Title
Verb-action versus role relations congruence effects: Evidence from ERPs in picture-sentence verification

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 32(32)

ISSN
1069-7977

Authors
Knoeferle, Pia
Urbach, Thomas P.
Kuas, Marta

Publication Date
2010

Peer reviewed

Verb-action versus role relations congruence effects: Evidence from ERPs in picture-sentence verification

Pia Knoeferle (knoeferl@cit-ec.uni-bielefeld.de)
Cognitive Interaction Technology Excellence Cluster
Bielefeld University, 33615 Bielefeld, Germany

Thomas P. Urbach (turbach@cogsci.ucsd.edu)
Department of Cognitive Science
University of California San Diego La Jolla, USA

Marta Kutas (kutas@cogsci.ucsd.edu)
Department of Cognitive Science
University of California San Diego La Jolla, USA

Abstract

Comprehenders can rapidly use both their linguistic knowledge and different kinds of information in visual context during language comprehension. Little is known, however, about the relative time courses and mechanisms by which different kinds of visual information influence language comprehension. We recorded event-related brain potentials (ERPs) as participants read a subject-verb-object sentence and verified whether or not it matched different (verb-action versus role relations) aspects of a recently viewed picture. When the verb-action did not match the depicted action, we replicated larger N400s (300-500ms) over centro-parietal scalp to the verb (300-500 ms) relative to the responses for matches. In contrast, ERP effects to role-relation mismatches (a person depicted as undergoing an action but described as performing it) qualitatively differed from and occurred prior to the verb-action congruence N400. Our findings implicate at least two temporally distinct mechanisms governing picture-sentence verification processes.

Keywords: sentence-picture verification; visual context effects; event-related brain potentials;

Introduction

Information in visual context can rapidly influence online language comprehension and ambiguity resolution (e.g., Altmann, 2004; Knoeferle, Habets, Crocker, & Münte, 2008; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). Recently, researchers also have begun to examine picture-sentence congruence processes even when sentences are structurally unambiguous and do not necessarily globally match visual context (e.g., Knoeferle, Urbach, & Kutas, 2009; Vissers, Kolk, van de Meerendonk, & Chwilla, 2008; Wassenaar & Hagoort, 2007). The motivation for these studies is that determining the correspondence between what is said and how things are, appears to play a central part in natural language processing: Positive verification may be readily inferred from, e.g., expressions of agreement (*So I heard*) while failures to verify may be inferred from corrections and expressions of disbelief, requests for clarification and the like (e.g., *Well no, actually what happened was ...*, *Are you sure?*).

Psycholinguistic research on verification processes is by no means a recent endeavour: Just and Carpenter (1971) found that participants' verification latencies were shorter when the color of the dots on an image (red vs. black) matched than mismatched the color adjective in a corresponding sentence (henceforth "congruence effects", see also, e.g., Clark & Chase, 1972). To account for these findings and a range of others, Carpenter and Just (1975) developed the Constituent Comparison Model (CCM) of sentence-picture comparison processes. The model operates via a serial mechanism that incrementally compares representations of sentence ([AFF, (RED, DOTS)]) and picture ("black dots") constituents. The comparison proceeds from inner to outer representations (in this case right to left). When a mismatch is found (here for the inner representations), it is indexed "-"; the truth value is changed to "false", and the comparison process is reinitialized, resulting in one extra comparison step (and hence longer response times) relative to a match (e.g., "red dots"). The output of that comparator process is the truth value of the comparison and response time values.

This verification model does not specify the time course of constructing the representations of verbal information as the sentences are read, and it is unclear to which extent the constituent-wise comparator mechanism implies incremental comprehension processes. Recent event-related brain potential (ERP) research, however, suggests that congruence processing is incremental, i.e., ongoing during (not merely after) word-by-word sentence processing, and furthermore can be systematically related to end-of-sentence verification times (Knoeferle et al., 2009). This was evidenced by finding (a) reliable congruence effects in ERPs as soon as a word (e.g., the verb) that mismatched aspects of a preceding visual context (e.g., a depicted action) was encountered; (b) reliable congruence effects in verification time response latencies; and (c) reliable correlations between a participant's ERP and verification time congruence effects.

The incremental ERP congruence effects at the verb are, in principle, compatible with the CCM comparator mechanism. One may question, however, to what extent the CCM can account for verification processes during (rather than after) the sentence. Specifically, it is unclear whether all aspects of picture-sentence mismatch processing are adequately accounted for by a single comparator mechanism as suggested by the CCM.

Findings from two recent studies can be viewed as supporting a single mechanism account: highly similar ERP congruence effects were observed in response to different picture-sentence mismatches (Vissers et al., 2008; Wassenaar & Hagoort, 2007). In Wassenaar and Hagoort (2007), healthy older adults inspected a line drawing of an agent-action-patient event (e.g., man pushing woman vs. a woman pushing a man), and then listened to a spoken utterance in Dutch (e.g,. 'The tall man on [sic] this picture pushes the young woman'). Participants indicated whether the thematic relations of the utterance were congruent with the depicted role relations or not. There was no reliable response time congruence effect[1]. In the ERPs, however, healthy adults exhibited congruence effects. For active sentences as in the above example, there was a larger posterior negativity (with a non-reliable late positivity) to mismatching than matching conditions in the verb region (centro-posterior from 50-450 ms; for anterior sites from ca. 50-300 ms). Irreversible active and reversible passive sentences showed an early negativity for incongruous relative to congruous trials and a subsequent (reliable) late positivity. These effects were interpreted as reflecting thematic role assignment.

In a different study, participants inspected a line drawing containing two objects (e.g., a square followed by a circle, Vissers et al., 2008). They then read a sentence via rapid serial visual presentation (e.g., *De cirkel staat achter het vierkant*, 'The circle stands in front of the square'), and verified whether or not the object arrangement described in the sentence matched the depiction. For a first condition the location mismatch occurred within the same (horizontal) dimension (e.g., the sentence would state that the circle was in front of the square while it was in fact behind it). For a second, mismatch, the incongruence occurred between the horizontal and vertical dimensions: The sentence stated that the circle is below the square while it was in fact behind it. The authors observed an N400-P600 ERP pattern as in Wassenaar and Hagoort (2007) despite differences in the mismatches (object location rather than role relations) and language modality (written versus spoken). They interpreted the mismatch effects as reflecting monitoring of potential processing errors. Cruci ally, mean amplitudes of the ERPs in Vissers et al. did not differ in response to the two kinds of picture-sentence mismatches (200-400 ms; 500-700 ms time-locked to the critical preposition).

Based on these findings, it appears that some picture-sentence mismatches (e.g., role relations versus object location mismatches) elicit similar ERP patterns, providing - at least tentative - support for a single functional brain mechanism dealing with these incongruences (though note the different interpretations of the ERP pattern in these two studies).

In contrast with the Vissers et al. and Wassenaar and Hagoort findings, tentative support for the alternative - multiple mechanism - view comes from Knoeferle et al. (2009) in which participants read a subject-verb-object sentence and verified whether or not the verb matched a previously viewed (depicted) action. When verbs mismatched a depicted action, speeded verification response latencies were reliably longer, N400s over centro-parietal scalp to the verb were larger, and post-verbal potentials up to the time of the response (including an anterior negativity to the object) were more negative relative to the responses for matches. These different negativities across the sentence differ from the ERP congruence effects in response to role relations and object location mismatches per the absence of an ensuing P600-like congruence effect. In either case, however, our knowledge of the *relative* time course and nature of different visual context effects during sentence comprehension is relatively limited and only few studies have directly compared different visual context effects.

The present research further investigates the nature and time course of picture-sentence verification processes by directly comparing visual context effects that require interpreting a written verb in relation to an action with effects that involve interpretation of sentential role relations in relation to depicted role relations. In two Experiments, we analyzed ERPs as participants read a subject-verb-object sentence and verified whether or not the sentence matched a recently viewed visual scene. The verb either matched the previously depicted action or not; and who-does-what-to-whom in the sentence was either congruous with the depicted role relations or not, resulting in 4 (fully counterbalanced) conditions (see Table 1).

If there is a single mechanism for congruence processing we would expect to see similar ERP patterns to role relations and verb-action mismatches. Alternatively, the action and role-relation mismatches involve different mechanisms. Processing a role-relations mismatch involves comparing depicted agents and patients with a compositional interpretation, perhaps requiring more time and processing effort than relating an action to a verb interpretation. Recall, that prior research observed a verb-action congruence N400 effect (Knoeferle et al., 2009). Assuming a larger negativity reflects greater processing difficulty (see Monetta, Tremblay, & Joanette, 2003), and given that Wassenaar and Hagoort (2007) observe their first congruence effects at the verb, such an account predicts greater negative mean ERP amplitudes during the N400 region at the verb for the role relations than verb-action mismatches (and most negative for the combined mismatches), and also later ERP and verification time congruence effects.

Alternatively, role-relation effects would precede verb-action congruence effects: People likely expect the first noun phrase of a sentence to be the agent. When they read the first noun and realize that it does not refer to the character depicted as the agent, they may begin to anticipate incongruence between picture and sentence even though there is no overt mismatch at the first noun phrase; the moment the verb confirms

[1]The failure to replicate the verification time congruence effect could be due to the fact that the verification response occurred well after sentence end (but essentially this requires further investigation.)

this expectancy (specifying agent-action relationships), role (in)congruence could be confirmed and thus might elicit earlier ERP congruence effects than verb-action mismatches. To better delineate any role relations and verb-action congruence effects, we varied stimulus onset asynchrony between Experiment 1 (500 ms) and Experiment 2 (300 ms). The timing of those congruence effects that depend soley on processing associated with the first noun phrase is not expected to change substantially as a function of SOA. Alternatively, the timing of congruence effects related to information provided by the verb, is expected to vary with the interval between the noun and verb.

Experiments 1 and 2

Methods

Participants Thirty-two students of UCSD took part in Experiment 1, and a further thirty-two participated in Experiment 2. All participants were native English speakers, right-handed (Edinburgh Handedness Inventory), and had normal or corrected-to-normal vision. All gave informed consent; the experiment protocol was approved by the UCSD IRB.

Materials, design, and procedure We derived materials for both experiments from a previous study (Knoeferle et al., 2009). The present two experiments had a 2-factor *role-relation congruence* (congruent, Picture 1a/b vs. incongruent, Picture 1c/d) x *action congruence* (congruent, Picture 1a/c vs. incongruent, Picture 1b/d) within-subjects design (Table 1).

Table 1: Example of the four experimental conditions

Condition	Picture		Sentence
full match	1a		*The gymnast punches the journalist*
action mismatch	1b		*The gymnast punches the journalist*
role mismatch	1c		*The gymnast punches the journalist*
combined mismatch	1d		*The gymnast punches the journalist*

The sentence, *The gymnast punches the journalist*, in Table 1 is congruent on both action and role dimensions with Picture 1a, (full match); it is incongruent on the action but congruent on the role-relation dimension with Picture 1b (action mismatch); it is congruent on the action but incongruent on the role relations dimension with Picture 1c (role mismatch); and it is incongruent on both of these dimension following Picture 1d (combined mismatch). The materials were counterbalanced to ensure that any congruency-based ERP differences were not spuriously due to stimuli or to their presentation. There were 80 item sets which, combined with the conditions and further counterbalancing, yielded 16 experimental lists. Each list contained one occurrence of an item sentence/picture, and an equal number of left-to-right and right-to-left action depictions. Each list also contained 160 filler items, of which half were mismatches. These filler sentences had different syntactic structures including negation, clause-level and noun phrase coordination, as well as locally ambiguous reduced relative clause constructions.

Participants inspected the picture on a CRT monitor for a minimum of 3000 ms terminated via a right thumb button press. Next, a fixation dot was presented for a random duration between 500 and 1000 ms, followed by the sentence, one word at a time. Word onset asynchrony was 500 ms in Experiment 1 and 300 ms in Experiment 2; word duration was 200 ms in both. Participants were instructed to examine the picture and then to read the sentence in the context of the preceding picture. Participants indicated via a button press as quickly and accurately as possible after each sentence whether it matched or did not match the preceding picture. After that button press, there was a randomly varying pause between 500 and 1000 ms prior to the next trial.

Analysis We report analyses of variance (ANOVA) on response latencies and mean amplitude ERPs. Time regions for the ERP analyses were: the first noun; the verb, and the post-verbal object noun. We performed omnibus repeated measures ANOVAs on mean ERP amplitudes (averaged by participants for each condition at each electrode site) with role congruence (mismatch vs. match), action congruence (mismatch vs. match), hemisphere (left vs. right electrodes), laterality (lateral vs. medial), and anteriority (5 levels) as factors. The pre-stimulus baseline for all analyses was 200 ms. Time windows (0-100, 100-300, 300-500) were chosen based on prior studies and visual inspection of waveforms.

Results Experiment 1 (500 ms SOA)

Behavioural results Repeated measures ANOVAS for the verification latencies showed that response times were marginally faster for the action match than mismatch conditions (1115 ms vs. 1163 ms, $p = 0.06$), while there was no reliable effect for the role relations factor ($p > 0.2$); the interaction between these two factors was reliable ($p < 0.01$).

The response latency data replicate findings of a verb-action congruence effect (Knoeferle et al., 2009) as well as the absence of verification time congruence effects for role relations mismatches (Wassenaar & Hagoort, 2007).

ERP results We present grand average ERPs at prefrontal, parietal, temporal, and occipital sites for all four conditions (Fig. 2) and for mean amplitude role mismatches versus matches (Fig. 3).

For the role relations factor, differences emerged early, during the first noun phrase. ERPs for role mismatches were more negative beginning about 200 ms after noun onset (Figure 3), with the effect more pronounced at lateral electrodes over right anterior scalp (100-300 ms, $p < 0.05$). In line

with early (200-400ms) mismatch effects observed by Vissers et al., we also measured ERPs from 200-400 ms at the first noun. Analyses revealed more negative going ERPs to role mismatches than matches ($p < 0.01$). Following the anterior negativity, a relative positivity for mismatches was observed, largest over posterior scalp, beginning around 400ms after noun onset and continuing beyond the onset of the subsequent verb. This effect was reliable from 0-100 ms and 100-300 ms following the verb ($p < 0.01$). These role congruence effects were also reliable when analyzed relative to a pre-noun baseline. They did, however, not last into the later portion of the verb (300-500 ms).

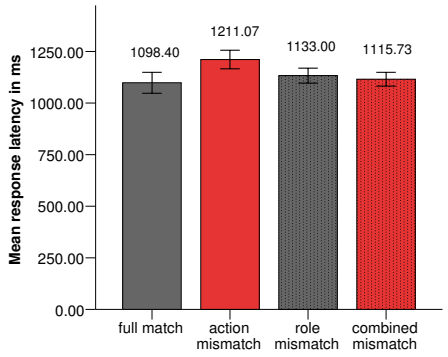

Figure 1: Experiment 1 verification response times (error bars indicate 95% confidence interval)

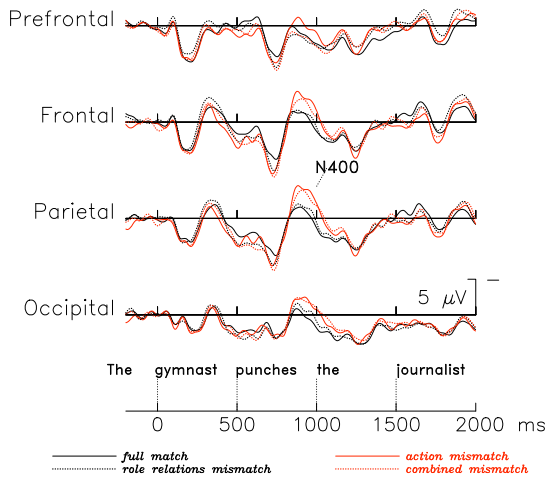

Figure 2: Grand average ERPs (mean amplitude) for all four conditions across the sentence at prefrontal, parietal, temporal, and occipital sites (Experiment 1)

For action mismatches, the first reliable effects occurred at the verb, where we replicated larger mean amplitude ERPs to action mismatches than matches with a a centro-parietal maximum (300-500 ms, $p < 0.001$, see Fig. 2 Knoeferle et al., 2009). The reliable verb-action congruence effect in this window (300-500 ms at the verb and the absence of a

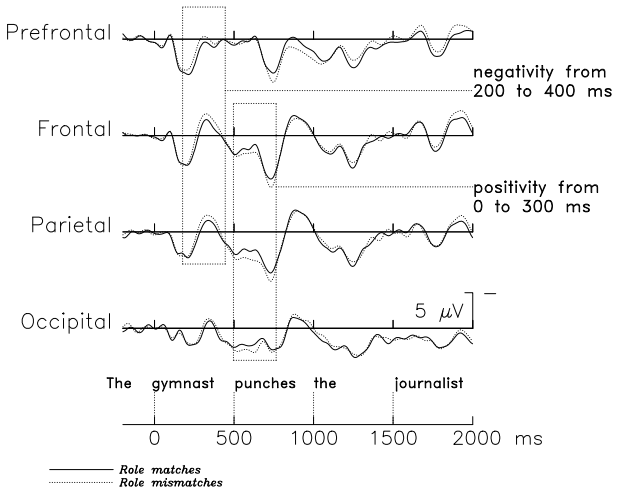

Figure 3: Grand average mean amplitude ERPs for role mismatching conditions versus role matching conditions across the sentence at prefrontal, parietal, temporal, and occipital sites (Experiment 1)

role-relation effect a lead to an interaction between these two factors ($p < 0.05$). During the second noun (300-500 ms), the role mismatches were more negative-going than the role matches ($p < 0.05$).

Results Experiment 2 (300 ms SOA)

Analyses of verification time latencies revealed no reliable effects of the manipulated factors (see Fig. 4).

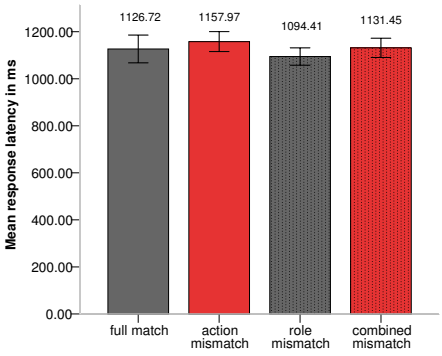

Figure 4: Response latencies in ms for Experiment 2

For the ERPs, the earliest effect of a role mismatch appears to be a broadly distributed relative negativity that reached a maximum between about 300 and 400ms, i.e., shortly after the verb onset (Figure 5). These role congruence effects and occurred from 0-100 and 100-300 ms after verb onset (i.e., 300-600 ms after noun onset).

In these early verb time windows (0-100, 100-300 ms) role mismatches were more negative than role matches ($p < 0.001$, see Fig. 5). That negativity is confirmed when analysing the data re-baselined relative to the first noun (300-500 ms and 200-400 ms ($p < 0.01$). Analysis of the time

window 300-500 ms post-verb found no role mismatch effects ($F < 0.2$). There were no further reliable role relations congruence effects except for more negative ERPs for mismatching than matching trials during the post-verbal object noun (second noun: 400-600 ms, $p < 0.05$).

For the action mismatches, the effects in Experiment 2 appeared after the verb (300-500 ms, $p < 0.001$, see Fig. 6) just as in Experiment1, leading to a reliable interaction of role and action congruence ($ps < 0.01$). Post-verbally, the verb-action congruence negativity continued into the determiner and object noun (see Fig. 6).

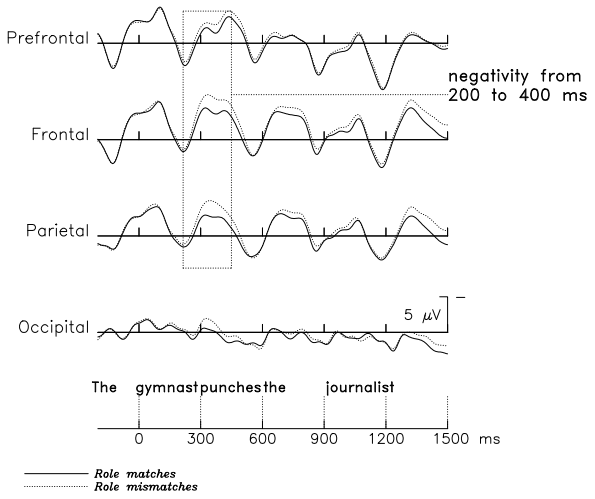

Figure 5: Grand average mean amplitude ERPs scores for role relations mismatches versus matches across the sentence at prefrontal, parietal, temporal, and occipital sites (Experiment 2)

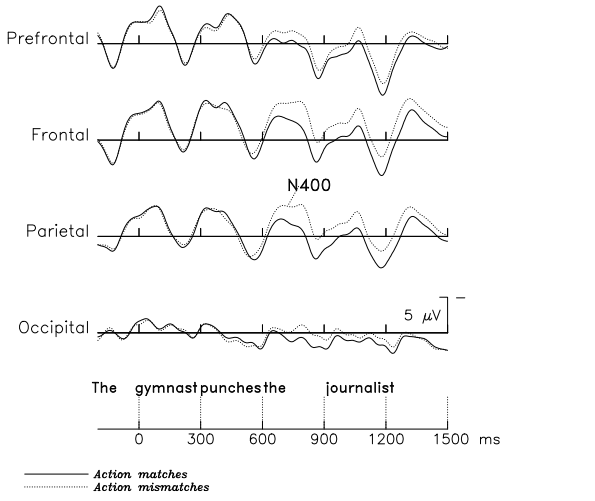

Figure 6: Experiment 2: Grand average mean amplitude ERPs scores for action mismatches versus matches across the sentence at prefrontal, parietal, temporal, and occipital sites (Experiment 2)

In Experiment 2, role mismatch effects again clearly preceded verb-action mismatch effects. Although the role mismatch effect was more broadly distributed and laterally symmetric than the early role congruence negativity in Experiment 1, both effects had the same polarity and a similar time course and neither exhibited the posterior, right lateralized maximum frequently observed for N400 effects.

General Discussion

The present ERP experiments compared role-relation and verb-action congruence processing in a picture-sentence verification task, and examined whether they differed in their natures and/or time courses. Verification time congruence effects for verb-action mismatches (longer response times for action mismatches relative to matches) were replicated (marginal effect) at the longer SOA (Exp 1) and were not reliable at the shorter SOA (Exp 2). By contrast, role match and mismatch response times did not differ at either SOA. ERPs, however, revealed reliable revealed reliable but different effects for both role and action mismatches (*vs* matches)

The earliest role mismatch effects were seen within a few hundred milliseconds of the first noun onset at both SOAs. By contrast, reliable effects of action mismatches were observed only later, a few hundred milliseconds after verb onset. Although the action mismatch effect also was a broadly distributed relative negativity to the mismatches, it tended to be larger over posterior scalp (as is characteristic of visual N400) whereas the role relation mismatch effect was not. At the longer SOA (only) the role relation congruence negativity was followed by a reliable positivity over posterior scalp that continued past the onset of the next word (verb).

As in Knoeferle et al. (2009) we find ongoing ERP congruence effects across the sentence suggest that verification-related processes are part of ongoing incremental sentence interpretation. We observe effects of the action-verb mismatch at the verb, continuing into the second determiner and object noun (see also Ferretti, Singer, & Patterson, 2008; Singer, 2006, for related evidence on text verification). The overall morphology, latency, and centro-parietal distribution of the N400 is similar to that for lexico-semantic anomalies or low cloze probability words in sentences read for comprehension (e.g., Kutas, 1993; Kutas, Van Petten, & Kluender, 2006; Berkum, Hagoort, & Brown, 1999).

Conclusions Our findings are consistent with verification models on which depicted information modulates processing of verbal information as sentences unfold word by word. In the context of a just-viewed depicted action in which a journalist is punching a gymnast, there is nothing incongruous or anomalous about a sentence that begins with *The gymnast* .... People could have waited until they read the verb before assigning a thematic role to the first noun phrase. It seems, however, that when they read the first noun and realized that it referred to a character that had not been depicted as the agent of an action, their expectations of thematic role assignment (i.e., that the first noun in a sentence often refers to the thematic agent) conflicted with their visual context representation (of that character as a patient). Such incongruence may

have led to the larger negativity for role relations mismatches at the first noun. The subsequent centro-parietal positivity elicited by role relations mismatches may be a P600, related to the revision of thematic role assignments though, if so, it is unclear why it did not replicate in Experiment 2.

Furthermore, although action and role relations mismatches were both evident at the verb (mismatching the action; identifying the first noun phrase as a role filler that mismatches its role in the picture, respectively), critically, the time course of their effects differed, as did - at least for the positivity during the early verb in Experiment 1 - polarity. Role mismatch effects were further absent in the later time window at the verb for which we found the verb-action congruence N400 effect. The reliable interaction of role and action congruence suggests these two effects are dissociable. To the extent that a single mechanism account does not straightforwardly predict this dissociation, our findings appear to accord better with the view that multiple functional brain mechanisms govern visual context effects during online language comprehension.

Neither the ERP nor verification time data confirmed the complexity account which predicted substantially longer verification latencies for role than action mismatches. In both studies, verification times to the role relations conditions were no longer than those to action mismatches. A complexity account also predicts larger (and possibly delayed) negative mean ERP amplitudes for role mismatches (combined mismatch and role mismatch) relative to action mismatches (action mismatch and combined mismatch, 300-500 ms at the verb, e.g., Fig. 2). This also was not what we observed.

Why then did we find a difference in ERP effects for a role relations mismatch relative to verb-action mismatch effects, while prior research has failed to find differences between ERP congruence effects in response to such - at first blush - different mismatches as object locations versus role relations (Vissers et al., 2008; Wassenaar & Hagoort, 2007)? First, prior studies did not compare object location with role relations mismatches directly. A theoretically more interesting possibility is that for both the role relations and object location mismatches, re-processing involves restructuring of mental representations (spatially and/or in terms of thematic role relations) whereas for our verb-action mismatches, re-processing concerned lexico-semantic content (rather than the structure) of mental representations.

In sum, we find that the time course of visual context influences on language comprehension can vary as a function of which aspects of a picture (role relations versus actions) mismatch corresponding aspects of a sentence. The findings best align with an incremental account of comprehension in picture-sentence verification.

Acknowledgments

This research was funded by a postdoctoral fellowship to PK (German research foundation, DFG) and by NIH grants HD-22614 and AG-08313 to Marta Kutas. The studies were conducted while PK was at UC San Diego.

References

Altmann, G. T. M. (2004). Language-mediated eye-movements in the absence of a visual world: the 'blank screen paradigm'. *Cognition*, *93*, B79–B87.

Berkum, J. van, Hagoort, P., & Brown, C. M. (1999). Semantic integration in sentences and discourse: evidence from the n400. *Journal of Cognitive Neuroscience*, *11*, 657–671.

Carpenter, P. A., & Just, M. A. (1975). Sentence comprehension: a psycholinguistic processing model of verification. *Psychological Review*, *82*, 45–73.

Clark, H. H., & Chase, W. G. (1972). On the process of comparing sentences against pictures. *Cognitive Psychology*, *3*, 472–517.

Ferretti, T. R., Singer, M., & Patterson, C. (2008). Electrophysiological evidence for the time course of verifying text ideas. *Cognition*, *108*, 881-888.

Just, M. A., & Carpenter, P. A. (1971). Comprehension of negation with qualification. *Journal of Verbal Learning and Verbal Behavior*, *10*, 244–253.

Knoeferle, P., Habets, B., Crocker, M. W., & Münte, T. F. (2008). Visual scenes trigger immediate syntactic reanalysis: evidence from ERPs during situated spoken comprehension. *Cerebral Cortex*, *18*, 789–795.

Knoeferle, P., Urbach, T. P., & Kutas, M. (2009). Is incremental semantic interpretation related to end-of-sentence verication?: Evidence from correlation analyses. In N. Taatgen, H. van Rijn, L. Schomaker, & J. Nerbonne (Eds.), *Proceedings of the Annual Conference of the Cognitive Science Society* (p. 1127-1132). Cognitive Science Society, Inc.

Kutas, M. (1993). In the company of other words: Electrophysiological evidence for single-word and sentence context effects. *Language and Cognitive Processes*, *8*, 533–572.

Kutas, M., Van Petten, C., & Kluender, R. (2006). Handbook of psycholinguistics. In M. Traxler & M. Gernsbacher (Eds.), (2nd Edition ed., p. 659-724). New York: Elsevier.

Monetta, L., Tremblay, T., & Joanette, Y. (2003). Semantic processing of words, cognitive resources and n400: An event-related potentials study. *Brain and Cognition*, *53*, 327–330.

Singer, M. (2006). Verification of text ideas during reading. *Journal of Memory and Language*, *54*, 574-591.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, *268*, 632–634.

Vissers, C., Kolk, H., van de Meerendonk, N., & Chwilla, D. (2008). Monitoring in language perception: evidence from erps in a picture-sentence matching task. *Neuropsychologia*, 967-982.

Wassenaar, M., & Hagoort, P. (2007). Thematic role assignment in patients with broca's aphasia: sentence-picture matching electrified. *Neuropsychologia*, *45*, 716-740.