


REPLICATION STUDY 

Second language knowledge can influence native language performance in exclusively native contexts

An approximate replication of Van Hell & Dijkstra (2002)

Eric Pelzl^{1,2} , Rafał Jończyk³ and Janet G. van Hell¹

¹The Pennsylvania State University, University Park, PA, USA; ²The Hong Kong Polytechnic University, Hong Kong, Hong Kong SAR, China and ³Adam Mickiewicz University, Poznań, Poland

Corresponding author: Eric Pelzl; Email: eric.pelzl@polyu.edu.hk

(Received 01 August 2023; Revised 30 April 2024; Accepted 30 May 2024)

Abstract

Over the past decades, bilingualism researchers have come to a consensus around a fairly strong view of nonselectivity in bilingual speakers, often citing Van Hell and Dijkstra (2002) as a critical piece of support for this position. Given the study's continuing relevance to bilingualism and its strong test of the influence of a bilingual's second language on their first language, we conducted an approximate replication of the lexical decision experiments in the original study (Experiments 2 and 3) using the same tasks and—to the extent possible—the same stimuli. Unlike the original study, our replication was conducted online with Dutch–English bilinguals (rather than in a lab with Dutch–English–French trilinguals). Despite these differences, results overall closely replicated the pattern of cognate facilitation effects observed in the original study. We discuss the replication of outcomes and possible interpretations of subtle differences in outcomes and make recommendations for future extensions of this line of research.

Keywords: bilingual; cognate facilitation; Dutch; English; replication

Introduction

Over the past decades, numerous studies on bilingual language processing have addressed the central question of how bilinguals access and activate words in each of their languages and whether this lexical activation is selective or nonselective with respect to language. According to the language–selective view, bilinguals exclusively activate word candidates in the language that corresponds to the language of the incoming information (in comprehension; e.g., Gerard & Scarborough, 1989) or with the language currently

© The Author(s), 2024. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike licence (<http://creativecommons.org/licenses/by-nc-sa/4.0>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the same Creative Commons licence is used to distribute the re-used or adapted article and the original article is properly cited. The written permission of Cambridge University Press must be obtained prior to any commercial use.

in use (in production; e.g., Costa & Caramazza, 1999; cf. Costa, La Hey, & Navarrete, 2006). In contrast, the language–nonselective view holds that words from both languages are activated, and linguistic input in one language induces the coactivation of both languages (e.g., Marian & Spivey, 2003; Thierry & Wu, 2007). Over time, bilingualism researchers have come to the consensus that bilingual lexical activation is fundamentally nonselective with respect to language, even when the social or linguistic context calls for only one language (for a review, see Van Hell & Tanner, 2012). Evidence for parallel coactivation of two languages has been found when bilinguals process words in their second language (L2): L2 words coactivate words in the first and often most dominant language (L1). It has also been found during L1 processing that L1 words coactivate words in the second, less dominant language. The bilingual memory system is permeable when the bilinguals' L1 and L2 systems share the same script, but also for bilinguals whose language systems have different scripts (e.g., Hoshino & Kroll, 2008; Poarch & Van Hell, 2014; Thierry & Wu, 2007), different gesture systems (e.g., Brown & Gullberg, 2008), or are from different modalities as in sign–speech bilinguals (e.g., Lee, Meade, Midgley, Holcomb, & Emmorey, 2019; Ormel, Giezen, & Van Hell, 2022).

In the literature on bilingual lexical activation, the work by Van Hell and Dijkstra (2002) is often cited as critical evidence for the language–nonselective activation view. In this paper, we report a replication study of Van Hell and Dijkstra's (2002) article "Foreign language knowledge can influence native language performance in exclusively native contexts" published in *Psychonomic Bulletin & Review*.

The original study: Van Hell and Dijkstra (2002)

We will focus on Van Hell and Dijkstra's (henceforth, VHD2002) second experiment, and to a lesser degree on the third experiment¹. In these experiments, trilingual speakers (L1 Dutch, L2 English, and L3 [third language] French) performed a lexical decision task (LDT) exclusively in their native and dominant L1 Dutch. The critical stimulus materials included L1 Dutch words that were either cognates with L2 English (e.g., Dutch: *bakker*; English: *baker*; French: *boulangier*), cognates with L3 French (e.g., Dutch: *meubel*; French: *meuble*; English: *piece of furniture*), or noncognate control words (e.g., Dutch: *tuin*; English: *garden*; French: *jardin*). In Experiment 2, trilinguals who were more proficient in English than in French recognized L1 Dutch words that were cognates with L2 English faster than noncognate controls; however, no such *cognate facilitation effect* was observed for the Dutch–French cognates. In Experiment 3, trilinguals with higher proficiency in French (similar to their proficiency in English) performed an LDT on the same stimulus materials demonstrating a cognate facilitation effect for Dutch–English cognates and Dutch–French cognates. These findings strongly support the language–nonselective view, namely, that lexical access and activation are nonselective with respect to a bilingual's languages. The finding that weaker L2 or L3 knowledge can influence L1 processing was interpreted as strong support for the fundamental permeability of language systems in bilingual (or multilingual) speakers.

¹The first experiment reported in VHD2002 asked L1 Dutch–L2 English–L3 French trilinguals to conduct a word association task on a series of L1 Dutch words that were either cognates with L2 English, cognates with L3 French, or noncognates. A cognate facilitation effect was observed for cognates with English (relative to noncognates with English), but not for cognates with French (French was their weakest language). This result pattern parallels the VHD2002 Experiment 2 lexical decision task (discussed in more detail) and also replicates the L1 cognate-facilitation effect in word association reported by Van Hell and De Groot (1998).

Why replicate VHD2002?

In addition to the general motivations for an increase in replications of second language/applied linguistics research highlighted by Marsden, Morgan-Short, Thompson, and Abugaber (2018), Porte and McManus (2018), and Marsden and Morgan-Short (2023), as well as ongoing discussions regarding replication failures (e.g., Nieuwland et al., 2018), there are several specific reasons to pursue replication of this study. First, among the many studies reporting bilingual cognate effects, VHD2002 stands out for providing a particularly strong test of language nonselectivity given its focus on effects *from the L2 to the L1*. Proficiency tests administered after the critical LDT demonstrated that L1 Dutch was the trilinguals' dominant and most proficient language. In other words, processing words in the dominant language coactivated words in the less dominant L2 and L3, but only when proficiency in the L2 and L3 was sufficiently high (as only trilinguals with higher L3 proficiency in Experiment 3, and not those with lower L3 proficiency in Experiment 2, demonstrated the coactivation of their L3). VHD2002's findings served as an empirical foundation for influential theoretical models of lexical processing in the bilingual mental lexicon (e.g., the Bilingual Interactive Activation Plus [BIA+] model: Dijkstra & van Heuven, 2002; the *Multilink* model: Dijkstra et al., 2019; and the Bilingual Language Interaction Network for Comprehension of Speech [BLINCS] model: Shook & Marian, 2013), as well as for critiques of models that did not incorporate cross-language interaction and the influence of L2 knowledge on L1 processing (e.g., Brysbaert & Duyck's [2010] critique of the Revised Hierarchical Model (Kroll & Stewart, 1994; and multiple commentaries on the recently proposed Ontogenesis model [Bordag, Gor, & Opitz, 2021]). Despite the general acknowledgment of language nonselectivity, debate continues about the centrality of language activation for theorizing about linguistic processes in L_n (any language beyond the first language) acquisition and processing. For this reason, efforts to understand the replicability and generalizability of cross-linguistic influence in bilinguals continue to be a pressing issue, and the strong test (from L2 to L1) originally presented in VHD2002 is particularly consequential.

VHD2002 is among the most influential studies in the bilingual word processing literature, and is a highly cited study in second language and bilingual processing research; at the time of writing, VHD2002 has 367 Web of Science citations and 862 Google Scholar citations (both accessed 25 April 2024), with new citations continuing to the present year (e.g., Wu, Van Heuven, Schiller, & Chen, 2024).

Type of replication

Following the typologies proposed by Marsden et al. (2018) and Porte and McManus (2018), we consider the current study an *approximate* replication of VHD2002, hewing close to the original study, but with some significant changes. First, the target population was Dutch–English bilinguals rather than Dutch–English–French trilinguals, and second, the study was conducted online rather than in a lab. Beyond these two substantive changes, we tried to stick as closely to the original study and procedures as possible, though some additional minor changes will be noted below.

We had two motivations for changing the target population. First, there is the theoretical motivation of understanding bilinguals (rather than trilinguals) to test whether the pattern of results observed in VHD2002 is attained when there is no L3. If results are comparable to those of VHD2002, this would support generalization to similar bilinguals. Second, there was the practical reality that recruiting Dutch–

English–French trilingual participants is challenging; VHD2002 had recruited their Experiment 2 trilinguals from a large participant pool of students enrolled at the University of Amsterdam who had volunteered to share in which subjects they had taken final exams in secondary school. Based on this information, a targeted invitation letter to participate in an experiment was sent only to those students who had taken English and French languages throughout secondary school and at their final exams, without revealing the multilingual nature of the experiment they were asked to participate in. However, this database and recruitment procedure were no longer available.

We also had several motivations for moving the study online, rather than trying to conduct it in a lab. As has often been noted, studies with university students (the original participants of VHD2002) may not be generalizable to the broader population. Additionally, the original samples recruited in VHD2002 were relatively small by modern standards (Exp. 2: $n = 19$, Exp. 3: $n = 24$). By using a web-based platform, we were able to reach a more diverse and much larger group of participants. Assuming this approach proves effective, web-based methods will also facilitate future related studies by allowing researchers to target bilinguals across the globe without being tethered to specific laboratory spaces, and they will make our experimental materials easy to share with other researchers.

Design of replication study

VHD2002 highlighted three factors that can impact bilingual language activation: (1) task demands and stimuli, (2) participant's expectations about a study's linguistic context, and (3) relative language fluency. These factors guided the original design and we endeavored to follow the same design principles in the present replication.

With respect to (1) stimuli and task, the original word stimuli were available in the appendix of the VHD2002. The nonwords of the LDT (not reported in VHD2002) were not listed in the appendix, but have now been provided by Van Hell.

With respect to (2) participant's expectations about the linguistic context of the study, we conducted targeted recruitment of participants from Prolific (additional details below). To do this, we used the already available filters on Prolific so that our study was only available to people who spoke Dutch as their L1 and English and/or French as L2s. There was no mention of bilingual status as a requirement for participation. A postexperiment questionnaire allowed us to confirm participant's knowledge of Dutch, English, and French.

With respect to (3) relative language fluency, after the critical experiment, we administered the same proficiency tests used in VHD2002 to measure relative language proficiency.

The complete list of changes made in the present replication is summarized in Table 1. Additional details can be found in the "Procedures" section.

Evaluating evidence of replication

We aim to test whether the (1) *direction* and (2) *significance* of effects replicate from VHD2002 (Exp. 2) to our present study. Evidence of replication would be that (1) bilingual Dutch–English participants were faster at recognizing Dutch–English cognates than either noncognates or Dutch–French cognates, and (2) the difference between Dutch–English cognates and noncognates, and between Dutch–English and

Table 1. Summary of all changes from VHD2002 (Experiment 2)

Variable	Original study	Replication study
Participants	Dutch–English–French trilinguals attending a university in The Netherlands	Dutch–English bilinguals based in The Netherlands who used Prolific
Recruitment	Targeted recruitment based on university records	Targeted recruitment based on Prolific filters
Experiment location	laboratory	online platform
Materials	real words and pseudowords for main experiment; real words and pseudowords for Dutch, English, and French proficiency tests	utilized real–word materials for main experiment reported in original study, recovered additional materials provided by Van Hell
Instructions	original instructions delivered in-person and onscreen	constructed to clearly guide participants online (i.e., without experimenter assistance)
Debriefing questions	not included	presented after the LDT
Statistical analysis	repeated measures ANOVAs	repeated measures ANOVAs and mixed–effects regression

Dutch–French cognates, are statistically significant (i.e., there is a cognate–facilitation effect for English, but not for French). Given that we recruited Dutch–English bilinguals, rather than Dutch–English–French trilinguals, we expect that our participants will be less proficient in French than in English (their L2), and most proficient in Dutch (their L1). Because of their limited French knowledge, we might also find our participants display stronger differences between Dutch–English and Dutch–French cognates than observed in VHD2002 (Exp. 2). Altogether, such results would indicate a replication of the cognate facilitation effect observed in VHD2002 (Exp. 2) and demonstrate that it occurs for Dutch–English bilinguals just as it did for Dutch–English–French trilinguals.

Impact

Replication (or failure of replication) will provide additional evidence for the ongoing discussion of the coactivation of a bilingual’s languages during word recognition tasks, and the nature of cognates in bilingual lexical representations. By testing the cognate facilitation effect in Dutch–English bilinguals, and by using a web–based platform to collect data online rather than in the lab, we will provide new evidence regarding the generalizability of previously observed effects.

Methods

Participants

We recruited 96 participants online using Prolific (50 male, 45 female, 1 nonbinary). Table 2 reports participant age and details of language experience (including self–rated proficiency) for English and French. We followed the “rule of thumb” recommendation from Brysbaert and Stevens (2018) to have at least 1600 observations per condition (20 items × 96 participants = 1920 observations per condition). While we did not conduct an a priori power analysis, this sample is considerably larger than that of VHD2002 (Exp. 2), which had 19 participants. Retrospective power sensitivity analyses suggest this sample was sufficient to reliably detect differences of roughly 20 ms or more between conditions (see Appendix B of supplementary materials for details).

Table 2. Age and language background information of participants ($n = 96$). Age of acquisition and self-rated proficiency for four language skills is reported for English and French (scale of 1-10)

	Mean (SD)	Range
Age (years)	28.4 (8.0)	18–53
English		
Age of acquisition (years)	8.75 (2.8)	2–15
Listening	8.9 (0.9)	7–10
Speaking	7.8 (1.1)	5–10
Reading	8.9 (1.0)	6–10
Writing	7.8 (1.3)	3–10
French		
Age of acquisition* (years)	13.4 (6.4)	8–42
Listening	2.9 (1.7)	1–7
Speaking	2.0 (1.4)	1–7
Reading	2.9 (1.7)	1–7
Writing	1.8 (1.4)	1–7

*age of acquisition only for the 23 participants who indicated having knowledge of French.

As described earlier, VHD2002 had a unique recruitment situation that allowed them to conduct targeted recruitment of trilingual participants (English L2, French L3) *without revealing the aims of the study*. We no longer had access to that original recruitment context. Given the difficulty of recruiting trilingual participants without asking if they were trilingual, we took a more modest approach that targeted Dutch–English bilinguals using preexisting screening filters available on the participant recruitment platform Prolific (www.prolific.co) [13 May 2023].

Eligibility was determined in a two-step process. First, the study was only available to people who identified their (1) nationality and (2) current place of residence as The Netherlands, their (3) primary language as Dutch, (4) identified as bilingual, and (5) identified their fluent language(s) as English and/or French. Note that the fluent language filter did not allow us to require *both* English and French (this would have required additional new screening and might have revealed the nature of the study). In this way, Prolific’s default settings served as a first step in determining eligibility. As a second step, we also used postexperiment survey questions to verify these requirements. Participants who answered questions in ways inconsistent with the first round of screening or indicated other linguistic situations inconsistent with the aims of the study (e.g., identified both Dutch and English as L1) were removed from further analysis and replaced (3 out of a total of 99 participants).

This recruitment approach allowed us to disguise the aims of the study as participants were never told *why* they were eligible, nor given any indication that the study would involve knowledge of English and French. On Prolific, the study had a Dutch title: “Geschreven taalverwerking in het Nederlands” (*Written Language Processing in Dutch*); and the description of the study was provided only in Dutch.

All procedures were approved by IRB. Participants provided informed consent and were compensated the equivalent of 5 USD for their participation.

Stimuli

Lexical decision task

As in VHD2002, the lexical decision task comprised 140 stimuli: 80 critical real Dutch words, and 60 pseudowords. Of the Dutch words, 20 were Dutch–English cognates

(Dutch: *bakker*; English: *baker*; French: *boulangier*), 20 were Dutch–French cognates (Dutch: *meubel*; French: *meuble*; English: *piece of furniture*), and the other 40 were Dutch noncognates (Dutch: *tuin*; English: *garden*; French: *jardin*).

As reported in VHD2002, all real words were words for concrete concepts and were controlled for length in letters (Dutch–English cognates: $M = 5.5$, $SD = 1.1$; Dutch–French cognates: $M = 5.4$, $SD = 1.0$; Dutch noncognates: $M = 5.4$, $SD = 0.7$), log word frequencies (occurrences per million based on the CELEX printed–lemma frequency counts, Baayen, Piepenbrock, & Van Rijn, 1993; Dutch–English cognates: $M = 1.17$, $SD = 0.53$; Dutch–French cognates: $M = 1.12$, $SD = 0.60$; Dutch noncognates: $M = 1.38$, $SD = 0.40$), and for the number of orthographic neighbors in Dutch (Dutch–English cognates: $M = 2.6$, $SD = 3.8$; Dutch–French cognates: $M = 2.4$, $SD = 2.2$; Dutch noncognates: $M = 2.6$, $SD = 2.4$). Pseudowords were created by changing one letter of a real Dutch word.

The critical cognates were included in an [appendix](#) of VHD2002 and were used exactly as listed. The pseudowords, however, were not included in that [appendix](#). We were able to recover a list of pseudowords from the original study documents used in VHD2002. Four pseudowords in those files occurred with two variants (jussen vs. jangen, schatel vs. inwoker, vonger vs. hoolte, doffie vs. oorlod). Of these, we selected jussen, schatel, vonger, and doffie.

An additional 35 stimuli (20 words, 15 pseudowords), also recovered from the study documents, were used as practice items.

Proficiency tests

Along with the critical lexical decision stimuli, VHD2002 also included three proficiency tests—all in lexical decision format—one each for Dutch, English, and French. Each test comprised 50 critical real words and 40 pseudowords, along with 18 practice items (10 real words, 8 pseudowords). The stimuli for these tests were not included in the [appendix](#) of VHD2002. We were able to recover Dutch, English, and French lexical decision stimuli which we believe are the same as those used in VHD2002.² All materials used in the present study are available in [Appendix A](#) of the supplementary materials.

Procedures

As described above, participants were recruited to the study using Prolific. Instructions and procedures were delivered entirely via computer, without any experimental staff present. For this reason, we took care to make instructions especially clear and easy to follow (a full transcript of instructions is available in [Appendix E](#) of the supplementary materials). As noted in [Table 1](#), this differed from VHD2002, where the experimenter was present to deliver instructions in person, as well as on the computer screen. After enrolling in the study, participants received a link directing them to the study hosted on the experimental platform Labvanced (Finger, Goeke, Diekamp, Standvoß, & König, 2017). On Labvanced the entire study was conducted in Dutch, with the exception of default Labvanced messages that appeared in English

²Each set of practice items, presented before each critical proficiency LDT test, had one fewer pseudoword (7 items) and one more real word (11 items) than reported in VHD2002 (8 and 10 respectively). To match the numbers of practice items reported in VHD2002, we changed one letter of a real word in each list to create the correct number of pseudowords.

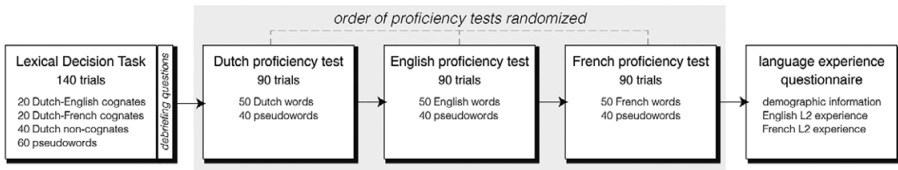


Figure 1. Order of tasks and summary of stimuli.

when the website loaded.³ Participants first saw a consent form in Dutch, and after consenting proceeded directly into the experiment. They were told they would complete a series of vocabulary tests and a brief survey. The procedure for the lexical decision task was introduced and, after completing the practice trials, the critical LDT began. The order of trials was pseudo-randomized uniquely for each participant, with no more than three consecutive trials of the same condition allowed. (The order of experimental tasks is illustrated in Figure 1.)

Upon completion of the LDT, participants answered two debriefing questions. The questions were (in Dutch): “During the first experiment, did you notice yourself thinking in English?” and the same question was also provided in French. Although this was not part of the original protocol, we believed it was a desirable and simple change that might provide additional evidence as to whether participants were indeed operating in a solely Dutch mode or not. As the questions came after the critical experiment, they could not affect the key experimental results.

Participants were next instructed that they would complete three vocabulary tests, one each for Dutch, English, and French. They were instructed that if they did not know English or French they should simply do their best. They then proceeded to complete the practice items and critical trials for the three proficiency tests. The order of the three tests was randomized across participants, and the order of trials within each test was pseudo-randomized for each participant, with no more than three consecutive words/pseudoword trials allowed.

After completing the proficiency tests, participants completed a short language experience questionnaire providing basic demographic information and answering targeted questions about their previous experience with English and French. All participants indicated whether they knew English and/or French, the age at which they began learning each language, and provided self-ratings (1 = none, 10 = perfect) for their listening, speaking, reading, and writing skills in English and French.

After completing all these tasks, they returned to Prolific to indicate they had completed the study. Median completion time was about 20 min.

Data processing and statistical analysis

Lexical decision task

All LDT data were processed and analyzed in *R* (version 4.0.3; R Core Team, 2022). Trials with missing values (nonresponses) were removed (0.7% of all real word data:

³The Labvanced messages were as follows: “Loading Complete! You can now start the experiment. This will switch your browser into fullscreen mode. Please note that during the experiment you should not press escape or use the ‘backward’ button in your browser.” This message was accompanied by a “Start” button. After the participant clicked “Start,” the experiment went into fullscreen mode and briefly displayed the message “Starting Experiment.” From then on all instructions and other messages were provided only in Dutch.

0.7% of Dutch–English cognates, 0.8% of Dutch–French cognates, and 0.7% of non-cognates). The accuracy of responses was scored “correct” (1) or “incorrect” (0). Trials with incorrect responses were removed prior to RT analyses (4.9% of all real word trials: 1.6% of Dutch–English cognates, 4.9% of Dutch–French cognates, and 6.5% of non-cognates). Following VHD2002, we computed a mean RT and standard deviation for each participant for each condition removing any responses that were faster than 100 ms or 2.5 standard deviations above a participant’s mean (3.2% of all real word data: 3.6% of Dutch–English cognates, 3.3% of Dutch–French cognates, and 3.0% of noncognates). Finally, we computed mean RTs and standard deviations for the trimmed data.

VHD2002 used repeated measures analyses of variance (RM ANOVAs) to analyze both the LDT and proficiency tests. For comparison, we also report RM ANOVA results here, however, we also deployed mixed-effects logistic regression models given their ability to simultaneously model both participants and items (Baayen et al., 2008).

The processed LDT data were first submitted to RM ANOVA (Type III) using the *afex* package (Singmann, Bolker, Westfall, Aust, & Ben-Shachar, 2022) in *R*. RTs were log-transformed to better meet the assumption of normality for our statistical models.⁴ Huynh–Feldt corrections were applied for violations of sphericity.

We fitted linear mixed-effects regression models using the *lme4* package (version 1.1-30; Bates, Mächler, Bolker, & Walker, 2015). The dependent variable was log RT, and the fixed effect of *cognate status* (Dutch–English, Dutch–French, noncognate) was sum-coded (1, 0, -1; 0, 1, -1). Random intercepts and slopes were included for the effects of participants and items. A maximal random effects model was fit first and we retained the maximal model that converged without warnings (Barr et al., 2013; Bates et al., 2015). Models including random slopes for participants did not converge smoothly (singular fit warnings) and were not selected. The final model included random intercepts for subject and items: *lmer model formula*: $\log(\text{RT}) \sim \text{cognate status} + (1 \mid \text{participant}) + (1 \mid \text{item})$. Post hoc comparisons were performed using the *emmeans* (Lenth, 2022) and *multcomp* (Hothorn et al., 2008) packages in *R*.

Proficiency tests in Dutch, English, and French

Trials from the proficiency tests with missing values were removed (1.3% of Dutch real word trials, 2.1% of English real word trials, 2.0% of French real word trials). The accuracy of responses was scored “correct” (1) or “incorrect” (0). Trials with incorrect responses were removed prior to reaction time (RT) analyses (2.78% of Dutch words, 3.1% of English words, 31.4% of French words). Following VHD2002, we computed mean RT and standard deviation values for each participant. We removed any responses faster than 100 ms or 2.5 standard deviations above a participant’s mean (3.3% of Dutch data, 3.2% of English data, 3.6% of French data). Finally, we computed mean RTs and standard deviations for the trimmed data.

The processed proficiency data were first submitted to repeated measures ANOVA (Type III) using the *afex* package (Singmann et al., 2022) in *R*. RTs were log-transformed to better meet the assumption of normality for our statistical models.⁵ Huynh–Feldt corrections were applied for violations of sphericity.

⁴We also ran models with untransformed RTs; there was no substantive difference in outcomes between models for raw and log-transformed RTs.

⁵We again ran models with untransformed RTs; there was no substantive difference in outcomes between models for raw and log-transformed RTs.

We again fitted linear mixed-effects regression models using the *lme4* package (version 1.1-30, Bates et al., 2015). The dependent variable was log RT, and the fixed effect of *language* (Dutch, English, French) was sum-coded (1, -1). Random intercepts and slopes were included for the effects of participants and items. A maximal random effects model was fit first and we retained the maximal model that converged without warnings (Barr et al., 2013; Bates et al., 2015). The final model included random slopes and intercepts for the subject and random intercepts for items: *lmer model formula*: $\log(\text{RT}) \sim \text{language} + (1 + \text{language} \mid \text{participant}) + (1 \mid \text{item})$. Post hoc comparisons were performed using the *emmeans* (Lenth, 2022) and *multcomp* (Hothorn et al., 2008) packages in R.

Results

Lexical decision task: descriptive results

Descriptive results are summarized in Table 3 with the results of VHD2002 (Exp 2) for comparison. Overall, the pattern of RTs and error rates across conditions mirrors that of the original study, but with average times roughly 90 ms faster in VHD2002 than in the present study.

Lexical decision task: ANOVA

There was a significant effect of cognate status on both by-participants ($F(1.91, 181.03)$ $\eta_p^2 = .535$, $p < .001$) and by-items ($F(2, 77) = 8.96$, $\eta_p^2 = .189$, $p < .001$).

Pairwise post hoc comparisons (with Tukey adjustments) for by-participant and by-item results indicated that Dutch-English cognates were recognized significantly faster than noncognates (participants, $p = .013$; items, $p = .001$). The faster RTs to Dutch-French cognates relative to noncognates and Dutch-English cognates reached significance in the by-participants analyses ($p < .001$ in both cases), but not in the by-items analyses ($p = .052$ and $p = .300$, respectively). For confidence intervals, see Appendix C.

These results mirror the direction and statistical significance of by-participant and by-item outcomes reported in VHD2002, with the exception that VHD2002 did not find a significant difference between Dutch-French cognates and noncognates in the by-participant analysis.

Lexical decision task: linear mixed-effects regression models

Model results are reported in Table 4 and indicate log RT effects for Dutch-English and Dutch-French cognates compared with the model intercept (i.e., the grand mean). Model results are depicted visually in Figure 2. Post hoc comparisons (Table 5) between

Table 3. Response times (correct answers only) and error rates for lexical decision task in current replication and original study

Condition	Replication		VHD2002	
	RT (ms)	ER (%)	RT (ms)	ER (%)
	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)
Dutch-English cognates	577 (53.2)	1.62 (2.7)	499 (48)	1.94 (3.0)
Dutch-French cognates	611 (70.3)	4.95 (5.9)	519 (46)	2.50 (3.5)
Noncognates	630 (63.3)	6.48 (6.0)	529 (41)	7.36 (3.7)

Table 4. Lexical decision task: mixed-effect regression model results

	Fixed effects						Random effects	
	Estimate	SE	t	df	p	By subjects	By item	
						SD	SD	
(Intercept)	6.39	0.01	428.00	162.34	<.001	0.12	0.07	
Cog.Status [S.Cog-E]	-0.05	0.01	-3.71	77.47	<.001	–	–	
Cog.Status [S.Cog-F]	0.01	0.01	0.76	77.75	.452	–	–	

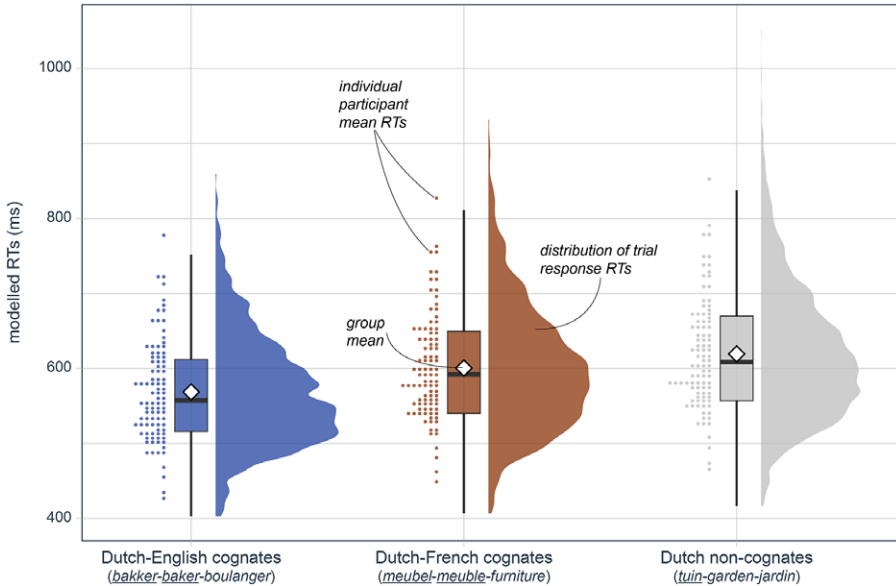


Figure 2. Model-estimated RTs (back-transformed from log RTs) for the lexical decision task. Group mean depicted with white diamonds. Participant means (binned in 5 ms intervals) depicted by shaded circles. Shaded areas to the right depict distribution of responses.

Table 5. Lexical decision task: post hoc comparisons for mixed-effects regression model (log RTs)

Contrast	Estimate	SE	df	95% CI		z	p
				LCL	UCL		
Dutch-English vs. Dutch-French	-0.06	0.02	81.7	-0.12	-0.001	-2.43	.045
Dutch-English vs. noncognates	-0.09	0.02	81.6	-0.14	-0.04	-4.18	<.001
Dutch-French vs. noncognates	-0.03	0.02	82.0	-0.08	0.02	-1.38	.358

*95% CIs estimated with *pbkrtest* package (Halekoh & Højsgaard, 2014).

log RTs of each cognate status condition indicated that responses to Dutch-English cognate words were significantly faster than to noncognates. The speed of responses to Dutch-French cognates did not differ significantly from either Dutch-English cognates or noncognates.

Table 6. Response times (correct answers only) and error rates for three proficiency tests in current replication and original study

Language	Replication		VHD2002	
	RT (ms)	ER (%)	RT (ms)	ER (%)
	Mean (SD)	Mean (SD)	Mean (SD)	Mean
Dutch	583 (97)	2.78 (254)	471 (31)	3.78
English	578 (79)	3.08 (3.7)	550 (43)	8.89
French	746 (175)	31.40 (12.7)	611 (64)	23.33

These results mirror the direction and statistical significance of cognate priming effects reported in VHD2002, but remove the ambiguity produced by inconsistencies between by-item and by-participant repeated-measures ANOVA results.

Proficiency tests: descriptive results

We now consider the results of the three proficiency tests that used lexical decision tasks to test vocabulary knowledge in Dutch, English, and French. Descriptive results are summarized in Table 6 with a comparison with the results of VHD2002.

Proficiency tests: ANOVA

There was a significant effect of language on both by-participants ($F(1.17, 111.13) = 47.00, \eta_p^2 = .331, p < .001$) and by-items ($F(2, 147) = 212.77, \eta_p^2 = .743, p < .001$).

Pairwise post hoc comparisons (with Tukey adjustments) for by-participant and by-item results indicated significantly faster RTs for Dutch compared with French (both participants and items, $p < .001$), and English compared with French (both participants and items, $p < .001$) (for confidence intervals, see Appendix C); this pattern is similar to that found by VHD2002. There was no significant difference in RTs for Dutch compared with English (participants, $p = .999$; items, $p = .989$); this differs from VHD2002 who found significantly faster RTs for Dutch compared with English.

Proficiency tests: linear mixed-effects regression models

Model results are reported in Table 7 and indicate the log RT effect for English and Dutch compared with the model intercept (grand mean). Model results are depicted visually in Figure 3. Post hoc comparisons (Table 8) between log RTs for each pair of

Table 7. Model results for proficiency tests (log RTs)

	Fixed effects					Random effects	
	Estimate	SE	t	df	p	By subjects	By item
						SD	SD
(Intercept)	6.41	0.01	432.75	116.39	<.001	0.14	0.06
Language [S.Dutch]	-0.07	0.01	-5.54	150.05	<.001	0.11	-
Language [S.English]	-0.07	0.01	-5.97	164.53	<.001	0.10	-

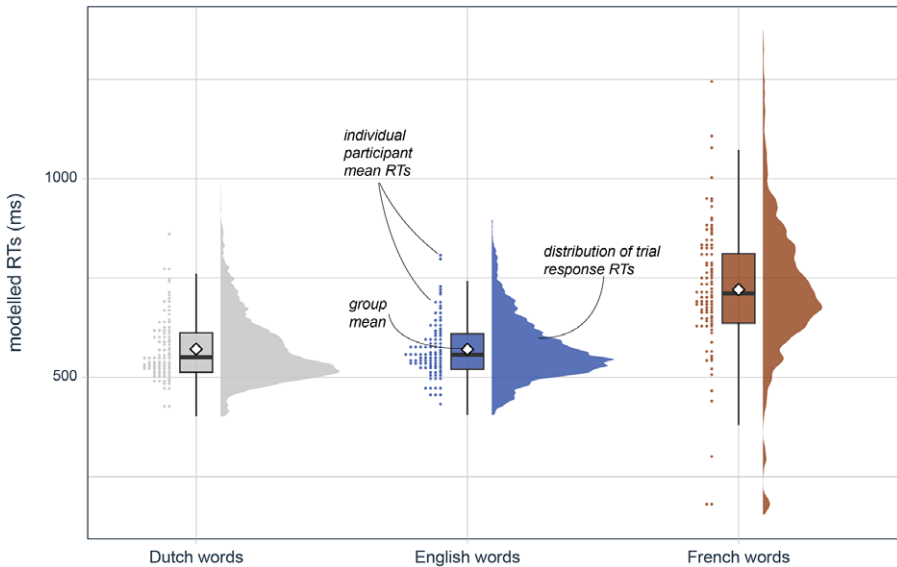


Figure 3. Model-estimated RTs (back-transformed from log RTs) for the three proficiency tests. Group mean depicted with white diamonds. Participant means (binned in 5 ms intervals) depicted by shaded circles. Shaded areas to the right depict distribution of responses.

Table 8. Post hoc comparisons for mixed-effects regression model of proficiency test results (log RTs)

Contrast	Ratio	SE	df	95% CI*		t	p
				LCL	UCL		
Dutch vs. English	-0.002	0.02	207	-0.04	0.03	-0.16	.987
Dutch vs. French	-0.222	0.03	123	-0.30	-0.14	-6.73	<.001
English vs. French	-0.220	0.03	126	-0.29	-0.15	-7.02	<.001

*95% CIs estimated with pbkrtest package (Halekoh & Højsgaard, 2014).

languages indicated that responses to both Dutch and English words were significantly faster than to French words, but that the speed of responses to Dutch and English words did not differ significantly.

This pattern of results partially parallels the proficiency test results of VHD2002 in that we observed that RTs to French words were significantly slower than to Dutch or English words. However, results differ from those of VHD2002 in that we did not observe significantly faster RTs to Dutch than to English words.

To explore the possible role of French proficiency, we conducted additional exploratory analyses (reported in Appendix D in the supplementary materials). None of these analyses indicated significant differences based on French proficiency.

Debriefing questions

The debriefing question data (in Dutch) were scored manually. Answers that indicated any level of awareness of English or French during the LDT were scored 1 (e.g., “Yes,” “A little,” “For a few words”). Negative responses were scored 0 (e.g.,

“No,” “Not really”). For the question regarding English, 32 participants indicated some level of awareness of thinking of English words during the task. For French, 11 participants indicated some level of awareness. In short, insofar as we can trust such retrospective reports, the resemblance of Dutch words to English or French cognates elicited some amount of awareness of the other language(s) in some Dutch participants, even though—with the exception of the default Labvanced messages that appeared in English when the website loaded—the entire task and all instructions were presented only in Dutch and made no mention of English or French up to that point of the study.

Discussion

We conducted an approximate replication of VHD2002 to test for cognate facilitation effects from Dutch–English bilinguals L2 (English) to their L1 (Dutch). As in the original study, we found facilitation for Dutch–English cognates compared with non-cognates and no statistically significant facilitation for Dutch–French cognates. Additionally, repeated measures ANOVA analyses by-items and using mixed-effects models failed to find differences between Dutch–English and Dutch–French cognates, though the by-person repeated measures ANOVA did find a difference. In short, these results are the same as those of the original study in terms of both the direction of effects and their statistical significance. The single difference is that in the post hoc test for the by-person repeated measures ANOVA (and only there), responses to Dutch–French cognates differed significantly from Dutch noncognates.

A second set of analyses considered RT results of three language proficiency tests for Dutch, English, and French words. Unlike VHD2002, we found no difference in proficiency for Dutch and English among our participants who were equally fast at recognizing words in both languages. They were, however, much slower (and much less accurate) at recognizing French words. These results are consistent with our expectation that, as Dutch–English bilinguals, the participants in the replication would not be as proficient in French as participants in VHD2002. We did not expect that participants would be equally as proficient in English as in Dutch (by the metric of vocabulary tests), though this might not be surprising given increases in the use of English in Dutch society and, especially, on the internet over the past decades.

Role of French proficiency

One significant change from the original study was our recruitment of bilingual rather than trilingual participants. Although some of our participants did claim some knowledge of French, as a group, they appear to have been less proficient than the trilinguals in VHD2002. In the proficiency tests, our participants made about 10% more errors in French than VHD2002’s participants, and the difference in RTs between French words and Dutch/English words was more than twice as large as in the original study (see Table 6 above).

Despite the difference in French proficiency, we replicated the cognate facilitation effect from L2 to L1 found in VHD2002. However, assuming our participants were indeed less proficient in French, there is one puzzle: Why did Dutch–French cognates trend towards a facilitation effect (there was a significant difference in the by-participants ANOVA, although not in the by-items ANOVA and linear mixed-effects regression modeling)?

One plausible reason why the present replication study observed a facilitation effect for French cognates in the by-participants ANOVA (and again, only in this analysis) is that the replication study included considerably more participants ($n = 96$) than the original VHD2002 study ($n = 19$). Numerically, VHD2002 observed a 10 ms facilitation effect for the cognates with French (mean RTs for French cognates and noncognates were 519 ms and 529 ms, respectively; see Table 3), but this numerical difference did not reach significance in both the by-participants and the by-items analyses. Using the same (number of) items, the French cognate facilitation effect in the replication study was 19 ms (mean RTs for French cognates and noncognates were 611 ms and 630 ms, respectively; see Table 3). Even though (still) not significant in the by-items ANOVA (and in linear mixed-effects regression modeling), this 19 ms facilitation effect did reach significance in the by-participants analysis that included 96 participants. The difficulty of interpreting inconsistencies between by-participant and by-item analyses in repeated measures ANOVAs is one reason to prefer mixed-effects regression that yields a single outcome that simultaneously models by-participant and by-item effects (Baayen, Davidson, & Bates, 2008).

Online vs. lab-based methods

Our results provide support for the utility of online methods for conducting lexical decision tasks examining cognate facilitation effects. Average RTs in the present study were roughly 90 ms slower and RT differences between conditions were about twice the size of those in the original study (difference between Dutch-English cognates and noncognates: replication = 53 ms, VHD2002 = 30 ms; between noncognates and Dutch-French cognates: replication = 19 ms, VHD2002 = 10 ms). At the same time, we recruited a substantially larger group of participants who were more diverse in terms of age and educational background. A larger sample like this may be necessary to get reliable effects using web-based methods (though, in the present case, the study was likely overpowered; see power sensitivity analysis in Appendix B of the supplementary materials). At the same time, a more diverse sample (at least more diverse than undergraduate populations often tested in psycholinguistic and neurolinguistic studies) will yield data that are more generalizable to the wider population of language users.

Future replication research

The rich body of empirical work supporting language nonselectivity (and refuting language selectivity) actually entails a limited set of languages, typically Western languages that share the same script (e.g., Dutch-English, Spanish-English; for exceptions, see, e.g., Allen & Conklin, 2013, Gollan, Forster, & Frost, 1997, Poarch & Van Hell, 2014, Nakayama, Verdonschot, Sears, & Lupker, 2014). This is particularly true for studies of cognate processing in L1. This significantly limits the generalizability of that literature, including the findings of VHD2002 and subsequent studies. Critically, studies including a wider mix of languages (including typologically different and less frequently studied languages) are needed to determine the extent of nonselectivity as a principled mechanism and the factors that moderate or constrain language selectivity.

Replication and reproducibility (Marsden et al., 2018; Porte & McManus, 2018), as well as open science practices (Marsden & Morgan-Short, 2023), have become critically

important in current research practice, including the field of second language learning and bilingualism. Replication studies are an important tool to ascertain the validity and reliability of the empirical basis for theoretical models on language processing, and for warranting societal and clinical implications these findings may have. The present replication study demonstrates that online platforms enable the collection of a solid set of RT data that can be reliably used in the field's replication efforts, without having to invest the extensive time and resources typically associated with in-person laboratory testing.

Conclusion

Replication is one of the key ways that scientists build confidence in the scientific merit of empirical data. In line with recent calls for replicating influential findings in the field of second language learning and bilingualism (e.g., Marsden et al., 2018; Porte & McManus, 2018), our approximate replication study paralleled the findings of Van Hell and Dijkstra (2002) that have served to shape theoretical models of bilingual word processing by showing that bilingual language activation is fundamentally nonselective with respect to language. Using an online platform for data collection rather than individually testing participants in the laboratory, and testing a more diverse participant sample, the present study confirmed that second language knowledge can influence native language performance in exclusively native contexts.

Supplementary material. The supplementary material for this article can be found at <http://doi.org/10.1017/S0272263124000457>.

Data availability statement. The experiment in this article earned Open Data badge for transparent practices. The data are available at <https://osf.io/xb8u3/>

Acknowledgments. The writing of this article was supported by National Science Foundation (NSF) grants SBE SMA-2004279 to Eric Pelzl and Janet van Hell and by DGE-NRT-2125865, BCS-2041264, DUE-IUSE-1726811, ECR-HER 2155079, and OISE-1545900 to Janet van Hell.

Competing interest. The authors declare none.

References

- Allen, D. B., & Conklin, K. (2013). Cross-linguistic similarity and task demands in Japanese-English bilingual processing. *PLoS ONE*, 8(8), e72631. <https://doi.org/10.1371/journal.pone.0072631>
- 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(4), 390–412. <https://doi.org/10.1016/j.jml.2007.12.005>
- Baayen, R. H., Piepenbrock, R., & Van Rijn, H. (1993). *The CELEX Lexical Database* [CD-ROM]. Linguistic Data Consortium, University of Pennsylvania.
- 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. <https://doi.org/10.1016/j.jml.2012.11.001>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Bordag, D., Gor, K., & Opitz, A. (2021). *Ontogenesis model of the L2 lexical representation*. Bilingualism: Language and Cognition, 1–17. <https://doi.org/10.1017/S1366728921000250>
- Brown, A. & Gullberg, M. (2008). Bidirectional cross-linguistic influence of L1-L2 encoding of manner in speech and gesture. *Studies in Second Language Acquisition*, 30, 225–251. <https://doi.org/10.1017/S0272263108080327>

- Brysaert, M., & Duyck, W. (2010). Is it time to leave behind the revised hierarchical model of bilingual language processing after fifteen years of service? *Bilingualism: Language and Cognition*, 13(3), 359–371. <https://doi.org/10.1017/S1366728909990344>
- Brysaert, M., & Stevens, M. (2018). Power analysis and effect size in mixed effects models: A tutorial. *Journal of Cognition*, 1(1):9, 1–20. <https://doi.org/10.5334/joc.10>
- Costa, A., & Caramazza, A. (1999). Is lexical selection in bilingual speech production language-specific? Further evidence from Spanish-English and English-Spanish bilinguals. *Bilingualism: Language and Cognition*, 2(3), 231–244. <https://doi.org/10.1017/S1366728999000334>
- Costa, A., La Hey, W., & Navarrete, E. (2006). The dynamics of bilingual lexical access. *Bilingualism: Language and Cognition*, 9(2), 137–151. <https://doi.org/10.1017/S1366728906002495>
- Dijkstra, T., & van Heuven, W. J. B. (2002). The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and Cognition*, 5(3), 175–197. <https://doi.org/10.1017/S1366728902003012>
- Dijkstra, T., Wahl, A., Buytenhuijs, F., Van Halem, N., Al-Jibouri, Z., De Korte, M., & Rekké, S. (2019). Multilink: A computational model for bilingual word recognition and word translation. *Bilingualism: Language and Cognition*, 22(4), 657–679. <https://doi.org/10.1017/S1366728918000287>
- Finger, H., Goeke, C., Diekamp, D., Standvoß, K., & König, P. (2017). *LabVanced: A unified JavaScript framework for online studies*. International Conference on Computational Social Science, Cologne.
- Gerard, L. D., & Scarborough, D. L. (1989). Language-specific lexical access of homographs by bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15(2), 305–315. <https://doi.org/10.1037/0278-7393.15.2.305>
- Gollan, T. H., Forster, K. I., & Frost, R. (1997). Translation priming with different scripts: Masked priming with cognates and noncognates in Hebrew-English bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23(5), 1122–1139. <https://doi.org/10.1037//0278-7393.23.5.1122>
- Halekoh, U., & Højsgaard, S. (2014). A Kenward-Roger approximation and parametric bootstrap methods for tests in linear mixed models—The R package. *Journal of Statistical Software*, 59(9), 1–20. <https://doi.org/10.18637/jss.v059.i09>
- Hoshino, N., & Kroll, J.F. (2008). Cognate effects in picture naming: Does cross-language activation survive a change of script? *Cognition*, 106(1), 501–511. <https://doi.org/10.1016/j.cognition.2007.02.001>
- Kroll, J. F., & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of Memory and Language*, 33(2), 149–174. <https://doi.org/10.1006/jmla.1994.1008>
- Lee, B., Meade, G., Midgley, K.J., Holcomb, P.J., & Emmorey, K. (2019). ERP evidence for co-activation of English words during recognition of American Sign Language signs. *Brain Sciences*, 9(148), 1–17. <https://doi.org/10.3390/brainsci9060148>
- Lenth, R. (2022). emmeans: Estimated Marginal Means, aka Least-Squares Means. (Version R package version 1.7.4-1) [Computer software]. <https://CRAN.R-project.org/package=emmeans>
- Marian, V., & Spivey, M. (2003). Competing activation in bilingual language processing: Within- and between-language competition. *Bilingualism: Language and Cognition*, 6(2), 97–115. doi:10.1017/S1366728903001068
- Marsden, E., & Morgan-Short, K. (2023). (Why) are open research practices the future for the study of language learning? *Language Learning*, 73(S2), 344–387. <https://doi.org/10.1111/lang.12568>
- Marsden, E., Morgan-Short, K., Thompson, S., & Abugaber, D. (2018). Replication in second language research: Narrative and systematic reviews and recommendations for the field. *Language Learning*, 68(2), 321–391. <https://doi.org/10.1111/lang.12286>
- Nakayama, M., Verdonschot, R. G., Sears, C. R., & Lupker, S. J. (2014). The masked cognate translation priming effect for different-script bilinguals is modulated by the phonological similarity of cognate words: Further support for the phonological account. *Journal of Cognitive Psychology*, 26(7), 714–724. <https://doi.org/10.1080/20445911.2014.953167>
- Nieuwland, M. S., PolitzerpAhles, S., Heyselaar, E., Segaert, K., Darley, E., Kazanina, N. ... Huettig, F. (2018). Large-scale replication study reveals a limit on probabilistic prediction in language comprehension. *eLife*, 7, e33468. <https://doi.org/10.7554/eLife.33468>
- Ormel, E., Giezen, M., & Van Hell, J. G. (2022). Cross-language activation in bimodal bilinguals: Do mouthings affect the coactivation of speech during sign recognition? *Bilingualism: Language and Cognition*, 25(4), 579–587. <https://doi.org/10.1017/S1366728921000845>

- Poarch, G. J., & Van Hell, J. G. (2014). Cross-language activation in same-script and different-script trilinguals. *International Journal of Bilingualism*, 18(6), 693–716. <https://doi.org/10.1177/1367006912472262>
- Porte, G., & McManus, K. (2018). *Doing replication research in applied linguistics*. Routledge.
- R Core Team. (2022). R: A language and environment for statistical computing. [Computer software]. R Foundation for Statistical Computing. <http://www.R-project.org/>
- Shook, A., & Marian, V. (2013). The bilingual language interaction network for comprehension of speech. *Bilingualism: Language and Cognition*, 16(2), 304–324. <https://doi.org/10.1017/S1366728912000466>
- Singmann, H., Bolker, B., Westfall, J., Aust, F., & Ben-Shachar, M. S. (2022). *afex: Analysis of factorial experiments* (R package version 1.1-1). <http://cran.r-project.org/package=afex>
- Thierry, G., & Wu, Y. J. (2007). Brain potentials reveal unconscious translation during foreign-language comprehension. *Proceedings of the National Academy of Sciences*, 104, 12530–12535. <https://doi.org/10.1073/pnas.0609927104>
- Van Hell, J. G., & De Groot, A. M. B. (1998). Conceptual representation in bilingual memory: Effects of concreteness and cognate status in word association. *Bilingualism: Language and Cognition*, 1, 193–211. <https://doi.org/10.1017/S1366728998000352>
- Van Hell, J. G., & Dijkstra, T. (2002). Foreign language knowledge can influence native language performance in exclusively native contexts. *Psychonomic Bulletin & Review*, 9(4), 780–789. <https://doi.org/10.3758/BF03196335>
- Van Hell, J. G., & Tanner, D. (2012). Second language proficiency and cross-language lexical activation. *Language Learning*, 62, 148–171. <https://doi.org/10.1111/j.1467-9922.2012.00710.x>
- Wu, J., Van Heuven, V. J., Schiller, N. O., & Chen, Y. (2024). Recognizing two dialects in one written form: A Stroop study. *Bilingualism: Language and Cognition*, 1–17. <https://doi.org/10.1017/S1366728924000142>

Cite this article: Pelzl, E., Jończyk, R., & van Hell, J. G. (2024). Second language knowledge can influence native language performance in exclusively native contexts: An approximate replication of Van Hell & Dijkstra (2002). *Studies in Second Language Acquisition*, 46: 1457–1474. <https://doi.org/10.1017/S0272263124000457>