference effect should be larger in Experiment 1 (100% spatial cues) than in Experiment 2 (50% spatial cues). That is, comparison of Experiments 1 and 2 will test whether the proportion of spatially-associated cues moderates the effect.

Experiment 2 was preregistered (http://aspredicted.org/blind.php?x=4bb6zh), and all data and code for both experiments are available at the Open Science Foundation (available at https://osf.io/fbm7d/?view_only=46603bd6ea574f6eb987b96fe4cce18a). Given their high similarity, we report Experiments 1 and 2 together.

Methods

Sampling. Simonsohn (2015) recommended that replication samples should be about 2.5 times larger than the original sample. Given that the present experiments were close conceptual replications of Estes et al. (2008, Experiment 3), where $N = 27$, we sought a target N of about 68 participants in each of these two replication studies.

Participants. Students at an Italian university participated in exchange for course credit or a small reimbursement. All participants were native speakers of Italian, and all participated in only one experiment reported herein. Sixty-eight students (43 females, $M = 21.76$ years, $SD = 1.24$, *range* = 19-26) participated in Experiment 1, but three participants whose overall error rate was 20% or more were excluded, leaving 65 valid participants. Seventy students (44 females, $M = 21.34$ years, $SD = 1.31$, *range* = 19-24) participated in Experiment 2, and no participant committed more than 20% errors, so all were included in analyses. In total, then, there were 135 participants included in the analyses.

Stimuli. See the Supplementary Material for the full set of stimuli. Experimental cues were 24 regular words with a high ($n = 12$) or low ($n = 12$) spatial association, all taken from Petrova et al.'s (2018) Experiments 1-4, and selected from a spatial rating pretest (see the Appendix). To note, two of them (*chioma, funivia*) were added after the pretest. The spatial ratings of the high-association cues ($M = 6.22$, $SD = 0.44$, *Range* = 5.48-6.70) did not overlap with the low-association cues (*M* = 1.58, *SD* = 0.20, *Range* = 1.33-2.00). The same experimental cues were used in both Experiments 1 and 2.

Each experiment also included 24 irregular filler cues, also selected from the spatial rating pretest (see the Appendix). In Experiment 1, twelve of the filler cues had a high spatial association (*M* = 5.81, *SD* = 0.62, *Range* = 4.88-6.77), and twelve had a low association (*M* = 1.85, *SD* = 0.15, *Range* = 1.55-2.11). In Experiment 2, all 24 filler cues had neutral associations (*M* = 3.55, *SD* = 0.32, *Range* = 3.03-4.22).

Apparatus. Stimulus presentation, response times (RTs) and accuracy were controlled and recorded by E-Prime 2 (Psychology Software Tools, Inc., Sharpsburg, PA). Participants completed the experiment on a Lenovo notebook running Windows 10 with a 15.6 *in* monitor and a display resolution of 1366 x 768 pixels.

Procedure. This research complied with APA ethical standards for the treatment of participants, and it was approved by the ethics committee of the host university. Participants were tested individually in a sound-attenuated, uniformly lit room. They were seated approximately 60 cm from the monitor. Participants initiated each trial by pressing the spacebar, which triggered a central fixation cross that appeared for 250 ms, followed by the cue word, which appeared centrally for 100 ms. After a 50 ms delay, a target object (either a circle or a square) subtending approximately 5° of visual angle appeared at the top or bottom of the screen. Thus, as in Estes et al. (2008) and Petrova et al. (2018, Experiments 5-9), the SOA was 150 ms. Cues were presented in black on a white background in Courier New 18-point font. The "top" and "bottom" locations were centered horizontally approximately 9° vertically from the center of the display. Circle and square targets were also used by Petrova et al. (2018, Experiments 1 and 2). Cue Association (high, low), Target Location (top, bottom), and Target Object (circle, square) were fully crossed and balanced, such that each target object was equally likely to appear at each target location within each cue condition. This counterbalancing yielded eight stimulus lists of 96 trials each, with each participant assigned randomly to one of those eight lists.

Participants were instructed to identify the target object as quickly as possible, without making errors, by pressing the appropriate key ("C" or "M", as in Petrova et al., Experiments 4, 6, 7, 8) on a QWERTY keyboard. Half of the participants responded by pressing the "C" key with their left index finger when a circle appeared on the monitor, and the "M" key with their right index finger when a square appeared on the monitor. The other half were assigned to the opposite mapping.

The experiment consisted of 16 practice trials on which new non-spatial cue words were presented, and two experimental blocks of 48 trials each. Each cue was presented twice, once in each block. Trials were randomly presented within each block. Blocks were separated by a self-paced break, and order of blocks was counterbalanced across participants. The task took about 7 minutes to complete.

Data Analysis. RTs from trials with incorrect responses were excluded from analyses. Outlying RTs, defined as those more than 2.5 *SD*s from the participant's mean, were also excluded from analyses (Experiment 1: 2.84% of trials; Experiment 2: 2.99%). After completion of the experiments, we discovered that we had inadvertently included two irregular words (i.e., "aereo" and "sommergibile") among our experimental cues. We therefore report results with those two cues removed from all analyses.

We combined the data from Experiment 1 and 2 and analyzed them via linear mixed-effects models (LMM) using IBM SPSS Statistics 25. Following Barr, Levy, Scheepers, and Tily (2013), we first attempted to fit a maximal random effects model, with unstructured covariance and random slopes for Congruence across both subjects and items. Because that maximal model failed to converge, we then used a "step-down" strategy to identify the maximal model supported by the data (Barr et al., 2013; Bates, Kliegl, Vasishth, & Baayen, 2015; Matuschek, Kliegl, Vasishth, Baayen, & Bates, 2017). The maximal convergent model for both error rates and RTs had Congruence and Experiment as fixed effects and random intercepts for subjects only (Matuschek et al., 2017). The models were specified as:

lmer(Target.ERR ~ Congruence * Experiment + (1 | Subj_n), data = data);

lmer(Target.RT ~ Congruence * Experiment + (1 | Subj_n), data = data).

We dummy coded both Congruence (congruent = 0, incongruent = 1) and Experiment (Experiment 1 = 0, Experiment 2 = 1).

Results

Error rates were generally low in both Experiment 1 (overall $M = 2.76\%$, $SE = .38$) and Experiment 2 ($M = 3.40\%$, $SE = .46$), and they exhibited no significant effect of Congruence ($p = .788$), Experiment ($p = .141$), or their interaction ($p = .481$). We therefore do not consider error rates in any further analyses.

Response time results are summarized in Table 2, and the spatial interference effect is illustrated in Figure 2. The main effect of Congruence was significant, $F(1, 5514.42) = 6.98$, $p = .008$. As predicted, cue words slowed identification of targets in their associated location (i.e., spatial interference). The effect of Experiment was not significant, $p = .453$, and the interaction also failed to approach significance, $p = .503$, thus providing no evidence that the spatial interference effect was moderated by the proportion of spatially-associated cues. See the Supplemental Analyses for full details of individual parameter estimates.

For thoroughness and transparency, we also conducted several supplemental analyses that were intended to either facilitate comparison to prior tests of spatial interference (i.e., *t*-tests, *ANOVA*s, and Bayesian hypothesis tests) or investigate the robustness of the effect (i.e., inclusion of filler trials, counterbalancing checks, and alternative outlier detection methods). The outcomes largely align with the results of the linear mixed models reported above. See the Supplementary Material for further details.

Figure 2. The spatial interference effect in Experiments 1 and 2, and in a combined analysis of Experiments 1 and 2. Bars indicate ± 1 *SE*, corrected for within-participant designs (Loftus & Masson, 1994).

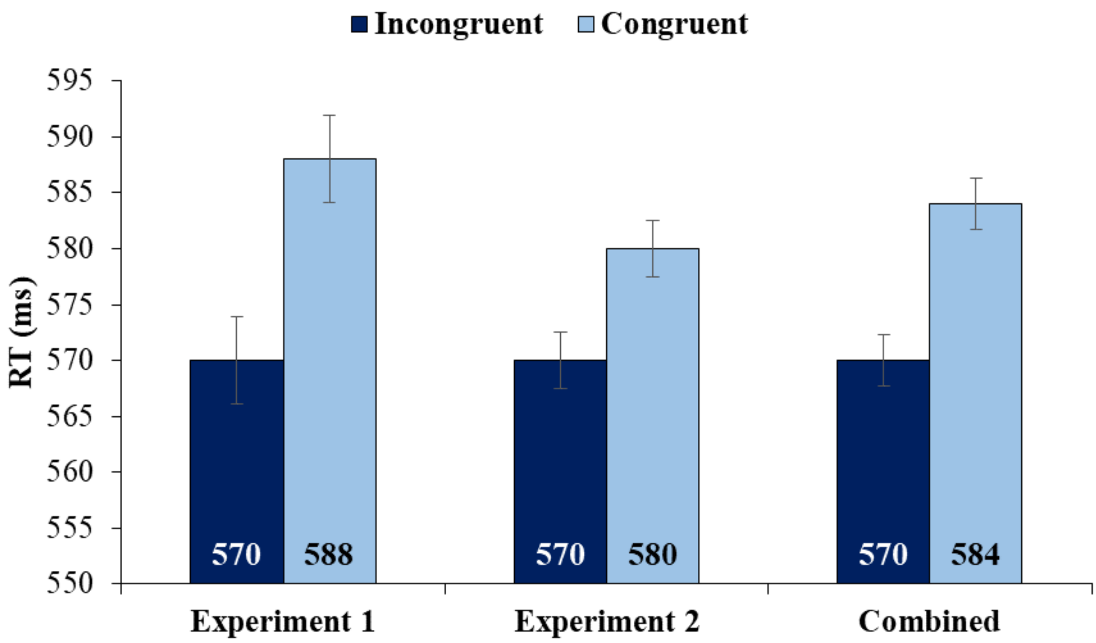


Table 2. Mean response times (RTs) and error rates (ERs; with standard deviations in parentheses) as a function of Condition (Incongruent, Congruent) in Experiments 1 and 2 and in the Combined Analysis of Experiments 1 and 2. * $p < .05$; ** $p < .01$.

	Experiment 1		**Experiment 2**		**Combined Analysis**	
Condition	**RTs (ms)**	**ERs (%)**	**RTs (ms)**	**ERs (%)**	**RTs (ms)**	**ERs (%)**
Incongruent	570 (135)	2.6 (3.8)	570 (117)	3.6 (5.1)	570 (126)	3.1 (4.5)
Congruent	588 (168)	2.8 (3.7)	580 (126)	3.1 (4.4)	584 (147)	3.0 (4.0)
Effect	**18***		**10***		**14****	

General Discussion

These results replicate the spatial interference effect. Notably, the effect was shown here in Italian. Inducing semantic processing of the *experimental* cues via a rather subtle manipulation of the regularity of the *filler* cues was sufficient to reveal spatial interference in an orthographically transparent language, where many prior attempts have failed (see Table 1). The standardized effect size was moderate, with spatial congruence accounting for 6.6% of the variance in target identification times. The raw effect size was 14 ms, which is comparable to the meta-analytic effect size observed in English under otherwise comparable conditions (i.e., 17 ms; Estes & Barsalou, 2018). These results provide the first demonstration that spatial interference can be obtained reliably in an orthographically transparent language.

Re-Weighing the Evidence. The spatial interference effect has been the subject of some controversy, having been demonstrated fourteen times by several independent research groups, and also having failed to replicate at least fourteen times by several other independent research groups (see Table 1). The present research, by adding two successful replications of the effect, does *not* tip the balance of evidence in favor of the effect's reliability simply as a matter of score-keeping: 16 for the defense to 14 for the challengers. Such simple counting is not how bodies of evidence are evaluated. There are other, more important factors such as methodological fidelity (i.e., "closeness" of the replication attempt), strength of the manipulation, sensitivity of the measurement, and statistical power (primarily affected by sample size) to detect the hypothesized effect. Aside from their small samples, prior tests of the spatial interference hypothesis generally were methodologically sound. Therefore, there is little point in trying to identify "better" or "worse" replication attempts among this literature.

The present research *does* tip the balance of evidence in favor of the effect's reliability, but for a reason other than simple counting: This research provides the first test of a previously-hidden moderator. Estes and Barsalou (2018) noted, post hoc, that most tests of spatial interference in orthographically opaque languages (e.g., English) were successes, whereas most tests in more transparent orthographies (i.e., Italian or German) were failures (see Table 1). However, because orthographic transparency is a property of languages, it cannot be manipulated experimentally, rendering direct tests of this hypothesized moderator impossible. In the present research, this methodological limitation was circumvented by inducing participants to process an orthographically transparent language as if it were an opaque language (i.e., by inducing deeper semantic processing). In this way, a reliable demonstration of spatial interference in a transparent language was obtained. Thus the present research explains why some prior tests successfully obtained spatial interference and others failed to do so. Consequently, this research strongly supports the reliability of the effect by providing a systematic explanation of the conditions under which it does or does not occur, as described next.

Theoretical Implications. These results suggest that prior failures to obtain spatial interference in more orthographically transparent languages such as Italian and German (i.e., Estes, 2016; Petrova et al., 2013, 2018; Renkewitz & Müller, 2015) were likely due to insufficient semantic processing. Because transparent languages have highly consistent spelling-to-sound correspondences, words can be read with relatively little semantic processing (Bates et al., 2001; Buchanan & Besner, 1993; Burani et al., 2007; Frost et al., 1987; Katz & Frost, 1992; Kwok et al., 2017; Pagliuca & Monaghan, 2010; Peressotti & Job, 2003; Schmalz et al., 2015). And because semantic processing appears to be necessary for spatial congruence effects to occur (Brookshire et al., 2010; Lebois et al., 2015; Santiago et al., 2012; Shaki & Fischer, 2023a, 2023b), these more orthographically transparent languages typically fail to elicit spatial interference.

In contrast, the inconsistency of spelling-to-sound correspondences in opaque orthographies such as English elicits a stronger reliance on semantic processing during word reading (Frost et al., 1987; Katz & Frost, 1992; Kwok et al., 2017; Schmalz et al., 2015). Critically however, in transparent languages, words with irregular stress also induce deeper semantic processing during reading (Tabossi & Laghi, 1992; Wilson et al., 2012). The present experiments demonstrate that, when such irregular words are included among the cue words, spatial interference is also observed in an orthographically transparent language. This observation suggests that prior failures to replicate spatial interference in more transparent languages may well have been attributable to insufficient semantic processing.

Consistent with prior research from other paradigms (Brookshire et al., 2010; Lebois et al., 2015; Santiago et al., 2012), the present results further suggest that relatively deep semantic processing is necessary for spatial interference. This conclusion also provides a unifying explanation of the previously observed moderators of the spatial interference effect. First, spatial interference does not occur when only a single cue category is used (Gozli et al., 2013). Presumably, with only one cue category, the perceptual simulation of that cued category becomes strongly activated within the first few trials of the experiment. After those first few trials, processing of the given scenario or event no longer requires as much neural and/or perceptual resources. Consequently, the cues produce less neural and perceptual competition with the target stimulus, thereby eliminating the spatial interference effect (Ostarek & Vigliocco, 2017). Second, spatial interference does not occur with long SOAs (Gozli et al., 2013). At short SOAs, the cue word is thought to evoke an attention shift toward the associated location, and a perceptual simulation of the denoted object or event. That location-specific simulation, in turn, is thought to perceptually or neurally compete with (or "mask") identification of the visual target in that location. At longer SOAs, however, the perceptual simulation begins to dissipate, leaving visual attention in the associated location without perceptual competition. Third, spatial interference occurs in identification tasks but not in detection tasks (Gozli et al., 2013). This is because spatial interference arises from the semantic incongruence of the cue and target, but the detection task does not require and hence may not always evoke deep semantic processing of the target. The common denominator among all these known moderators of spatial interference, including orthographic transparency (Estes & Barsalou, 2018), is more or less semantic processing. Collectively, these moderations can be summarized as follows: If the cues are processed semantically and the targets are unrelated to those cues, spatial interference tends to occur. If the cues are not processed at a sufficiently deep semantic level, then interference does not occur.

Future Directions. Six promising directions for further research that may be theoretically informative of spatial interference could be identified. One striking aspect of the spatial interference effect is that it occurs despite the fact that the cue words are entirely unrelated to the target identification task. The task can be completed error-free without even reading the cue words, and hence reading the cues is purely incidental to task performance. In fact, the cue words and their spatial associations do not predict the location or identity of the subsequent target, so reading the cues could not possibly improve performance. Some studies, however, required semantic judgment of the cue words (e.g., Amer et al., 2017; Gozli et al., 2013; for related tasks see also Brookshire et al., 2010; Lebois et al., 2015; Peressotti & Job, 2003; Petrova et al., 2018; Santiago et al., 2012; Shaki & Fischer, 2023a). It may be informative to examine whether the presence or absence of such semantic judgment moderates the magnitude of spatial interference, especially in orthographically transparent languages, where the cue words may otherwise be processed semantically only weakly.

A second avenue for further research is to re-examine the boundary or generality of spatial interference. In the introduction, based on prior tests of the hypothesis, spatial interference was operationally defined as occurring with nonsemantic targets such as geometric shapes (e.g., circle and square) and alphanumeric characters (e.g., p and q). It remains an open question, however, whether the targets must actually be nonsemantic, or whether they need be only semantically unrelated. For instance, Estes et al. (2015) included targets that were semantically unrelated to the cues, such as the word “bird” followed by an image of a wrench at either the top or bottom of the display. They found significant spatial interference to those semantically unrelated targets. On the other hand, Ostarek and Vigliocco (2017) similarly tested semantically unrelated targets, but there the targets also had their own spatial associations. For instance, “sky” preceded an image of a hat (which also has a high association). And in that case, there was no spatial interference, though it may simply have been overshadowed by the target’s own spatial association. Thus, it is currently not entirely clear whether spatial interference requires nonsemantic targets, or semantically unrelated targets.

Third, the salience of the cues’ spatial associations might be an important direction for additional research. Petrova et al. (2018), in their Experiment 7, explicitly informed participants that the cue words had spatial associations. Those “biased” instructions, which render spatial associations highly salient, produced the only prior demonstration of spatial interference in an orthographically shallow language (see Table 1). However, when Petrova et al. conducted an exact replication in their Experiment 8, they obtained the exact opposite result, finding instead a tendency toward spatial *facilitation* (p = .061). In the present research, the salience of spatial associations was manipulated by varying the proportion of cue words that had spatial associations across experiments, but no difference in spatial interference across those experiments emerged. Given this empirical ambiguity, these results collectively suggest that the salience of spatial associations may indeed be relevant to the spatial interference effect, but its effect (if any) appears to be complex.

A fourth direction for theoretical advance is to more thoroughly examine the spatial interference effect in semi-transparent languages such as German. As far as we are aware, only a single test of spatial interference has been conducted in German: Renkewitz and Müller (2015) failed to obtain spatial interference in German, but their study might be considered underpowered. Semi-transparent languages are theoretically interesting to study because they tend to entail a moderate amount of semantic processing during language comprehension. Would a semi-transparent language such as German produce a spatial interference effect midway between Italian (a transparent language) and English (an opaque language)? Or does the mere presence of some moderate amount of orthographic complexity in the language induce deep semantic processing, such that spatial interference effects are equally large in German and English? Large-scale, cross-language tests of linguistic cueing are needed to address this question.

It would also be theoretically informative to test for spatial interference with auditory presentation of linguistic cues. Twenty-four of the 28 prior tests of spatial interference (see Table 1) used visual presentation of written cues, as in the present experiments. Given that orthographic transparency is a property of *written* language – i.e., the consistency of spelling-to-sound correspondence – we see no reason why orthographic transparency would moderate spatial interference with *auditory* presentation of linguistic cues. So, would spatial interference occur in transparent languages (e.g., Italian) with auditory presentation of cues? By demonstrating that spatial interference can also occur in an orthographically shallow language, the present experiments reveal that it is not *orthographic* depth per se that moderates spatial interference. Rather, the true hidden moderator is *semantic* depth, or the extent to which the linguistic cues elicit semantic processing (see also Shaki & Fischer, 2023a). And critically, spoken language also elicits varying degrees of semantic processing (Sanford & Sturt, 2002). The present experiments thus suggest that spatial interference from auditory cues likely depends on the semantic depth of cue processing both within and across languages. Indeed, of all prior studies of the spatial interference effect (see Table 1), only Bergen et al. (2007) presented the linguistic cues auditorily, and they obtained spatial interference twice with literal cues (e.g., "The mule climbed"; Experiments 1 and 2), but failed to obtain spatial interference twice with metaphorical cues (e.g., "The market sank"; Experiments 3 and 4). More research with auditory cue presentation is needed to disentangle the potential roles of orthographic and semantic depth in spatial interference.

Finally, these results raise implications for linguistic cueing effects more generally, beyond spatial interference. If orthographically transparent languages do not typically induce deep enough semantic processing of cue words to elicit the spatial interference effect, as we argue, then presumably such languages may also fail to elicit the more common spatial congruence effect, whereby cue words instead facilitate perception at the associated location (e.g., Hommel, Pratt, Colzato, & Godijn, 2001). Aside from the tests of spatial interference in Italian that we reviewed extensively above, our literature search revealed only one other investigation of linguistic cueing in an orthographically transparent language: Ouellet, Santiago, Funes, and Lupiánez (2010) centrally presented time-related words such as "before" and "after" in Spanish, and then tested perception of visual targets on the left or right of the display. Across three experiments, they obtained spatial congruence effects, such that past-related cues facilitated perception at the left location and future-related words sped perception on the right. At face value, this finding seems to contradict the implication that transparent languages do not typically elicit linguistic cueing effects. Crucially, however, Ouellet et al. explicitly required participants to semantically process the cue words during all three of their experiments. Thus, Ouellet et al. did obtain a linguistic cueing effect in a transparent language, but as in the present experiments, it occurred with relatively deep semantic processing of the cues. As for why Ouellet et al. found a congruence effect instead of interference, we note that their experiments did not have the conditions under which spatial interference tends to occur. Specifically, Ouellet et al. used long SOAs, which are known to elicit facilitation instead of interference (Goodhew, McGraw, & Kidd, 2014; Gozli et al., 2013). As explained in our introduction, the spatial interference effect tends to occur only with short SOAs (see "Defining the Spatial Interference Effect"), before the perceptual simulation of the cue word has dissipated (see "Theoretical Explanation of Spatial Interference"). Thus, more research is warranted to investigate more fully the conditions under which linguistic cueing effects in general (i.e., both congruence and incongruence effects) may occur in orthographically transparent languages.

Concluding Remark. The spatial interference effect has attracted relatively many replication attempts. Given the counter-intuitive nature of this effect, such replication attempts are not merely justified, but *necessary* for the integrity of the field. Surprising effects *should* be subjected to replication attempts, because surprising effects are relatively likely to be false-positive, Type I errors (Forstmeier, Wagenmakers, & Parker, 2017). Our knowledge of the underlying process(es) may become deeper and broader only if other, independent researchers continue testing for spatial interference via direct and conceptual replications. Spatial interference is a positive example of how counter-intuitive effects – and their subsequent replication failures – can advance theoretical understanding of the phenomenon further than mere confirmations of the effect.

References

Altmann, G. T., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, *73*(3), 247-264.

Amer, T., Gozli, D. G., & Pratt, J. (2017). Biasing spatial attention with semantic information: an event coding approach. *Psychological Research*, 1-19.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59, 390–412.

Bakhtiar, M., & Weekes, B. (2015). Lexico-semantic effects on word naming in Persian: Does age of acquisition have an effect?. *Memory & Cognition*, *43*(2), 298-313.

Balota, D. A., Cortese, M. J., Sergent-Marshall, S. D., Spieler, D. H., & Yap, M. J. (2004). Visual word recognition of single-syllable words. *Journal of Experimental Psychology: General*, *133*, 283-316.

Barca, L., Burani, C., & Arduino, L. S. (2002). Word naming times and psycholinguistic norms for Italian nouns. *Behavior Research Methods, Instruments, & Computers*, *34*(3), 424-434.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of memory and language, 68*(3), 255-278.

Barsalou. L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences, 22,* 577-660.

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology*, *59*, 617-645.

Barsalou, L. W. (2016). On staying grounded and avoiding Quixotic dead ends. *Psychonomic Bulletin & Review*, *23*, 1122–1142.

Bates, E., Burani, C., D'Amico, S., & Barca, L. (2001). Word reading and picture naming in Italian. *Memory & Cognition*, *29*(7), 986-999.

Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). Parsimonious mixed models. *arXiv preprint arXiv:1506.04967*.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of statistical software*, *67*, 1-48.

Bergen, B. K., Lindsay, S., Matlock, T., & Narayanan, S. (2007). Spatial and linguistic aspects of visual imagery in sentence comprehension. *Cognitive Science*, *31*(5), 733-764.

Brookshire, G., Ivry, R., & Casasanto, D. (2010). Modulation of motor-meaning congruity effects for valenced words. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 32, No. 32).

Buchanan, L., & Besner, D. (1993). Reading aloud: Evidence for the use of a whole word nonsemantic pathway. *Canadian Journal of Experimental Psychology*, *47*(2), 133-152.

Burani, C., Arduino, L. S., & Barca, L. (2007). Frequency, not age of acquisition, affects Italian word naming. *European Journal of Cognitive Psychology*, *19*(6), 828-866.

Chasteen, A. L., Burdzy, D. C., & Pratt, J. (2010). Thinking of God moves attention. *Neuropsychologia, 48*(2), 627-630.

Colombo, L. (1991). The role of lexical stress in word recognition and pronunciation. *Psychological Research*, *53*(1), 71-79.

Colombo, L., & Tabossi, P. (1992). Strategies and stress assignment: Evidence from a shallow orthography. In *Advances in psychology* (Vol. 94, pp. 319-340). North-Holland.

Colombo, L., & Zevin, J. (2009). Stress priming in reading and the selective modulation of lexical and sub-lexical pathways. *PloS one*, *4*(9), e7219.
Dienes, Z. (2008). Understanding Psychology as a Science: An Introduction to Scientific and Statistical Inference. Basingstoke: Palgrave Macmillan.
Dudschig, C., Lachmair, M., de la Vega, I., De Filippis, M., & Kaup, B. (2012). From top to bottom: spatial shifts of attention caused by linguistic stimuli. *Cognitive Processing*, *13*(1), 151-154.
Dudschig, C., Souman, J., Lachmair, M., de la Vega, I., & Kaup, B. (2013). Reading "sun" and looking up: The influence of language on saccadic eye movements in the vertical dimension. *PloS one*, *8*(2), e56872.
Dunn, B., Kamide, Y., & Scheepers, C. (2014, January). Hearing "moon" and looking up: Word-related spatial associations facilitate saccades to congruent locations. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 36, No. 36).
Estes, Z. (2016). Test of spatial interference. Unpublished data.
Estes, Z., & Barsalou, L. W. (2018). A Comprehensive Meta-Analysis of Spatial Interference From Linguistic Cues: Beyond Petrova et al. (2018). *Psychological Science*, *29*(9), 1558-1564.
Estes, Z., Verges, M., & Adelman, J. S. (2015). Words, objects, and locations: Perceptual matching explains spatial interference and facilitation. *Journal of Memory and Language*, *84*, 167-189.
Estes, Z., Verges, M., & Barsalou, L. W. (2008). Head up, foot down: Object words orient attention to the objects' typical location. *Psychological Science*, *19*(2), 93-97.
Forstmeier, W., Wagenmakers, E. J., & Parker, T. H. (2017). Detecting and avoiding likely false-positive findings–a practical guide. *Biological Reviews*, *92*(4), 1941-1968.
Frost, R., Katz, L., & Bentin, S. (1987). Strategies for visual word recognition and orthographical depth: a multilingual comparison. *Journal of Experimental Psychology: Human Perception and Performance*, *13*(1), 104-115.
Gibson, B. S., & Kingstone, A. (2006). Visual attention and the semantics of space: Beyond central and peripheral cues. *Psychological Science, 17*(7), 622-627.
Gibson, B. S., & Sztybel, P. (2014). The spatial semantics of symbolic attention control. *Current Directions in Psychological Science, 23*(4), 271-276.
Goodhew, S. C., McGaw, B., & Kidd, E. (2014). Why is the sunny side always up? Explaining the spatial mapping of concepts by language use. *Psychonomic Bulletin & Review*, *21*(5), 1287-1293.
Goslin, J., Galluzzi, C., & Romani, C. (2014). PhonItalia: A phonological lexicon for Italian. *Behavior Research Methods*, 46, 872–886. doi: 10.3758/s13428-013-0400-8.
Gozli, D. G., Chasteen, A. L., & Pratt, J. (2013). The cost and benefit of implicit spatial cues for visual attention. *Journal of Experimental Psychology: General*, *142*(4), 1028-1046.
Gozli, D. G., Pratt, J., Martin, K. Z., & Chasteen, A. L. (2016). Implied spatial meaning and visuospatial bias: Conceptual processing influences processing of visual targets and distractors. *PloS one*, *11*(3), e0150928.
Hommel, B., Pratt, J., Colzato, L., & Godijn, R. (2001). Symbolic control of visual attention. *Psychological science, 12*(5), 360-365.
Huettig, F., Rommers, J., & Meyer, A. S. (2011). Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta psychologica*, *137*(2), 151-171.
Hutchison, K. A., Balota, D. A., Neely, J. H., Cortese, M. J., Cohen-Shikora, E. R., Tse, C. S., ... & Buchanan, E. (2013). The semantic priming project. *Behavior Research Methods*, *45*(4), 1099-1114.

Katz, L., & Frost, R. (1992). The reading process is different for different orthographies: The orthographic depth hypothesis. In *Advances in psychology* (Vol. 94, pp. 67-84). North-Holland.

Kwok, R. K. W., Cuetos, F., Avdyli, R., & Ellis, A. W. (2017). Reading and lexicalization in opaque and transparent orthographies: Word naming and word learning in English and Spanish. *The Quarterly Journal of Experimental Psychology*, 70(10), 2105-2129.

Lebois, L. A. M., Wilson-Mendenhall, C. D., & Barsalou, L. W. (2015). Are automatic conceptual cores the gold standard of semantic processing? The context-dependence of spatial meaning in grounded congruency effects. *Cognitive Science*, *39*, 1764–1801.

Loftus, G. R., & Masson, M. E. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin & Review*, 1(4), 476-490.

Logan, G. D. (1995). Linguistic and conceptual control of visual spatial attention. *Cognitive psychology, 28*(2), 103-174.

Luke, S. G. (2017). Evaluating significance in linear mixed-effects models in R. *Behavior research methods*, *49*, 1494-1502.

Lupyan, G., & Ward, E. J. (2013). Language can boost otherwise unseen objects into visual awareness. *Proceedings of the National Academy of Sciences*, *110*(35), 14196-14201.

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of memory and language, 94*, 305-315.

Meier, B. P., & Robinson, M. D. (2004). Why the sunny side is up: Associations between affect and vertical position. *Psychological Science*, *15*(4), 243-247.

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251), aac4716.

Ostarek, M., & Vigliocco, G. (2017). Reading sky and seeing a cloud: On the relevance of events for perceptual simulation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(4), 579-590.

Ouellet, M., Santiago, J., Funes, M. J., & Lupiánez, J. (2010). Thinking about the future moves attention to the right. *Journal of Experimental Psychology: Human Perception and Performance, 36*(1), 17-24.

Pagliuca, G., & Monaghan, P. (2010). Discovering large grain sizes in a transparent orthography: Insights from a connectionist model of Italian word naming. *European Journal of Cognitive Psychology*, *22*(5), 813-835.

Paulesu, E., McCrory, E., Fazio, F., Menoncello, L., Brunswick, N., Cappa, S.F., Cotelli, M., Cossu, G., Corte, F., Lorusso, M., Pesenti, S., Gallagher, A., Perani, D., Price, C., Frith, C.D., & Frith, U. (2000). A cultural effect on brain function. *Nature Neuroscience*, 3, 91-96.

Pauszek, J. R., & Gibson, B. S. (2018). The Least Costs Hypothesis: A rational analysis approach to the voluntary symbolic control of attention. *Journal of experimental psychology: human perception and performance, 44*(8), 1199.

Peressotti, F., & Job, R. (2003). Reading aloud: Dissociating the semantic pathway from the non-semantic pathway of the lexical route. *Reading and Writing*, *16*(3), 179-194.

Petrova, A., Navarrete, E., Suitner, C., Sulpizio, S., Reynolds, M., Job, R., & Peressotti, F. (2018). Spatial congruency effects exist, just not for words: Looking into Estes, Verges, Barsalou (2008). *Psychological Science, 29,* 1195-1199.

Petrova A., Sulpizio S., Navarrete E., Job R., Suitner C. & Peressotti, F. (2013). Foot or X down? Response compatibility but no effects of the object's typical location. Several failures to replicate Estes, Verges & Barsalou (2008). Poster presented at the 18th Conference of the European Society for Cognitive Psychology (ESCOP), Budapest, Hungary, 29 August – 1 September.

Posner, M. I. (1980). Orienting of attention. *Quarterly journal of experimental psychology, 32*(1), 3-25.

Renkewitz, F. & Müller, S. (2015). Replication of Study 1 by Estes, Z., Verges, M., & Barsalou, L.W. (*Psychological Science*, 2008). Retrieved September 20, 2015 from https://osf.io/vwnit/

Richardson, D. C., Spivey, M. J., Barsalou, L. W., & McRae, K. (2003). Spatial representations activated during real-time comprehension of verbs. *Cognitive Science*, *27*(5), 767-780.

Sanford, A. J., & Sturt, P. (2002). Depth of processing in language comprehension: Not noticing the evidence. *Trends in cognitive sciences*, *6*(9), 382-386.

Santiago, J., Ouellet, M., Román, A., & Valenzuela, J. (2012). Attentional factors in conceptual congruency. *Cognitive Science*, *36*(6), 1051-1077.

Schmalz, X., Marinus, E., Coltheart, M., & Castles, A. (2015). Getting to the bottom of orthographic depth. *Psychonomic Bulletin & Review*, *22*(6), 1614-1629.

Shaki, S., & Fischer, M.H. (2023a). How does language affect spatial attention? Deconstructing the prime-target relationship. *Memory & Cognition, 51*(5), 115-1124.

Shaki, S., & Fischer, M. H. (2023b). Language directs spatial attention differently in explicit and implicit tasks. *Plos one, 18*(11), e0291518.

Simonsohn, U. (2015). Small telescopes: Detectability and the evaluation of replication results. *Psychological Science*, *26*(5), 559-569.

Spinelli, G., Sulpizio, S., & Burani, C. (2017). Q2Stress: A database for multiple cues to stress assignment in Italian. *Behavior research methods*, 49, 2113-2126.

Sulpizio, S., Burani, C., & Colombo, L. (2015). The process of stress assignment in reading aloud: Critical issues from studies on Italian. *Scientific Studies of Reading*, *19*(1), 5-20.

Tabossi, P., & Laghi, L. (1992). Semantic priming in the pronunciation of words in two writing systems: Italian and English. *Memory & Cognition*, *20*(3), 303-313.

van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. Academic Press.

Verges, M., & Duffy, S. (2009). Spatial representations elicit dual-coding effects in mental imagery. *Cognitive Science*, *33*(6), 1157-1172.

Wilson, M. A., Ellis, A. W., & Burani, C. (2012). Age-of-acquisition affects word naming in Italian only when stress is irregular. *Acta Psychologica*, *139*(3), 417-424.

Zacks, J. M., & Tversky, B. (2001). Event structure in perception and conception. *Psychological Bulletin, 127*(1), 3-21.

Zhang, E., Luo, J., Zhang, J., Wang, Y., Zhong, J., & Li, Q. (2013). Neural mechanisms of shifts of spatial attention induced by object words with spatial associations: an ERP study. *Experimental Brain Research*, *227*(2), 199-209.

Zwaan, R. A. (2016). Situation models, mental simulations, and abstract concepts in discourse comprehension. *Psychonomic Bulletin & Review, 23*(4), 1028-1034.

Zwaan, R. A., Etz, A., Lucas, R. E., & Donnellan, M. B. (2018). Making replication mainstream. *Behavioral and Brain Sciences*, *41*, 1-61.

Appendix

Spatial Rating Pretest. Regular and irregular Italian words, with or without spatial associations, were selected from a pretest for use as cue words in the experiments. Thirty participants (18 males, mean age = 33, SD = 8.6, range = 20-49) were recruited from the Prolific online panel. All declared to be Italian native speakers with acceptable to excellent competence in the Italian language. Additionally, a brief vocabulary test was administered at the beginning of the study. Four target words were each presented with four different words as response options, and participants were required to indicate which of the four option words is a synonym of the given target word (e.g. *tessuto-stoffa*). Only participants who correctly selected the synonym word for all four target words were included in the analysis.

Pretest stimuli were 162 words: 70 with regular stress and 92 with irregular stress. The 70 regular words, each containing two to five syllables with stress on the penultimate syllable, were sampled from Petrova et al. (2018). Twenty-two of those were spatial cues from their Experiments 1-4, and 48 were non-spatial cues from their Experiments 6-9. To note, Petrova et al.'s Experiments 1-3 included 24 spatial cues. However, two of them were homographs (*sole = sun, alone; formica = ant; Formica*) and two had irregular stress (*aquila, nuvola*). Those four were replaced with regular spatial cues from their Experiment 4 (*chioma, funivia, nido, tombino*). The 92 irregular words, each containing three syllables with stress on the first syllable, were retrieved from the PhonItalia database (Goslin, Galluzzi, & Romani, 2014). We included some words with high or low spatial associations, and some without spatial associations.

The pretest was administered online via Qualtrics survey software. Participants' task was to judge whether each word's referent (i.e., the denoted object or event) typically occurs high or low in the visual field, on a scale from 1 (low) to 7 (high). As an attention check, the number words from "one" to "seven" were also included. Participants were instructed to click the correspondent number on the scale when a number word appeared. The 169 words (162 cue words + 7 number words) were presented in random order. Overall, the task took about 10 minutes to complete. Three participants were excluded for failing to respond correctly on the number-word catch trials (fewer than 3 out of 7 correct responses), leaving 27 participants in the analyses. Results for the selected stimuli are reported in the main text (see *Stimuli* section).

Supplemental Analyses

The data were first analyzed via linear mixed-effects models (LMM) using IBM SPSS Statistics 25. Following Barr, Levy, Scheepers, and Tily (2013), we first attempted to fit a maximal random effects model, with unstructured covariance and random slopes for Congruence across both subjects and items. Because that maximal model failed to converge in both Experiment 1 and Experiment 2, we then used a "step-down" strategy to identify for each analysis the maximal model supported by the data (Barr et al., 2013; Bates, Kliegl, Vasishth, & Baayen, 2015; Matuschek, Kliegl, Vasishth, Baayen, & Bates, 2017). The maximal convergent model varied across analyses, as described next. (i) Error rates were analyzed with Congruence as a fixed effect and random intercepts for subjects, as more complex random effects structures did not converge (Matuschek et al., 2017). (ii) In Experiment 1, RTs were analyzed using variance components covariance structure (i.e., removing correlation parameters between random effects while maintaining random slopes), with Congruence as a fixed effect, random slopes for Congruence for both subjects and items, and random intercepts for subjects and items. (iii) In Experiment 2, RTs were analyzed with Congruence as a fixed effect and random intercepts for subjects and items (Matuschek et al., 2017). In all analyses, Congruence is dummy coded (congruent = 0, incongruent = 1), and in combined analyses of both experiments, Experiment is also dummy coded (Experiment 1 = 0, Experiment 2 = 1).

Separate Analyses of Experiments 1 and 2. The experimental trials (i.e., those with regular cues) were identical across experiments. When analysing those regular cues, the effect of Congruence was significant in both Experiment 1, $F(1, 20.73) = 4.47$, $p = .044$, and Experiment 2, $F(1, 2796.16) = 3.92$, $p = .048$. Results are summarized in Table 2 in the main text, and the spatial interference effect is illustrated in Figure 2 in the main text. In both experiments, as predicted, cue words slowed identification of targets in their associated location (i.e., spatial interference).

Combined Analysis Parameter Estimates. For completeness, we provide the individual parameter estimates for the combined linear mixed-effects model reported in the main text.

Table 1. *Summary of the model output*

Parameter	Estimate	SE	df	t	p	95% CI
(Intercept)	579.59	11.48	95.21	50.50	<.001	[556.80, 602.37]
Congruence [incongruent]	-10.12	7.14	5514.58	-1.42	.156	[-24.12, 3.87]
Experiment [1]	7.38	7.34	5541.74	1.01	.315	[-7.01, 21.77]
Congruence × Experiment	-6.87	10.26	5514.54	-0.67	.503	[-26.99, 13.25]

Note that while the omnibus Type III F-test for Congruence is significant ($p = .008$; as reported in the main text), the individual parameter estimate is not ($p = .156$). This discrepancy is a well-documented phenomenon in mixed-effects models with categorical predictors and interactions. The omnibus Type III F-test evaluates the overall effect of Congruence across all levels of the model, while the individual parameter estimate tests the specific contrast at the reference level of Experiment (Baayen, Davidson, & Bates, 2008; Luke, 2017). Additionally, the presence of interaction terms in the model affects the estimation of standard errors and degrees of freedom for individual contrasts, even when the interactions themselves are not significant (Bates et al., 2015). Following established practice in mixed-effects modeling, we interpret the Type III F-test as the primary test of our hypothesis regarding spatial interference.

In addition to the linear mixed-effects models reported in the main text and above, we also conducted several supplemental analyses that were intended to either facilitate comparison to prior tests of spatial interference that used alternative analyses (see analyses i and iv below), or investigate the robustness of the spatial interference effect (see analyses ii, iii, and v below). Unless otherwise stated, all analyses were conducted in IBM SPSS Statistics 25.

(i) Most prior studies of the spatial interference effect analyzed the data via the simpler approach of *t*-tests and *ANOVA*s. We therefore also analyzed the present data via that approach, and the effect of Congruence was significant in both Experiment 1, paired $t(64) = 2.28$, $p = .026$, $\eta p2 = .075$, and Experiment 2, $t(69) = 2.06$, $p = .043$, $\eta p2 = .058$. Cue words slowed identification of targets in their associated location (i.e., spatial interference). When combining the data from both experiments and analyzing RTs via 2 (Experiment: 1, 2; between-participants) × 2 (Congruence: congruent, incongruent; within-participants) mixed ANOVA, the main effect of Congruence (i.e., spatial interference) was significant, $F(1, 133)$

= 9.45, *p* = .003, *ηp*2 =.066. The main effect of Experiment was not significant, *p* = .862, and the interaction also failed to approach significance, *p* = .422.

The preceding analyses simplify interpretation by combining two different congruent conditions (i.e., high cue with top target and low cue with bottom target) and combining two different incongruent conditions (i.e., high cue with bottom target and low cue with top target).

The spatial interference effect can alternatively be tested via a 2 (Cue Association: high, low) × 2 (Target Location: top, bottom) repeated measures ANOVA, where spatial interference is manifest as the interaction. In fact, this was our preregistered hypothesis in Experiment 2. As predicted, this interaction was significant in both Experiment 1, *F*(1, 64) = 5.05, *p* = .028, *ηp*2 =.073, and Experiment 2, *F*(1, 69) = 4.22, *p* = .044, *ηp*2 =.058.

(ii) In Experiment 1, all experimental (regular) and filler (irregular) cues had spatial associations. When testing for spatial interference with both experimental and filler cues included in the analysis, the effect of Congruence again remained significant, paired *t*(64) = 2.43, *p* = .018, *ηp*2 =.084. Target identification was slower at the cue word's associated location (*M* = 585, *SD* = 166) than at the opposite location (*M* = 574, *SD* = 147), indicating spatial interference.

(iii) Petrova et al. (2018) suggested that the counterbalancing of stimuli across different experimental lists could inadvertently induce a spatial interference effect. To test this possibility, we conducted a mixed ANOVA with Congruence (congruent, incongruent) within-participants and List (1, 2, 3, 4, 5, 6, 7, 8) between-participants. List did not interact with Congruence in either Experiment 1 (*p* = .329) or Experiment 2 (*p* = .253). Thus, the magnitude of the spatial interference effect was similar across the eight experimental lists, and the overall spatial interference effect was not attributable to the counterbalancing of stimuli across lists.

(iv) In addition to the *t*-test and *ANOVA* approach reported above, Petrova et al. (2018) also reported Bayesian analyses. We therefore also conducted a Bayesian hypothesis test using Dienes' (2008) online calculator.[1] The alternative hypothesis was represented as a normal distribution with the effect size reported by Estes et al. (2008, Experiment 3) as the mean of the distribution, and a default of mean/2 as the *SD*. With the data of Experiments 1 and 2 combined, the Bayes factor (*B*) was 15.52, indicating that the observed data are 15 times more likely under the alternative than under the null hypothesis. Thus, there is substantial evidence of spatial interference in these experiments.

(v) Like most RT effects, the spatial interference effect is likely to be sensitive to outliers. In prior studies of spatial interference, some researchers have removed outliers on the basis of each participant's overall mean RT (e.g., Gozli et al., 2013), just as was done in the analyses reported above. And indeed, by that method, spatial interference was robust across both experiments when analyzed separately and when combined. However, other researchers have removed outliers on the basis of each participant's mean RT within each of the four conditions of the 2 (cue association) x 2 (target location) design (e.g., Estes et al., 2015). Because this method includes only 25% of the trials in each outlier calculation (i.e., 25% in each of the four conditions), any extreme values largely distort the variance, thereby leading to fewer RTs being identified as outliers. Indeed, whereas the analyses reported above (i.e., based on the overall mean) excluded 2.84% and 2.99% of trials as outliers in Experiments 1 and 2 respectively, identifying outliers on the basis of the condition mean led to the exclusion of only 1.94% and 2.11% of trials as outliers. Thus, this method of outlier removal retains more moderate outliers in the analyses, thereby increasing the variance. Consequently, the spatial interference effect was not significant in either experiment when

1 http://www.lifesci.sussex.ac.uk/home/Zoltan_Dienes/inference/bayes_factor.swf

analyzed separately (Experiment 1: $p = .147$; Experiment 2: $p = .256$), though it was marginal in the combined analysis, $p = .063$. Thus, although spatial interference was robust across a large number of analyses reported above, a cautionary note is also in order: The effect is highly sensitive to outliers, and hence to methods of outlier removal.