



Detecting cognitive interactions through eye movement transitions

Lisa Vandenberg^{1,2,*}, Samantha Bouwmeester¹, Bruno R. Bocanegra², Rolf A. Zwaan

Department of Psychology, Erasmus University Rotterdam, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands

ARTICLE INFO

Article history:

Received 24 January 2013
revision received 16 May 2013
Available online 21 June 2013

Keywords:

Eye tracking
Multilevel markov models
Transitions
Temporal dependencies
Co-activation

ABSTRACT

Many eye tracking studies are designed to reveal the co-activation of representations in interactive cognitive systems, such as lexical candidates in the human language system. Such co-activation is presumed to occur within participants on a trial-level. However, traditional analyses mostly use the viewing tendency of participants over trials (e.g., average fixation proportions to visual referents), rather than individual fixation patterns within trials (e.g., consecutive fixations across visual referents). Instead, we argue that assessing temporal dependencies of eye movements between relevant referents is better suited for detecting co-activation in an interactive system, compared to other oft-used methods that may falsely accept or reject interaction hypotheses. We demonstrate how to analyze eye movement *transitions* with a multilevel markov modeling approach using a relevant experimental example (bilingual co-activation in a visual world paradigm), and discuss the practical applications and theoretical implications when analyzing transitions in any type of eye tracking data.

© 2013 Elsevier Inc. All rights reserved.

Introduction

To make sense of the world around us, we need to constantly process the sensory information in our environment. By studying how we process this continuous input, researchers attempt to understand the mechanisms that are involved in human cognition. For example, studying people's eye movements while they process spoken language, perceive visual scenes, and/or interact with their environment, can reveal the underlying processes that are called upon when performing these actions. For this purpose, eye tracking is a popular and widely used method

in cognitive psychology. A wide variety of research domains have adopted the eye tracking methodology to study such topics as language acquisition and comprehension, visuo-motor learning, memory, perception, visual cognition, social interactions, cognitive development, and action. Several methods for assessing people's viewing behavior have been used, all of which study eye movements to regions of interest (pictures, words, or locations) in a visual scene.

The *visual world paradigm* (Cooper, 1974) is an exemplary method. This paradigm is based on the notion that looking to real-world objects while simultaneously listening to spoken language might reveal cognitive mechanisms involved in language processing (Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). Measuring the eye fixations to a relevant visual scene while sentences incrementally unfold provides a sensitive real-time index of how language is being processed. By recording eye movements to real-world referents (often pictures on a screen), researchers aim to study which representations are activated during language processing, when they are activated, and to what extent. Another well-known paradigm used

* Corresponding author. Present address: University of Amsterdam, Kloveniersburgwal 48, Room D.0.11, 1012 CX Amsterdam, The Netherlands.

E-mail addresses: l.vandenberg@uva.nl (L. Vandenberg), bouwmeester@fsw.eur.nl (S. Bouwmeester), b.r.bocanegra@fsw.leidenuniv.nl (B.R. Bocanegra), zwaan@fsw.eur.nl (R.A. Zwaan).

¹ These authors have equally contributed to this paper.

² These authors are currently affiliated with different universities, but this manuscript is the result of research at the Erasmus University.

with eye tracking is *visual search* (Treisman & Gelade, 1980; Williams, 1966, as described by Findlay, 2004). Here, participants are instructed to search for certain objects on the screen. By manipulating aspects of the visual scene (such as object features, or the number of distractors), researchers can, for example, investigate how bottom up stimulus features and top down volitional control influence visual search (e.g., Henderson & Ferreira, 2004; Theeuwes, 2010; Wolfe, 1994; also see Hartsuiker, Huettig, & Olivers, 2011).

Cognitive interactions

Researchers have used eye movements to study whether information processing in the human representational system occurs in an interactive or non-interactive fashion (for a discussion see e.g., Stephen & Mirman, 2010). Cognitive systems in general are assumed to contain several processing levels representing different levels of abstraction, within which each level consists of various representational units. In non-interactive systems the information stream flows unidirectional through encapsulated levels (with information from each level affecting the next, i.e., bottom-up), whereas it is multidirectional in interactive systems (i.e., simultaneously bottom-up, top-down, and laterally within a level, cf. McClelland & Rumelhart, 1981; Rumelhart & McClelland, 1982). Such a multidirectional communication within and between adjacent representational levels is often implemented through excitatory and inhibitory connections between representational units. This results in *interactive activation* of representational units (McClelland & Rumelhart, 1981), meaning that the (extent of) activation of one representational unit depends on the (extent of) activation of another representational unit. As a result, one important property of an interactive system is that it allows for *co-activation* of multiple representational units, meaning that multiple representations are activated, after which one representation can be selected through a combination of bottom-up, top-down, and lateral (inhibitory) processing streams.

The detection of co-activation of representations has important consequences for the interpretation of human cognition and behavior. Successful performance on behavioral or mental activities (such as producing or comprehending language) depends on the efficiency with which mental representations can be retrieved from memory. In a non-interactive processing system such representations are activated independently, whereas in an interactive processing system, activation occurs interdependently.

The idea of using eye movements to assess interactive activation of representations is that it may influence where observers fixate during the viewing of a scene, i.e., which referents they attend to. Therefore, researchers who study people's eye movements in visual world paradigms are often interested in whether activation of one representation co-occurs with activation of another representation, which would suggest co-activation (and therefore interactive activation) within the cognitive system that underlies viewing behavior.

One way of investigating interactive activation is to assess the distributional signature of latent cognitive

processes on viewing behavior. For example, Stephen and Mirman (2010) have analyzed the distributions of gaze steps (i.e., the Euclidian distance between consecutive gaze positions) in visual search and visual world tasks and found that these distributions showed evidence for interacting processes that drive visual cognition (but also see Bogartz & Staub, 2012, for a substantial criticism). Although this distributional approach is able to test the nature (either interactive or not) of the underlying cognitive systems, it cannot reveal specific interactions between hypothesized representational units.

The present paper is concerned with a more fine-grained approach to assess whether the activation of specific representations are contingent upon each other or not. With this purpose, researchers usually define several regions of interest, using visual referents to elicit fixations that are hypothesized to reveal activation of specific underlying representations. Generally, they assess the fixation durations or fixation frequencies to each region of interest for several time intervals. Next, the average proportion of fixations is calculated for each time interval within participants or within trials and they are compared for the visual referents of interest. We will argue that comparing the fixation proportions over trials between referents may lead to substantial problems, because they can lead one to erroneously conclude the presence and absence of co-activation of representations in a cognitive system. As we will argue throughout this paper, the assessment of co-activated representations ideally occurs within participants *and within trials*.

Analyzing *transitions* between regions of interest circumvents the disadvantages of proportions because they are calculated within trials. They can therefore provide the information needed to detect co-activation within trials. Transition analyses (i.e., assessing dependencies between different states) have been performed in different eye tracking paradigms (e.g., Althoff & Cohen, 1999; Henderson, Falk, Minut, Dyer, & Mahadevan, 2000; Simola, Salojärvi, & Kojo, 2008) and are arguably preferable over fixation proportions³ for answering different research questions. In this paper, we will argue that transitions are ideally suited for assessing the presence or absence of dependencies between the activation of representations.

A relevant framework to demonstrate this is that of co-activated languages in a visual world paradigm. Because this framework leads to relatively simple predictions about first-order transitions, this example will be used throughout the paper. However, the points we make generalize to any type of eye tracking paradigm in which researchers assess the activation of representations in interactive systems that influence eye movements to visual referents within a single predefined event. First, we will outline the theoretical framework from our example and some conceptual and methodological issues that arise when conceptualizing co-activated representations. Next, we will illustrate the importance of assessing individual transition patterns with two hypothetical situations.

³ We thank Florian T. Jaeger for pointing out related discussions at CUNY conferences (Frank, Salverda, Jaeger, & Tanenhaus, 2009; Tanenhaus, Frank, Jaeger, Masharov, & Salverda, 2008).

Theoretical framework

Several studies that have adopted the visual world paradigm concluded that auditory linguistic input can lead to co-activated representations within different representational levels, for example in physical shapes (e.g., *snake – rope*, Dahan & Tanenhaus, 2005), semantic associates (e.g., *piano – trumpet*, Huettig & Altmann, 2005; Yee & Sedivy, 2006), lexical candidates within a language (e.g., *beaker – beetle*, Allopenna, Magnuson, & Tanenhaus, 1998; e.g., *candle – candy*, Spivey-Knowlton, Tanenhaus, Eberhard, & Sedivy, 1998, also see Huettig & McQueen, 2007), and across languages (e.g., *marku – marker*, Spivey & Marian, 1999). Most of these conclusions were based on fixation proportions that were averaged across participants, trials, and time (but also see Huettig & Altmann, 2005). For example, when fluent Russian-English bilinguals were instructed in Russian to manipulate a target picture of a “*marku*” (meaning *stamp* in English), a higher proportion of fixations was directed at a competitor picture of a *marker* than at distractor pictures with no phonetic overlap in either language (Spivey & Marian, 1999).

Such differences in fixation proportions (see Marian, Blumenfeld, & Boukrina, 2007; Marian & Spivey, 2003; Marian, Spivey, & Hirsch, 2003 for similar findings) are taken as support for co-activation of lexical representations from different languages, or cross-language lexical access. This is the leading view that when bilinguals are presented with a word that has (semantic, orthographic, or phonological) overlap across languages, lexical candidates from both languages are activated (e.g., Brysbaert, Van Dyck, & Van de Poel, 1999; Caramazza & Brones, 1979; Grainger & Beauvillain, 1987; Macnamara & Kushnir, 1971; Schulpen, Dijkstra, Schriefers, & Hasper, 2003; Vandenberg, Guadalupe, & Zwaan, 2011; von Studnitz & Green, 2002). This view is reflected in interactive models of speech recognition (cf. the interactive activation model by McClelland & Rumelhart, 1981; Rumelhart & McClelland, 1982), such as the cohort model (Marslen-Wilson, 1987; Marslen-Wilson & Welsh, 1978), the TRACE model (McClelland & Elman, 1986), and the Bilingual Interactive Activation (BIA+) model (Dijkstra & Van Heuven, 2002). Generally, these models describe one integrated language system containing the lexical representations of multiple languages which can be co-activated.

These theoretical frameworks suggest that the activation of one representation depends on the activation of another representation when processing language. When using eye movements to measure such dependencies in activation during the processing of a linguistic utterance, we argue that one should ideally assess dependencies between critical fixations. Consider the following experimental example. If a Dutch-English bilingual hears the English word “*mice*”, the phonologically similar Dutch word form “*mais*” (meaning *corn*) may also be activated. Assuming that eye movements to relevant objects indeed reveal lexical activation, this co-activation of lexical candidates should result in more looks to both the English depiction (*mice*) as well as the Dutch depiction (*corn*) of the word, relative to the irrelevant distractors (pictures of e.g., *leaves* and *apples*). Importantly, the two representations should not be activated at just any independent points during

the experiment, but their activation should be *temporally contingent* (co-activated) within a trial. Such temporal dependencies on a representational level (activation of the English and Dutch representation) can be detected by temporal dependencies in viewing behavior (fixations on the English and Dutch depictions). Transitions (switches) between the two critical pictures within a trial are ideally suited to capture such dependencies. We therefore argue that transitions are an ideal measure for co-activation.

In other words, if the English and Dutch word forms are both activated upon hearing the word “*mice*” (see e.g., Dijkstra & Van Heuven, 2002), the word “*mice*” should evoke fixations to both the English and the Dutch depiction. Because it is impossible to fixate two pictures simultaneously given the visual angle at which the pictures are presented, the way to tap into such temporal dependencies in lexical activation is to assess whether critical fixations occur alternately for the critical referent pictures. Such alternations would provide a strong indication of co-activation on a representational level.

With this claim, it is not our intention to define co-activation within a specific architectural framework. Though representations in an interactive system are by definition co-activated, other frameworks may implement dependencies in the activation of representations in other ways. As such, different architectures might result in the same temporal dependencies in overt viewing behavior. Most importantly however, researchers use visual world or visual search paradigms to assess the theoretical claim that the activation of a representation is dependent or independent of the activation of another representation. In this paper we propose a method for detecting such dependencies by assessing whether fixating one picture is conditionally dependent on fixating another picture.

Conceptual and methodological issues

Many eye tracking studies, such as those described in the current theoretical framework, are analyzed by performing *t*-tests (ANOVAs) on fixation durations, fixation proportions, or fixation counts to each region of interest within a certain time interval. However, several researchers have recently proposed alternative analyses, such as modeling growth curves on fixation proportions (Magnuson, Dixon, Tanenhaus, & Aslin, 2007; Mirman, Dixon, & Magnuson, 2008) or performing binomial multilevel analyses on log odds (Barr, 2008). The latter analyses are perfectly suited for establishing activation of representations as an aggregate property over events, e.g., within participants but across trials. However, many theories—including that of cross-lexical activation—specifically predict co-activation within participants *and* within trials. In other words, a minimal prerequisite for claiming bilingual co-activation is that both the English lexical candidate *mice* and the Dutch lexical candidate *mais* are activated in the “*mice*” trial (rather than activating the English candidate in some trials and the Dutch candidate in others, which may lead to an aggregated number that would have suggested the activation of both languages). We therefore argue that hypotheses based on such a *momentary* co-activation are ideally addressed within trials, because the description level of

the analysis is in concordance with that of the theoretical claim.

An alternative solution to the problem of aggregated fixations would be to compare fixation *latencies* to the English referent (“mice”) in the presence of the Dutch referent (“mais”), compared to a control condition in which the Dutch referent is not present in the display (cf. the design explained by Barr, 2008, in the special issue of the Journal of Memory and Language on emerging data analysis). A delayed fixation to the English referent in the presence (versus absence) of the Dutch referent should indicate co-activation of lexical candidates. Although this approach is capable of detecting co-activation, it does so at a coarser level. First, delays have to be measured across trials. That is, this method always compares a trial in which a referent was present to a different trial in which it was absent. As a result, assessing delays always occurs (a) across items, in which a participant responds to different items in different conditions, which does not allow the researcher to detect momentary interactions within a certain participant on any certain item, or (b) within items but across trials, in which a participant responds to the same item in different conditions; in this case, one item is presented multiple times which is undesirable in, for example, language research. Second, delays infer the activation of a representation based on absent behavior, i.e., the duration of not looking at the target. As a result, there is not much specific information about to what extent the other representation is activated at which point in time. Thus, unlike delays, transitions have the advantage of being able to measure within trials and to explicitly uncover activation of a certain representation at a certain point in time. The advantage of transitions therefore is that they can specifically pinpoint *which* representations are activated and *when* this occurs. Furthermore, the proposed transition analysis allows the assessment of a multinomial dependent variable. This allows us to assess the viewing behavior to all four regions within a single model, which is a stronger fit to the actual data than modeling the viewing behavior to the one or two critical regions only. And, as we will argue in the discussion, a transitional approach has the potential to assess the relative strength of the connections between representations, which can give insight into the structural hierarchy of the representational system.

In the present paper we argue that assessing transitions within participants and within trials to investigate co-activation can overcome many pitfalls of conventional approaches. To illustrate this, consider panels A and B from Table 1. Both panels reflect the hypothetical fixation pattern of a participant over time to four regions on a screen in ten different trials. Region EN represents the referent for the English representation (e.g., the picture of the English *mice* from our example), region DU refers to the referent for the Dutch representation (e.g., the picture of corn, the Dutch *mais*), and regions d1 and d2 refer to distractor pictures that do not represent a relevant representation (e.g., apples and leaves). In trial 1 from panel A, for example, the individual first fixated a distractor (d1), then made a transition to the referent for the Dutch representation (DU), kept fixating this referent, and finally fixated the other distractor (d2).

Critically, Panel A shows a situation in which analyzing the commonly used fixation proportions (or log odds) per time frame would lead to the erroneous conclusion of co-activated representations (or, in terms of our example, that hearing the English word form *mice* activated the Dutch word form *mais*, which corresponds to the picture of corn). This conclusion would be based on a higher proportion of fixations on critical regions EN and DU relative to distractor regions d1 and d2. However, the individual trials show that there were only three out of ten trials in which our hypothetical participant fixated both critical regions, and importantly, that these critical fixations were never sequential. In terms of our example this shows that there were no trials in which the participant made a transition between the English and Dutch referents upon hearing the target word (e.g., ‘*mice*’). For this reason, these data do not provide convincing evidence that *both word meanings were activated upon hearing the linguistic utterance*. However, analyzing fixation proportions by means of ANOVAs would have led to the erroneous conclusion that the representations were co-activated. Approaches that do not average across trials do not have this problem (also see the points made by Barr (2008)).

On the other hand, Panel B shows a situation in which analyses over aggregated data would erroneously conclude that representations were not co-activated, given the equal proportions of fixating the referent for the Dutch representation (region DU) and the distractors (regions d1 and d2): Not only did participants fixate *both* critical pictures on seven out of ten trials, they also did this in a consistent and temporally contingent manner (switching from region EN to DU and DU to EN). This would provide convincing evidence for the momentary co-activation of the English and Dutch representations, i.e., for cross-language activation. Essentially, the temporal contingencies between target and competitor fixations are indicative that both the English and Dutch lexical candidates are activated upon hearing the word “*mice*”.

These examples show that, because fixation patterns vary considerably between individuals and over trials, analyses based on fixation proportions can lead to erroneous conclusions concerning both the presence *and* absence of co-activation. For this reason, we argue for a fine-grained test for co-activation between representations using fixation transitions between critical pictures relative to distractor pictures within individual trials. As illustrated by our hypothetical examples, it is not sufficient to conclude that the critical referents were fixated across trials or across participants (i.e., over events). Rather, one should ideally establish whether a fixation to one referent is temporally dependent on a fixation to the other referent (i.e., within events, e.g., during the trial “*mice*”). Of course, as a result of visual search for the target, distractors will be fixated at times and switches from and to these pictures will occur. However, co-activation inherently implies a temporal dependency between the activation of multiple representations, which in general terms should also result in a temporal dependency between the fixations to the visual referents.

An ideal way to assess individual transition patterns is by means of *markov modeling*, because this technique

Table 1
Hypothetical viewing patterns.

New	Timeframe (ms)					Total
	0–250	251–500	501–750	751–1000		
<i>A. No temporal contingency between critical regions 1 and 2</i>						
Trials	1.	d1	DU	DU	d2	
	2.	DU	DU	d1	EN	
	3.	EN	d1	DU	d1	
	4.	EN	EN	EN	EN	
	5.	d1	EN	EN	EN	
	6.	d2	DU	DU	DU	
	7.	DU	DU	d2	EN	
	8.	d1	EN	EN	EN	
	9.	d2	d2	EN	EN	
	10.	EN	EN	EN	EN	
Proportions	Region EN	0.3	0.4	0.5	0.7	0.48
	Region DU	0.2	0.4	0.3	0.1	0.25
	Region d1	0.3	0.1	0.1	0.1	0.15
	Region d2	0.2	0.1	0.1	0.1	0.13
		0–250	251–500	501–750	751–1000	Total
<i>B. Temporal contingency between critical regions 1 and 2</i>						
Trials	1.	DU	EN	d2	EN	
	2.	EN	DU	d2	EN	
	3.	d1	EN	EN	DU	
	4.	DU	EN	d1	EN	
	5.	d2	d2	DU	EN	
	6.	EN	d2	EN	EN	
	7.	d1	EN	DU	d2	
	8.	d1	DU	EN	EN	
	9.	EN	d1	EN	d1	
	10.	d2	d1	EN	EN	
Proportions	Region EN	0.3	0.4	0.5	0.7	0.48
	Region DU	0.2	0.2	0.2	0.1	0.18
	Region d1	0.3	0.2	0.1	0.1	0.18
	Region d2	0.2	0.2	0.2	0.1	0.18

Note. This table illustrates the hypothetical fixation pattern of one participant in ten trials after onset of the target word in case of temporal contingency between critical fixations (panel B; transitions from region EN to DU and vice versa) or no temporal contingency (panel A; no transitions between regions EN and DU within trials). The text within a cell represents fixations on respective regions of interest. Each column represents a time interval; each row represents a trial (in the upper part of each panel) or averaged proportion of fixations across trials (in the lower part of each panel).

assesses changes in fixations (states) over time, calculates transition probabilities, and properly analyzes the data by addressing the temporal dependencies of the switches (see Cristino, Mathot, Theeuwes, & Gilchrist, 2010; Hwang, Wang, & Pomplun, 2011; von der Malsburg & Vasishth, 2011 for different methods of assessing direct transitions or sequences in eye movements). Also, *multilevel* markov models are able to account for the multilevel structure of the data (trials are nested within participants). The effects that can be tested are either first-order transitions (assessing direct transitions from fixating one region of interest to the next, as we hypothesize for this study) or higher-order transitions (reflecting a chain of two or more transitions to different regions of interest, i.e., the sequence of fixations). In the following section we will describe a study that is tailored to the present discussion, in order to illustrate the type of predictions that can be made in markov modeling and demonstrate how this technique can be used in analyzing eye tracking data.

The present study will assess cross-language activation in the operationalization that is explained throughout the introduction. As discussed in *Future development and*

considerations, there are conceivable circumstances where temporal dependencies between representations might not necessarily result in direct (first-order) transitions, but may affect indirect (higher-order) transitions. However, considering that we have no a priori-hypotheses about such higher order transitions, we only assess first-order transitions.

Application

Method

Participants

Forty-eight undergraduate students at the Erasmus University Rotterdam participated in the experiment for course credit or payment. The description of the Experiment stated that native Dutch (L1) speakers with second language English (L2) could sign up for an English experiment testing the English proficiency in Dutch students. Their proficiency in English was assessed by the Language Experience and Proficiency Questionnaire (LEAP-Q:

Marian, Blumenfeld, & Kaushanskaya, 2007). The mean self-rated English skills were 7.0 ($SD = 1.15$) on a scale from 1 to 10 (averaged over speaking skills, reading skills, and understanding spoken language). The mean age at which these unbalanced bilinguals acquired English was 9.4 ($SD = 2.1$). All had normal or corrected-to-normal vision.

Stimuli and apparatus

The auditory stimulus materials were derived from a pilot experiment that provided an empirical measure of phonological overlap. They were recorded by a male native speaker of Dutch who had lived in the US for 15 years, because the pilot experiment required the homophones to be pronounced both in English and Dutch. A native speaker of American English assessed the stimuli to ensure that they were pronounced correctly. We selected the data of 14 homophone-items from a pool of 43 English one- or two-syllable homophones (e.g., *mice*) and 43 English one- or two-syllable filler words (e.g., *grapes*). Each word was preceded by the carrier phrase “Click on the...”, which lasted 1078 ms. The auditory stimuli were sampled at 44.1 kHz.

The visual world, consisting of four quadrants on the 21 inch display of a remote Tobii 2150 eye tracker, was shown during each auditorily presented sentence. The participants' eye positions were continually registered at a sampling rate of 50 Hz. Each quadrant in the visual world contained one picture. In the homophone condition, the target picture (e.g., *mice*) was accompanied by a competitor reflecting the Dutch meaning of the homophone word (e.g., *corn*) and two distractor pictures, with no phonological overlap with the target (e.g., *apples* and *leaves*).⁴ Each picture occurred in one visual display only, and each visual display was presented once. The positioning of the pictures from quadrants 1 to 4 was randomized across trials and across participants. In addition, the order of trials was randomized across participants.

Procedure and design

The experimental session consisted of three parts. First, participants performed the visual world eye tracking task. They were seated in front of the eye tracker. After calibration, they read the English instructions on the computer

screen. They were instructed to listen to the presented sentences carefully and then click on the picture as instructed in the sentence. The participants performed five practice trials that were similar to the experimental trials but included different words and pictures. If there were no further questions after the practice trials, the experiment started. Before each trial, a fixation cross and the mouse cursor were presented in the middle of the screen. When the eye tracker recorded 500 ms of fixations to the cross, the trial started. The fixation cross disappeared and the visual world appeared on the screen simultaneously with the onset of the sentence. Each trial was terminated at the moment the participant clicked on a picture. 1000 ms after the trial had ended the fixation cross again appeared on the screen and the mouse cursor returned to its central position, leaving an equal distance between the mouse cursor and each of the four pictures at the onset of the trial.

Afterwards, participants filled out a Dutch translation of the validated Language Experience And Proficiency – Questionnaire (Marian, Blumenfeld, & Kaushanskaya, 2007) on the computer. This questionnaire was designed to assess the participants' language profile. At the end of the experimental session, participants performed a vocabulary test. During this test, they were auditorily presented with the English homophone target items to translate into Dutch. The trials in which the participants did not know the correct translations would be removed from the eye tracking data for further analyses, making the vocabulary test a useful selection mechanism both at participant and at item level. The entire experiment lasted approximately 45 min.

Exclusion criteria

Participants were excluded from the sample when they did not report English as being their real L2 ($N = 2$), had extremely high error rates in the vocabulary test (above .50, $N = 9$), or as a result of software malfunction ($N = 2$). Furthermore, trials that were defined erroneously in the vocabulary test or had extremely long response latencies (over 3500 ms after target onset) were excluded from the analysis.

Predictions

Importantly, the current hypothesis about cross-language activation should be formulated in terms of transitions towards targets, competitors, and distractors within a person within a trial. For this purpose, markov modeling provides an ideal approach. The results of a markov model analysis can simply be summarized in a transition matrix containing estimations of the transition probabilities to each picture. Markov modeling allows us to test detailed predictions regarding the content of transition matrices if co-activation occurs versus if it does not occur. The transition matrix contains (1) the probability of staying fixated on a certain picture (referred to as “staying probabilities”, resembling the fixation proportions that are used in traditional analyses) and (2) the probability of switching from one picture to another (i.e., transitions between pictures, referred to as “switching probabilities”). Given that we were

⁴ Because the lexical properties of the word forms for each picture in a trial might affect cross-language activation, we controlled for several of such properties. First, we calculated the frequency per million for both the English and Dutch word form belonging to each picture (assessed in CLEARPOND, Marian, Bartolotti, Chabal, & Shook, 2012). A repeated measures ANOVA on the frequencies per picture type (EN, DU, d1, d2) showed no significant differences in both English ($F < 1$) and Dutch ($F < 1$) word form frequencies. Next, we calculated a measure of semantic overlap between the lexical candidates belonging to each possible picture pair in a trial (EN-DU; EN-d1; EN-d2; DU-d1;DU-d2;d1-d2) by Latent Semantic Analysis (<http://lsa.colorado.edu/>). A repeated measures ANOVA showed that the semantic overlap between each picture pair within a trial was low ($M = 0.09$, $SD = 0.10$) and similar across picture types ($F(5,9) = 0.30$, $p = .90$). Finally, we assessed the English–Dutch cognate status of the distractors (calculated with the formula used by Midgley, Holcomb, and Grainger (2011)). Paired-samples *t*-tests showed that the orthographic ($M = 0.28$, $SD = 0.34$) and phonological ($M = 0.24$, $SD = 0.29$) overlap between the distractors' translation equivalents was low and did not differ for the different distractor categories ($t_s < 1.23$, $p_s > .24$, $r_s < .11$).

Table 2

Prediction matrix depicting the relative staying and switching probabilities for correct trials in the case of co-activation.

		Time frame = t			
		English	Dutch	Distractor 1	Distractor 2
Time frame = $t - 1$	English	A	a	b	b
	Dutch	c	B	d	d
	Distractor 1	e	f	C	f
	Distractor 2	e	f	f	C

Note. Identical letters represent equal (expected) proportions. The diagonal cells (containing capital letters) represent the staying predictions; the off-diagonal cells (containing lowercase letters) represent the switching predictions. The critical hypothesized relationships are: $a > b$, $c > d$.

interested in the pattern of probabilities in our model rather than their absolute sizes, we formulated all staying probabilities relative to the other staying probabilities (using capital letters), and all switching probabilities relative to the other switching probabilities (using lowercase letters).

Table 2 shows the hypothetical transition matrix, depicting the relative staying and switching probabilities in the case that languages are co-activated and the referents of both languages show temporal dependencies. The predictions are formulated as the probability of switching to a certain picture (at a certain time frame t , which is represented by the columns in the matrices) given a fixation on a certain other picture (at the preceding time frame $t - 1$, as represented by the rows in the matrices). In other words, the hypotheses pertaining to co-activation are based on the previously attended picture and comparisons between switching probabilities are made on a row level in the matrices. The diagonal cells in the matrix depict the staying probabilities and the off-diagonal cells depict the switching probabilities. This prediction matrix can be mapped directly onto the actual transition matrix that results from the markov model analyses.

Previous studies (e.g., Dahan & Tanenhaus, 2005; Huettig & Altmann, 2005; Spivey & Marian, 1999; Spivey-Knowlton et al., 1998) have found that participants showed a higher fixation proportion on the target than on the other pictures, as a result of the task demands. Furthermore, they found that participants showed a higher proportion of fixations on the competitor than on the irrelevant distractor pictures, but that this fixation proportion was smaller than that on the target. This asymmetry is depicted in Table 2 on the diagonal axis, showing the highest probability of a participant fixating the English target and remaining fixated on the English target in the successive time frame (A), a lower probability of fixating the Dutch competitor and staying fixated there in the successive time frame ($A > B$) and the lowest probability of fixating one of the distractor pictures and staying fixated ($A > B > C$).

Importantly, English target and Dutch competitor fixations should be temporally dependent if both lexical candidates are co-activated, as reflected in the off-diagonal cells of the upper two rows. If the English target is fixated by the participant, the probability of switching to the Dutch competitor should be larger than the probability of switching to one of the distractor pictures ($a > b$). The probability of switching from the Dutch competitor to the English target should be larger than the probability of switching to a distractor ($c > d$). However, because the task instructed to

click on the English target, we expect that every switch would be most likely made to the English target rather than any other picture. If a participant fixated a distractor, the probability of switching to the English target should also be greater than the probability of switching to the Dutch competitor or the other irrelevant distractor, because of the task demands ($e > f$). In terms of our critical hypothesis, this shows that the probability of switching from the Dutch competitor to the English target (relative to switching to a distractor, $c > d$), is not the most informative measure of a temporal dependency between the English and Dutch depiction. For this reason, the most critical comparison is that in row one of Table 2: Are people more likely to switch from the English target to the Dutch competitor than to a distractor ($a > b$)? This comparison is depicted in Fig. 1.

Equations of the multilevel markov model

The multilevel markov model can be described by two multinomial regression equations which describe the distribution of the initial probabilities on the first time frame and the transition probabilities from time $t - 1$ to time t . The two sections below describe the two sets of multinomial regression equations that are involved in the current first-order multilevel markov models.⁵ In our design, we did not include a predictor. However, many experimental designs are aimed at detecting differences in viewing behavior based on an independent variable, such as age (continuous), language group (nominal: e.g., monolingual/bilingual), or time (either continuous or in intervals). We will therefore demonstrate the inclusion of such a predictor in Validation of the models, which is why the exposition of the equations already includes a predictor. The equations will be familiar regression functions like $Y = a + bX + c$, in which the dependent variable Y is interpreted as the probability of fixating one of four pictures (EN [1], DU [2], d1 [3], d2 [4]: multinomial), fixed intercept a is interpreted as the baseline probability of fixating a certain picture region, coefficient b is interpreted as the additional probability of fixating that picture depending on the level of the predictor, X determines the level of the predictor, and random intercept c is interpreted as the additional probability of fixating the picture

⁵ For an introduction into multilevel modeling techniques see Baayen, Davidson, and Bates (2008). For categorical (multinomial) multilevel modeling in specific, see Agresti (2002) or Jaeger (2008), for example.

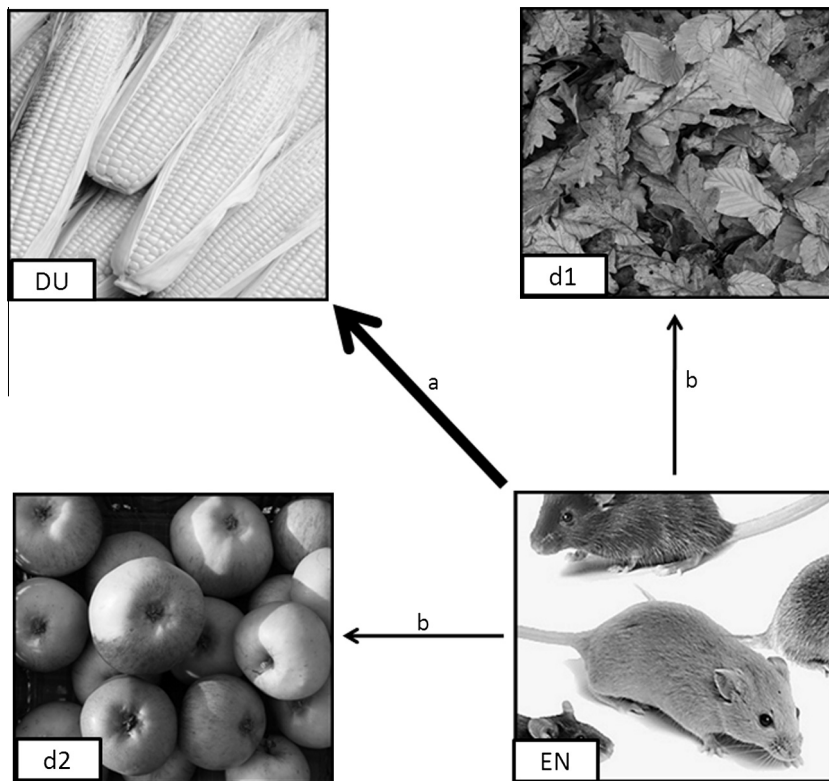


Fig. 1. Exemplary trial of the Visual World Paradigm, in which the critical switching probabilities are depicted by the arrows.

for a certain participant (which is the multilevel component of the equation).

The supplemental material (doi) contains an example of the data structure (Section A), the syntax (Section B), and the restrictions in the equations (Section C) we used for markov modeling in Latent Gold 4.5 (Vermunt & Magidson, 2005; Vermunt & Magidson, 2008).

Initial probabilities

In the following equations, the letter i represents a participant ($i = 1, \dots, N$), k represents a trial ($k = 1, \dots, K$), s represents a fixation on a particular picture [$s = 1, 2, 3, 4$; $\sum_{s=1}^4 p(s_t) = 1$], and t represents a time frame of 100 ms⁶ ($t = 1, \dots, T$, in which T represents the last time frame of a trial, i.e., the moment of the click response which may differ across participants and trials). The log odds of the initial fixation probabilities ($t = 1$) on a picture (s) by a participant (i) on a trial is defined as: $\pi_{is,t=1}$.

Because multiple trials are nested within participants, the fixation patterns of trials are dependent on the participant. To correct for this dependency, coefficient $\pi_{is,t=1}$ is modeled as a function of fixed and random effects:

$$\pi_{is,t=1} = \alpha_{s,t=1} + \sum_{r=1}^R \gamma_{v|ul} \cdot r + g_{is,t=1}.$$

⁶ The data were aggregated to time frames of 100 ms, for reasons that will be thoroughly discussed in the discussion (Future development and considerations).

$\alpha_{s,t=1}$ is the fixed intercept, reflecting the log odds of a fixation on picture s at $t = 1$. $\alpha_{sr,t=1}$ reflects the fixed effect of a predictor r ($r = 1, \dots, R$) on a fixation on picture s at $t = 1$, and allows initial fixations to be different for different levels of the predictor (e.g., for monolinguals versus bilinguals, or at different predefined time intervals within a trial). Finally, we have included a random intercept $g_{is,t=1}$ for individual participants.⁷ This effect of participants is assumed to be normally distributed and centered around zero: $\sim N(0, \sigma^2)$.

Because the coefficients $\pi_{is,t=1}$ sum to zero over s pictures, we estimated $s - 1$ ($=3$) parameters for each effect of $\alpha_{s,t=1}$, $\alpha_{sr,t=1}$ and $g_{is,t=1}$. The estimated log odds $\pi_{is,t=1}$ in the equation can be transformed into an initial fixation probability $\phi_{is,t=1}$ which has a multinomial distribution:

$$\phi_{is,t=1} = \frac{\exp(\pi_{is,t=1})}{\sum_{s=1}^S \exp(\pi_{is,t=1})}.$$

Transition probabilities

In the following equations, the letter u represents the fixation on a picture s at timeframe $t - 1$ [$u = 1, \dots, 4$; $\sum_{u=1}^4 p(u) = 1$], whereas v represents the fixation on a picture s at timeframe t , [$v = 1, \dots, 4$; $\sum_{v=1}^4 p(v) = 1$]. The log odds that a participant i on a trial k fixated picture v at time t , given that (s)he fixated picture u in the previous time

⁷ No random slopes were included because we assumed that the effect of the predictors on the initial fixation was equal for participants.

frame $t - 1$ is defined as $\pi_{i,v|u}$. This coefficient is in turn modeled as a function of fixed and random effects:

$$\pi_{i,v|u} = \gamma_{v|u0} + \sum_{r=1}^R \gamma_{v|ur} \cdot r + l_{i,v|u}$$

$\gamma_{v|u0}$ is the fixed intercept, representing the log odds of a transition from picture u to picture v . $\gamma_{v|ur}$ represents the effect of a predictor r ($r = 1, \dots, R$) on a transition from picture u to picture v . Intercept $l_{i,v|u} \sim N(0, \sigma^2)$ represents the random effect of individual participants on transition probabilities.

Because the coefficients $\gamma_{v|u0}$, $\gamma_{v|ur}$, and $l_{i,v|u}$ sum to zero over v , we estimated $s * (s - 1)$ ($=12$) parameters for each effect of $\gamma_{v|u0}$, $\gamma_{v|ur}$, and $l_{i,v|u}$. The estimated log odds $\pi_{i,v|u}$ in the equation above can be transformed into a transition probability $\phi_{i,v|u}$ which has a multinomial distribution:

$$\phi_{i,v|u} = \frac{\exp(\pi_{i,v|u})}{\sum_{u=1}^U \exp(\pi_{i,v|u})}$$

Results

The mean accuracy on the 14 homophone trials was 0.86 ($SD = 0.14$). The mean reaction time for the correct responses was 1913 ms ($SD = 913$) after target onset. Out of the false responses (corresponding to 55 trials), 92.73% was made to the Dutch competitor, whereas 1.82% was made to a distractor picture and 5.45% was not made to any picture at all (participants clicked next to a picture). These false responses were excluded from further analyses. The analyses were performed on the data after 200 ms from the onset of the target up until the click response, because it takes approximately 200–300 ms to plan and fully execute an eye movement⁸ (Hallet, 1986; Matin, Shao, & Boff, 1993). As a result, the dataset contained the data of 35 participants on a total of 339 correct homophone trials, resulting in 5199 data points (measurements in time frames of 100 ms).

On average, participants mainly fixated the English target (proportion averaged over participants, trials, and time $M = 0.58$, $SE = 0.02$), fixated the Dutch competitor not as much ($M = 0.19$, $SE = 0.01$), and fixated each of the distractor pictures even less (for both distractors $M = 0.11$, $SE = 0.01$).⁹ Fixation patterns were analyzed using the software package Latent Gold 4.5 (Vermunt & Magidson, 2005; Vermunt &

Magidson, 2008). The parameters of the models were obtained by maximum-likelihood estimation. Because the models were nested, we could use the difference between two models in log likelihood times -2 ($-2LL_{diff}$) to test the difference in fit of the two models using a chi square distribution, in which the degrees of freedom are equal to the difference in number of parameters of the two nested models.

Hypothesis testing

In order to test the cross-language activation hypothesis we estimated a number of predetermined models (based on the theory) and compared their fit to the data. First, we performed a manipulation check to assess whether some important preconditions were met. Next, we built towards a model that performed the critical comparison for assessing co-activation between languages. Finally, we performed an additional test with a predictor (time interval) to validate our findings.

Manipulation check

The manipulation check was performed to test the assumption that both distractors were equally salient in our setup. Equal saliency would confirm that (a) the random positioning of the pictures was successful and (b) the contents of both distractor pictures were treated equal in this task. If this were not the case, any of the following results might have been caused by a failed manipulation rather than by cross-language activation.¹⁰

The check consisted of the comparison of two models. In model 0, the distractor probabilities were based on freely estimated parameters and therefore allowed to vary. In model 1, the probabilities of the distractors in the transition matrix were constrained to be equal. If the distractors were equally salient in our setup we would expect that the fit of model 0 with the free distractor parameters was not significantly better than the fit of model 1 with the constrained distractor parameters. As expected, the fit of model 0 was not statistically better than the fit of the constrained model 1 ($-2LL_{diff}(16) = 23.12$, $p = .11$). This shows that the switching and staying probabilities did not differ for both distractor pictures, suggesting equal saliency for both distractors. For this reason, we took model 1 as the baseline multilevel markov model against which to compare the critical models.

Critical tests

Next, we estimated several models to perform the critical tests for co-activation. These models assess whether participants who were fixating the English target item were more likely to transition to the Dutch competitor than to either unrelated distractor item. A discussion of these critical models is provided below, their fit to the data is summarized in Table 3.

⁸ Altmann (2011) has reported the earliest meaningful saccade launch latencies (discriminating signal from noise) at about 100 ms. These launch times refer to the shortest possible time it takes to plan and initiate a saccade to the target picture, whereas we refer to the time it takes to plan and fully execute a saccade up until the eyes first land on the target.

⁹ To draw an analogy to the more traditional analyses and for the sake of illustration, we performed a repeated measures ANOVA on the fixation proportions with picture type (EN, DU, d1, d2) as within-subjects factors. The results showed a significant effect of picture type ($F(2.14, 72.65) = 316.48$, $p < .0001$, with Greenhouse–Geisser corrected degrees of freedom). Follow-up paired-samples t -tests showed significant differences between each comparison ($ts > 4.80$, $ps < .0001$) except between the two distractors ($t = 0.10$, $p = .92$). The significant difference between competitor and distractor fixations would traditionally be taken as support for co-activation. However, as Table 1A demonstrates these results are not conclusive. We therefore performed markov analyses on the fixation patterns.

¹⁰ For example, temporal dependencies between target and competitor pictures might have been caused by unbalanced positioning of the pictures (e.g., target and competitor were always presented next to each other), whereas no temporal dependencies between target and competitor pictures might have been caused by a different saliency of the content of distractor pictures (e.g., one distractor picture was so salient that it artificially reduced competitor fixations).

Table 3

Fit statistics of the critical markov models.

Model	LL	#par	Comp.	$-2LL_{diff}$	p
Model 1: Baseline markov model ($d1 = d2$)	-3615.65	23			
Model 2: Restriction $dis = DU$	-3636.67	10	1-2	42.04	<.0001
Model 3: Restriction $EN \rightarrow DU = EN \rightarrow dis$	-3620.73	19	1-3	10.16	<.05

Note. EN = English target; DU = Dutch competitor; dis = distractor pictures with equal probabilities. LL = log likelihood of the model; #par = number of estimated parameters; Comp. = comparison of the models; $-2LL_{diff}$ = difference between the current model and the reference model (see *Comp.*) in log likelihood times -2 ; p = the p -value.

In order to check the assumption that the viewing behavior with respect to the Dutch competitor differs from the viewing behavior to the distractor pictures (as should be the case when both languages are co-activated) we created model 2, in which the Dutch competitor probabilities were constrained to be equal to the distractor probabilities. If viewing behavior to the competitor differs from that to the distractors, the fit of model 2 should be significantly worse than the fit of model 1. The results indeed showed that the fit of model 2 was significantly worse than the fit of model 1 ($-2LL_{diff}(13) = 42.04$, $p < .001$). This demonstrates that the staying and switching probabilities of the Dutch competitor differed significantly from the distractor probabilities.

In model 3, the probability of switching from the English target to the Dutch competitor was constrained to be equal to the probability of switching from the English target to a distractor. The comparison of the fit of model 1 and model 3 is critical for co-activation. Because co-activation should be reflected in a temporal contingency between critical fixations – resulting in a higher probability of switching from the English target to the Dutch competitor than from the English target to a distractor, we expected model 3 to provide a significantly worse fit than model 1. Indeed, the fit of model 3 was significantly worse than the fit of model 1 ($-2LL_{diff}(4) = 10.16$, $p < .05$). This demonstrates that the probability of switching from the English target to the Dutch competitor was significantly different from the probability of switching from the English target to one of the distractor pictures.

Table 4

Estimated transition matrix of the staying probabilities and switching probabilities.

	Time frame = t				
	English	Dutch	Dis1	Dis2	
Time frame = $t - 1$	English	.908	.040	.026	.026
	Dutch	.137	.763	.050	.050
	Dis1	.169	.079	.671	.080
	Dis2	.169	.079	.080	.671

Note. The probabilities in each row add up to 1 (The presented probabilities in the bottom rows of the matrix add up to .99 due to rounding to 3 decimals.). The probabilities of the eye gaze staying fixated upon a certain picture are always greater than the probabilities of switching between two pictures. This is a result of calculating the transition probabilities per time frame of 100 ms. It is impossible to switch from one picture to another every 100 ms, given that it takes 200–300 ms to plan an eye movement. Therefore, the staying probabilities can only be interpreted relative to the other staying probabilities, and switching probabilities relative to the other switching probabilities within a row, as pointed out previously in the prediction matrix.

Table 4 shows the transition matrix belonging to the best fitting model, model 1.

Comparison of the transition matrix from the correct trials in Table 4 to the prediction matrix in Table 2 reveals an identical match. The observed transition matrix shows that when participants fixated the English target, they were more likely to switch to the Dutch competitor (.04) than to one of the two distractor pictures (.026). This is supported by the fact that model 1 provided a significantly better fit to the data than a model in which the probability of switching from English to Dutch was constrained to be equal to the probability of switching from English to a distractor (model 3; $-2LL_{diff}(4) = 10.16$, $p < .05$). This finding demonstrates that fixations to the critical English target and Dutch distractor were temporally dependent in correct trials. This convincingly supports the notion that both languages were co-activated within trials in which participants responded correctly to the English target picture.

Though the difference between the Dutch switching probability of .04 and the distractor switching probability of .026 is significant, it may seem small in overall magnitude. However, it shows that out of all transitions that are made from the English picture ($1 - .908 = .092$), an estimated 43.5% (.04 out of .092) went to the Dutch competitor whereas only 28.3% (.026 out of .092) went to one of both distractor pictures. Thus, the relative differences between the transition probabilities are substantial (43.5% vs. 28.3%), even though the absolute values of the probabilities may suggest otherwise.

As mentioned in Predictions, the temporal dependency between English and Dutch fixations should be bidirectional, meaning that participants should also be more likely to switch from the Dutch to the English depiction (compared to a Dutch-distractor switch). However, this comparison is confounded by task demands. As a result of the task instructions to click on the English target, participants will be more likely to switch to the English target than a filler even in the absence of co-activation. A way to deal with this is to estimate a markov model on the data in which time is reversed. In this case, the transition probabilities are defined in terms of the probability that a switch was made *from* a certain picture (backward), which is conceptually different from switching *to* a certain picture (forward, as in the analyses above). With this approach we are able to assess whether a switch to English was more likely to come from a fixation on a Dutch competitor than from a fixation on a distractor, which would show a temporal dependency between the English and Dutch depiction beyond the dependency that was caused by task demands. The results of this additional analysis indeed show

evidence for bidirectional temporal dependencies between the critical fixations ($-2LL_{diff}(4) = 11.73$, $p < .05$). When participants switched to the English target, they were more likely to come from the Dutch competitor (.058) than from one of the two distractor pictures (.045).

Validation of the models

It is possible to include predictors on all levels of the multilevel markov model, for example to test the effect of an experimental manipulation on the outcome of a transition matrix. Such predictors may be added on a participant, trial, and/or time level. For example, one could add a variable such as language experience (monolingual/bilingual) or age (continuous) on a participant level, because people from different groups or ages might show different fixation patterns. On a trial level, item features could be added as predictors, such as word length or word frequency. Also, accuracy of the click response could be added as a predictor on a trial level within an individual, because the viewing behavior on correct trials might be different from that on incorrect trials. On a time level, time intervals (either continuous or nominal) within a trial could be added as a predictor, because it is highly plausible that fixation patterns change over time during a trial. Due to the fact that we were able to make a priori predictions about such a time change within the current paradigm, we have included a nominal predictor *time intervals* in an additional analysis. With this analysis we can investigate the validity of the previous model and demonstrate the feature of adding a predictor to a markov model.

Model 4a was based on model 1 and included the predictor *time interval*. The variable was divided into two parts; interval 1 (containing all data from 200 ms after the onset of the target up to 300 ms before the click response) and interval 2 (containing data from the last 300 ms before the click response). The selection for these time intervals was made because it would result in a priori contrasting predictions with respect to co-activation. We simply hypothesized that fixation patterns would be different at the end of a trial when a response was being made, relative to the rest of the trial when the response was being selected. Fixations at the end of a trial should likely be directed at the picture that is selected for responding, whereas fixations before response selection and execution should likely be more diverse. We therefore predicted critical temporal contingencies in interval 1, in which co-activation should be able to occur and manifest. Furthermore, we predicted no such dependencies in interval 2, given that participants would be finished evaluating the alternative responses (i.e., co-activation should be resolved) and would fixate the picture they were going to click on while planning and executing their response.

In model 4b we put a restriction on the first time interval. We constrained the probability of an English–Dutch switch to be equal to the probability of an English–distractor switch, to test the specific hypothesis that co-activation should occur in the first time interval. If the fit of model 4b (with identical critical switching probabilities) would be worse than the fit of model 4a (which allowed for different critical switching probabilities), it would confirm our hypothesis that co-activation occurred the first time

interval. The difference in fit between model 4a and model 4b was significant, $-2LL_{diff}(4) = 17.72$, $p < .005$, showing that model 4b provided a worse fit to the data than model 4a. The transition matrix belonging to model 4a showed a higher probability of switching from the English target to the Dutch competitor (.053) than to a distractor (.032) in time interval 1, thereby confirming the temporal contingency between the critical pictures in time interval 1. Next, we created model 4c, in which we added the same constraint to time interval 2. Thus, now the probability of an English–Dutch switch was constrained to be equal to the probability of an English–distractor switch for the two time intervals. Comparison of models 4b (with restrictions on time interval 1) and 4c (with restrictions on both time intervals) would test the hypothesis that no evidence for co-activation could be found in the second time interval. If model 4c would not have a worse fit to the data than model 4b, this hypothesis would be confirmed. Indeed, the models did not have a significantly different fit to the data, $-2LL_{diff}(1) = 0.01$, $p > .90$, which showed that the critical switching probabilities in interval 2 were not significantly different. The transition matrix for the best time model (4a) showed that, on time interval 2, the probability of a transition from the English target to a Dutch competitor (.014) was equal to that of a transition from the English target to a distractor (.014). This shows that there are no temporal dependencies between the critical pictures in time interval 2, as hypothesized.

These results show that the time course within a trial can be divided into small time frames – as small as the researcher wishes to examine and the sampling rate of the eye tracking device allows¹¹ – thereby providing a more detailed look in the temporal aspect of language processing.

General discussion

In the present study, we have described and demonstrated a method for assessing transitions (first-order shifts in overt attention) in an eye tracking paradigm. The main idea motivating this approach is that by analyzing transitions between relevant visual referents one can investigate the co-activation of representations in the interactive systems that are assumed to underlie viewing behavior. The crucial point is that, in an interactive system, the activation of one representation is temporally contingent with the activation of another representation. If fixations to a relevant visual referent are indicative of the activation of representations, then a fine-grained and sensitive way to assess cognitive interactions is to search for temporal dependencies in eye movements. Furthermore, transitions enable the detection of co-activation where other methods might either miss (e.g. Table 1B) or spuriously find (e.g. Table 1A) signs of cognitive interactions.

¹¹ The more nominal time intervals one uses, the more this may lead to anticonservative inferences about parameters. In this case it may be useful to estimate continuous models for time according to a certain function (e.g., an \ln function). This way, it is not necessary to define relevant time intervals a priori and one can assess the decay of the effects over time, for example. Because we tested a priori hypotheses about time to validate our models, we assessed predefined time intervals.

However, when using the current approach one needs to consider a number of important factors. Although transition analyses are extremely helpful to detect cognitive interactions, there are, as is the case for any method, circumstances under which they might fail to detect co-activation. For example, it is possible that co-activation is too transient or dispersed to be detected through first-order transitions. In the following section, we therefore discuss (a) conditions under which direct transitions may or may not be expected, and (b) possible solutions and applications in case direct transitions may not occur. We also discuss other methodological issues that should be taken into account when using transition analyses. Finally, we will embed this approach within a broader context by discussing applications in other theoretical frameworks and paradigms.

Future development and considerations

Even though the analysis of transitions provides a more fine-grained window into interactions between representational units compared to the analysis of proportions or delays, the critical assumption that co-activation will result in temporal dependencies in eye-movements may not hold for every type of experimental paradigm. What if there are *no* direct transitions between critical fixations? It may be either the case that there are no interactions, or that there are interactions but they are too transient or dispersed to be detected.

Importantly, the following methodological assumptions have to hold when detecting interactions between representations using transition analyses in eye-movements. If these assumptions are not met, interactions might not be detected. First, the duration of the hypothesized interaction between representations has to be sufficiently long to result in overt eye movements. Second, the relevant representations have to be both sufficiently activated: if one representation is much more strongly activated than the other this will likely only result in fixations to the strongest referent. Third, the time to program and execute a saccade should not be longer than the temporal shifts in relative activation of both representations. Fourth, the sequential fixations have to be long enough to be detected by the eye-tracking method, but short enough not to mask the underlying hypothesized interaction. Fifth, the rate of new bottom-up input has to be slow enough for the hypothesized interactions to drive the eye movements.¹² Of course, these assumptions (and perhaps even more) implicitly underlie any eye tracking paradigm, and are therefore also relevant for the proposed transition analyses.

Furthermore, we assessed direct switches under the theoretical assumption that there were no additional processes (e.g., suppression of lexically or semantically related competitors) and effects (e.g., word frequency) that would temporarily interfere with co-activation. Any such interference might have dispersed the relative activation of targets and competitors in time, which would not have led to

direct switches between the target and competitor. For example, when instructed in English to manipulate a picture of “mice”, a bilingual might first activate lexically related words in English (such as the rhyme “rice”) before the Dutch counterpart “mais” is sufficiently activated to direct eye movements (see e.g., Marian, Blumenfeld, & Boukrina, 2007, who demonstrate early effects of such lexical relations in the eye movements of bilinguals). In this case, cross-language activation might not be manifested in direct transitions between the target and competitor pictures. This clearly was not the case in the current design in which we observed direct switches between the critical pictures. However, a transition analysis is specifically sensitive to such interfering processes because they give a more fine-grained index of temporal dependencies over time compared to the discussed traditional methods.

In general, there are two strategies that researchers may want to use in order to deal with these issues. First, like in any design, potentially interfering factors can be controlled for as much as possible. Alternatively, such interfering processes can be specifically tested with transition analyses, by hypothesizing about higher-order transitions (indirect switches) rather than first-order transitions (direct switches). For example, inclusion of an additional competitor depiction (such as the rhyme competitor “rice”) could explicitly show whether bilinguals fixate this competitor in between fixations to the English target and Dutch competitor (EN-rhyme-DU). This directly shows a great advantage of the current method, namely that transitions can serve as a proxy for the strength of connections between representations. This way, they might be able to expose hierarchical features of the representational system that underlies viewing behavior.

Several previous eye tracking studies were designed for this specific purpose. For example, Allopenna et al. (1998) designed a visual world experiment in which participants were instructed to manipulate a target object (e.g., a beaker), in the presence of a cohort competitor with initial word overlap (e.g., a beetle), and/or a rhyme competitor with final word overlap (e.g., a speaker), and an unrelated distractor (e.g., a carriage). Analyses on the fixation proportions per time window showed that participants activated the cohort competitor prior to the rhyme competitor. This supported the notion that the speech signal is continuously mapped onto potential representations as it unfolds over time, and it thereby revealed features of the hierarchy in which representations are accessed during language processing. In a comparable fashion, Huettig and McQueen (2007) used proportion analyses to demonstrate a different time course for the activation of phonological (e.g., beaver), shape (e.g., a bobbin), and semantic (e.g., a fork) competitors, respectively, when manipulating a target item (e.g., beaker). These results were taken as evidence for a cascaded organization of people’s lexical and visual knowledge systems.

Although these studies are indicative of a hierarchy in which representations are activated, they are sensitive to the errors we discussed in the introduction because they cannot assess temporal dependencies within trials. Higher-order transitions, on the other hand, could give a detailed indication of the sequential order in which

¹² We thank one of the reviewers of this paper for pointing out these considerations.

representations are activated within participants and within trials. In fact, higher-order transitions may even show whether processes mask co-activation in the absence of referents of the interfering process. Consider a situation in which a rhyme effect temporarily interferes with co-activation (i.e., activation of *mice-rice-mais*), but in which the visual world paradigm only consisted of depictions of the English *mice*, the Dutch *mais*, and two unrelated distractors (not of *rice*). In this case, one could use higher-order transitions to test whether EN-distractor-DU sequences were more likely to occur than EN-distractor-other sequences. This would show that the English and Dutch fixations were slightly separated in time but still contingent within a trial.¹³ To what extent a separation of critical fixations could still hold as co-activation is up for discussion, but even at the coarsest level our argument still holds: Assessing temporal dependencies of fixations within trials gives a better proxy of temporal dependencies in the activation of representations than assessing aggregated fixation data.

Despite these higher-order solutions, it is difficult to test or control for all possible (potentially unknown) factors that might temporally interfere with co-activation and/or to make a priori hypotheses about the possible sequences. Further exploration of this transition paradigm is necessary to examine its possible sensitivity for false-negatives in more detail. Here, one first step could be to use transition analyses in conjunction with the more traditional (multilevel) proportion analyses. Any correspondence between the two would bolster the case for or against co-activation in an interactive cognitive system. Also, if one observes a discrepancy between the analyses such as in Table 1B, this would show that the temporal co-activation did occur, but that it was missed by proportions that are less sensitive to temporal tests of interactivity. Finally, if one observes the situation in Table 1A, this could either imply that the proportion analysis was too liberal or that the transition analysis was too conservative, for example due to interfering processes. This could then be further examined, for example by assessing higher-order transitions as explained above.

Also, some practical considerations should be taken into account when using the proposed transition paradigm. First, one should be careful when specifying the time frames for assessing transitions. In the present example, time frames were set at 100 ms. However, the sampling rate of an eye tracker allows researchers to use smaller time frames (of up to 1 ms, depending on the eye tracking device). The reason for selecting larger time frames in our exemplary data lies in the relation between staying and switching probabilities. The time frames that are included in a markov model should reflect transition opportunities, i.e., the case in which it would have been possible to make an eye movement. The constraints of the human oculomotor system are such that individuals cannot make a

saccade to a different picture every few milliseconds; it takes a minimum of 100 ms to plan and start an eye movement (Altmann, 2011) and an average of about 200–300 ms to plan and fully execute an eye movement until it lands on the target location. Therefore, if very small time frames had been used in the analyses (e.g., much smaller than 100 ms), the results would have shown relatively higher staying probabilities and lower switching probabilities. In other words, when using smaller time frames, the staying probabilities would have been inflated relative to switching probabilities, possibly pushing switching probabilities toward floor. On the other hand, if larger time frames (e.g., well over 100 ms) had been analyzed, transitions could be missed given that more than one transition might have occurred within a single time frame. Loss of transition information inherently results in a loss of power. By selecting time frames of 100 ms, we have attempted to find a balance between keeping the switching probabilities above floor without compromising the power of the measure. However, the time frames could also have been set at somewhat smaller or larger sizes (of e.g., 50 ms or 150 ms). Exactly what size of time frames is optimal when assessing transitions is open to examination.¹⁴

One should further be alert for possible anticipatory effects due to context (cf., Barr, 2008; Barr, Gann, & Pierce, 2011; Dahan & Tanenhaus, 2005; Dahan, Tanenhaus, & Salverda, 2007; McMurray, Tanenhaus, & Aslin, 2009; Wolter, Gorman, & Tanenhaus, 2011). One of the primary uses of visual world studies is to investigate issues related to information integration, in which constraining information is presented prior to presentation of the target word (see Barr, 2008). If, for example, the preceding context in our exemplary study had biased towards the Dutch depiction, the probability of fixating these critical pictures would have been elevated. Researchers should handle item selection with great care and make sure that no constraining context is presented prior to the target word. Fortunately, our proposed method provides an effective way to handle possible anticipatory effects. For example, by estimating a model over a baseline period before presentation of the target word, one can assess how transition probabilities after target presentation change with respect to this baseline. If the experimental design does not allow for such a baseline period, one might use transition probabilities from a control condition with identical visual displays as a baseline for transition probabilities in the critical condition (see Barr et al., 2011 for a discussion on handling anticipatory baseline effects).

In sum, there are several considerations to take into account when assessing transitions in eye movements. However, the proposed transition analysis has important advantages over analyses on proportions or delays, and can provide critical novel insights in viewing behavior

¹³ If one wants to use transition probabilities as a proxy for the strength of activation, one should add to the model the possibility that participants did not fixate any picture (e.g., coded as picture 0). In this case, the probabilities of $s=0, 1, 2, 3, 4$ in the formulas in Equations of the multilevel markov model should sum to one.

¹⁴ An alternative to this time-based approach is to select an event-based approach, in which the staying probability is treated as the probability of making the next fixation on the same picture, and the switching probability is treated as the probability of making the next fixation on a different picture. In this case, the transition probabilities are calculated on event boundaries, rather than on time boundaries. However, in such an approach it remains a matter of definition as to what would count as an event (fixation or saccade) and what would not, which is open for discussion.

and the human representational system that underlies this behavior. The most important advantage is that the transition approach is able to reveal temporal dependencies in viewing behavior, because it does not require aggregation of data over participants, trials, or time. In our current application, we used the transition method to demonstrate that the activation of certain representations was co-dependent on the activation of other representations. However, transitions allow for a wide range of further novel and interesting assessments of viewing behavior. Within the visual world paradigm, for example, one may use transitions as a proxy for the relative distance or strength of connections between representations, as explained. Or one may want to use these analyses to assess other temporal aspects of viewing behavior, such as the duration of fixations before or after a critical switch was made. One could, for example, assess whether post-switch staying probabilities change as a result of a manipulation or over time, because this might reveal different processes that evolve from temporal dependencies. However, transition analyses are not limited to visual world studies. In the next section, we will provide potential applications of the current analyses in different paradigms.

Applications in other paradigms

Even though we have illustrated the transition approach by assessing lexical competition in a visual world paradigm, its scope is much broader. This approach can be used in any eye tracking paradigm in which shifts in attention within a trial are of interest, such as scene perception, visual search, change blindness studies, or reading. The majority of studies on sentence and text reading, for example, often analyze averaged measures of forward or backward fixations as an index for comprehension, learning, and/or memory processes. For example, the proportion of regressions to previously fixated regions during sentence reading has been used as a measure for disrupted language processing, in which a higher proportion of regressions indicates greater processing difficulties (e.g., Frazier & Rayner, 1982). Furthermore, averaged regression durations on previous or following (parts of) sentences have been used for this purpose, with longer durations indicating greater processing difficulties (e.g., Just & Carpenter, 1978). Instead of assessing averaged regression durations or proportions across participants and trials, our current approach can provide estimations on the regression (i.e., transition) probabilities at an individual trial level. This enables researchers to distinguish different types of regression patterns across trials and/or individuals. By performing latent class markov analyses, one could, for example, distinguish between latent regression patterns that are typical for skilled versus poor readers, or even unravel transition patterns between reading strategies within participants (see Simola et al., 2008).

In general, readers are more likely to move forward in a sentence or text than to regress backward. Therefore, the successive locations of fixations (states) are not mutually independent. When such state dependencies are expected to occur in an eye tracking paradigm, predictions about viewing patterns should be formulated as a chain of

transitions to different parts of the sentence or text, and should therefore include higher-order effects. With some minor adaptations to the syntax (see the supplemental material, doi, section B), higher-order hypotheses can easily be tested using the present markov modeling technique (see Althoff & Cohen, 1999, for an application of assessing second-order effects, or Simola et al., 2008).

There are previous applications of similar approaches in several fields. For example, Simola et al. (2008) distinguished three latent states (classes) of viewing behavior in a reading paradigm, referred to as scanning, reading, and decision making. In this study, each participant performed three types of information search tasks on a list of titles: simple word search, a question–answer task, and a task in which they had to select the subjectively most interesting topic. The authors found that the different tasks were characterized by different transition patterns between states, suggesting that readers switch between processing strategies (states), and that they do this in a different order depending on the information search task at hand. Other researchers have used this approach to distinguish local (reflecting the extraction of detailed information) from global (reflecting the redirection of attention) states of visual attention when people study advertisements in magazines (Liechty, Pieters, & Wedel, 2003). Furthermore, Althoff and Cohen (1999) and Henderson et al. (2000) have used markov model analyses to study sequences of fixations (referred to as “scan patterns”) that accompany face perception and recognition. Both studies found dependencies in successive fixation positions when perceiving an unknown face. For example, when fixating one of the two eye locations of the face, it was more likely that the successive fixation was directed to the other eye than to any other location in the face (Henderson et al., 2000).

These examples not only show that transition patterns can be used in different eye tracking paradigms, but also within different theoretical frameworks.¹⁵ One could envisage numerous types of research domains that would benefit from assessing transitions in a fixation pattern, such as studies assessing attentional biases, approach–avoidance mechanisms, object permanence in young children, or social interactions.

Conclusion

Whenever the aim is to study interactive activation of representations in a cognitive system, it is not sufficient to conclude that the visual referents of both systems were fixated across trials and participants. Rather, assessing whether a fixation to the referent of one representation is temporally dependent on a fixation to the referent of the other representation can provide a good indication of interactivity. Analyzing eye movement transitions using a multilevel markov approach is an ideal way to tap into such temporal dependencies. In this paper, we have

¹⁵ The multilevel markov modeling approach is not even restricted to eye tracking data. It could, for example, be used to assess transitions in any type of responses (e.g., behavioral or ERP) on different trials within a block, different blocks within a task, different tasks within a participant, et cetera.

demonstrated that analyzing transitions cannot only provide useful, but also essential, novel insights in human viewing behavior and its underlying cognitive processes.

Acknowledgments

We would like to thank Dale Barr, Erik Reichle, Florian Jaeger, and Viorica Marian, as well as several anonymous reviewers, for their valuable comments on earlier versions of the manuscript. We further thank the members of the Language and Cognition Group and the Memory Lab at the Erasmus University Rotterdam for their feedback. Finally, we thank Mark Adriaans, Marcel Boom, and Christiaan Tieman of the Erasmus Behavioral Lab for their help with the data.

A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.jml.2013.05.006>.

References

- Agresti, A. (2002). *Categorical data analysis* (2nd ed.). New York: Wiley.
- Alloppenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language*, 38, 419–439.
- Althoff, R. R., & Cohen, N. J. (1999). Eye-movement-based memory effect: A reprocessing effect in face perception. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25, 997–1010.
- Altmann, G. (2011). Language can mediate eye movement control within 100 milliseconds, regardless of whether there is anything to move the eyes to. *Acta Psychologica*, 137, 190–200.
- 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.
- Barr, D. J. (2008). Analyzing 'visual world' eyetracking data using multilevel logistic regression. *Journal of Memory and Language*, 59, 457–474.
- Barr, D. J., Gann, T. M., & Pierce, R. S. (2011). Anticipatory baseline effects and information integration in visual world studies. *Acta Psychologica*, 137, 201–207.
- Bogartz, R. S., & Staub, A. (2012). Gaze step distributions reflect fixations and saccades: A comment on Stephen and Mirman (2010). *Cognition*, 123, 325–334.
- Brysbaert, M., Van Dyck, G., & Van de Poel, M. (1999). Visual word recognition in bilinguals: Evidence from masked phonological priming. *Journal of Experimental Psychology: Human Perception and Performance*, 25, 137–148.
- Caramazza, A., & Brones, I. (1979). Lexical access in bilinguals. *Bulletin of the Psychonomic Society*, 13, 212–214.
- Cooper, R. M. (1974). The control of eye fixation by the meaning of spoken language: A new methodology for the real-time investigation of speech perception, memory, and language processing. *Cognitive Psychology*, 6, 84–107.
- Cristino, F., Mathot, S., Theeuwes, J., & Gilchrist, I. D. (2010). ScanMatch: A novel method for comparing fixation sequences. *Behaviour Research Methods*, 42, 692–700.
- Dahan, D., & Tanenhaus, M. K. (2005). Looking at the rope when looking for the snake: Conceptually mediated eye movements during spoken-word recognition. *Psychological Bulletin & Review*, 12, 453–459.
- Dahan, D., Tanenhaus, M. K., & Salverda, A. P. (2007). The influence of visual processing on phonetically driven saccades in the "visual world" paradigm. In R. P. G. Van Gompel, M. H. Fischer, W. S. Murray, & R. L. Hill (Eds.), *Eye movements: A window on mind and brain* (pp. 471–486). Oxford, England: Elsevier.
- Dijkstra, A., & Van Heuven, W. J. B. (2002). The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and Cognition*, 5, 175–197.
- Findlay, J. M. (2004). Eye scanning and visual search. In J. M. Henderson & F. Ferreira (Eds.), *The interface of language, vision, and action: Eye movements and the visual world* (pp. 135–159). New York: Psychology Press.
- Frank, A., Salverda, A. P., Jaeger, T. F., & Tanenhaus, M. (2009). Analyzing multinomial timeseries data with "state" dependencies. In *The 22nd CUNY sentence processing conference*, Davis, CA.
- Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology*, 14, 178–210.
- Grainger, J., & Beauvillain, C. (1987). Accessing interlexical homographs: Some limitations of a language-selective access. *Journal of Memory and Language*, 26, 658–672.
- Hallet, P. E. (1986). Eye movements. In K. Boff, L. Kaufman, & J. Thomas (Eds.), *Handbook of perception and human performance* (Vol. 1). New York: Wiley-Interscience (pp. 10–11, 10–112).
- Hartsuiker, R. J., Huettig, F., & Olivers, C. N. (2011). Visual search and visual world: Interactions among visual attention, language, and working memory (introduction to the special issue). *Acta Psychologica*, 137, 123–137.
- Henderson, J. M., Falk, R., Minut, S., Dyer, F. C., & Mahadevan, S. (2000). Gaze control for face learning and recognition by humans and machines. *Michigan State University Eye Movement Laboratory Technical Report*, 4, 1–14.
- Henderson, J. M., & Ferreira, F. (2004). Scene perception for psycholinguists. In J. M. Henderson & F. Ferreira (Eds.), *The interface of language, vision, and action: Eye movements and the visual world* (pp. 1–58). New York: Psychology Press.
- Huettig, F., & Altmann, G. T. M. (2005). Word meaning and the control of eye fixation: Semantic competitor effects and the visual world paradigm. *Cognition*, 96, B23–B32.
- Huettig, F., & McQueen, J. M. (2007). The tug of war between phonological, semantic, and shape information in language-mediated visual search. *Journal of Memory and Language*, 57, 460–482.
- Hwang, A., Wang, H., & Pomplun, M. (2011). Semantic guidance of eye movements in real-world scenes. *Vision Research*, 51, 1192–1205.
- Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, 59, 434–446.
- Just, M. A., & Carpenter, P. A. (1978). Inference processes during reading: Reflections from eye fixations. In J. M. Senders, D. F. Fisher, & R. A. Monty (Eds.), *Eye movements and the higher psychological functions*. New Jersey: Lawrence Erlbaum.
- Liechty, J., Pieters, R., & Wedel, M. (2003). Global and local covert visual attention: Evidence from a Bayesian Hidden Markov Model. *Psychometrika*, 68, 519–541.
- Macnamara, J., & Kushnir, S. L. (1971). Linguistic independence of bilinguals: The input switch. *Journal of Verbal Learning and Verbal Behavior*, 10, 480–487.
- Magnuson, J. S., Dixon, J. A., Tanenhaus, M. K., & Aslin, R. N. (2007). The dynamics of lexical competition during spoken word recognition. *Cognitive Science*, 31, 1–24.
- Marian, V., Bartolotti, J., Chabal, S., & Shook, A. (2012). CLEARPOND: Cross-linguistic easy-access resource for phonological and orthographic neighborhood densities. *PLoS ONE*, 7, e43230. <http://dx.doi.org/10.1371/journal.pone.0043230>.
- Marian, V., Blumenfeld, H. K., & Boukrina, O. V. (2007). Sensitivity to phonological similarity within and across languages. *Journal of Psycholinguistic Research*, 37.
- Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech, Language, and Hearing Research*, 50, 940–967.
- Marian, V., & Spivey, M. J. (2003). Competing activation in bilingual language processing: Within- and between- language competition. *Bilingualism: Language and Cognition*, 6, 97–115.
- Marian, V., Spivey, M. J., & Hirsch, J. (2003). Shared and separate systems in bilingual language processing: Converging evidence from eye tracking and brain imaging. *Brain and Language*, 86, 70–82.
- Marslen-Wilson, W. D., & Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology*, 10, 29–63.
- Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition*, 25, 71–102.
- Matin, E., Shao, K. C., & Boff, K. R. (1993). Saccadic overhead: Information-processing time with and without saccades. *Perception & Psychophysics*, 53, 372–380.
- McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, 18, 1–86.

- McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: I. An account of basic findings. *Psychological Review*, 88, 375.
- McMurray, B., Tanenhaus, M. K., & Aslin, R. N. (2009). Within-category VOT affects recovery from "lexical" garden-paths: Evidence against phoneme-level inhibition. *Journal of Memory and Language*, 60, 65–91.
- Midgley, K. J., Holcomb, P. J., & Grainger, J. (2011). Language learners: An ERP investigation. *Journal of Cognitive Neuroscience*, 23, 1634–1647.
- Mirman, D., Dixon, J. A., & Magnuson, J. S. (2008). Statistical and computational models of the visual world paradigm: Growth curves and individual differences. *Journal of Memory and Language*, 59, 475–494.
- Rumelhart, D. E., & McClelland, J. L. (1982). An interactive activation model of context effects in letter perception: II. The contextual enhancement effect and some tests and extensions of the model. *Psychological Review*, 89, 60.
- Schulpen, B., Dijkstra, T., Schriefers, H. J., & Hasper, M. (2003). Recognition of interlingual homophones in bilingual auditory word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, 29, 1155–1178.
- Simola, J., Salojärvi, J., & Kojo, I. (2008). Using hidden Markov model to uncover processing states from eye movements in information search tasks. *Cognitive Systems Research*, 9, 237–251.
- Spivey, M. J., & Marian, V. (1999). Cross talk between native and second languages: Partial activation of an irrelevant lexicon. *Psychological Science*, 10, 281–284.
- Spivey-Knowlton, M. J., Tanenhaus, M. K., Eberhard, K. M., & Sedivy, J. C. (1998). Integration of visuospatial and linguistic information: Language-comprehension in real-time and real-space. In P. Olivier & K. Gapp (Eds.), *Representation and processing of spatial expressions* (pp. 201–214). Hillsdale, NJ: Erlbaum.
- Stephen, D. G., & Mirman, D. (2010). Interactions dominate the dynamics of visual cognition. *Cognition*, 115, 154–165.
- Tanenhaus, M., Frank, A., Jaeger, T. F., Masharov, M., & Salverda, A. P. (2008). The art of the state: Mixed effect regression modeling in the Visual World. In *The 21st CUNY sentence processing conference*, Chapel Hill, NC.
- Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, 268, 1632–1634.
- Theeuwes, J. (2010). Top-down and bottom-up control of visual selection. *Acta Psychologica*, 135, 77–99.
- Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, 12, 97–136.
- Vandeberg, L., Guadalupe, T., & Zwaan, R. A. (2011). How verbs can activate things: Cross-language activation across word classes. *Acta Psychologica*, 138, 68–73. <http://dx.doi.org/10.1016/j.actpsy.2011.05.007>.
- Vermunt, J. K., & Magidson, J. (2005). *Technical guide for latent GOLD 4.0: Basic and advanced*. Belmont, MA: Statistical Innovations Inc.
- Vermunt, J. K., & Magidson, J. (2008). *LG-Syntax user's guide: Manual for latent GOLD 4.5 syntax module*. Belmont, MA: Statistical Innovations Inc.
- Von der Malsburg, T., & Vasishth, S. (2011). What is the scanpath signature of syntactic reanalysis? *Journal of Memory and Language*, 65, 109–127.
- von Studnitz, R. E., & Green, D. W. (2002). Interlingual homograph interference in German–English bilingual: Its modulation and locus of control. *Bilingualism: Language and Cognition*, 5, 1–23.
- Williams, L. G. (1966). The effect of target specification on objects fixated during visual search. *Perception and Psychophysics*, 1, 315–318.
- Wolfe, J. M. (1994). Guided search 2.0. A revised model of visual search. *Psychonomic Bulletin & Review*, 1, 202–238.
- Wolter, L., Gorman, K. S., & Tanenhaus, M. K. (2011). Scalar reference, contrast and discourse: Separating effects of linguistic discourse from availability of the referent. *Journal of Memory and Language*, 65, 299–317.
- Yee, E., & Sedivy, J. C. (2006). Eye movements to pictures reveal transient semantic activation during spoken word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32, 1–14.