The Quarterly Journal of Experimental Psychology

ISSN: 1747-0218 (Print) 1747-0226 (Online) Journal homepage: http://www.tandfonline.com/loi/pqje20

The influence of linguistic and cognitive factors on the time course of verb-based implicit causality

Arnout Koornneef, Jakub Dotlačil, Paul van den Broek & Ted Sanders

THE QUARTERLY JOURNAL OF EXPERIMENTAL PSYCHOLOGY, 2016
Vol. 69, No. 3, 455–481, http://dx.doi.org/10.1080/17470218.2015.1055282

The influence of linguistic and cognitive factors on the time course of verb-based implicit causality

Arnout Koornneef[1,2,3], Jakub Dotlačil[4,5], Paul van den Broek[3], and Ted Sanders[1,2]

[1]Department of Languages, Literature and Communication, Utrecht University, Utrecht, The Netherlands
[2]Utrecht Institute of Linguistics OTS, Utrecht University, Utrecht, The Netherlands
[3]Department of Education and Child Studies, Leiden University, Leiden, The Netherlands
[4]General Linguistics Department, University of Groningen, Groningen, The Netherlands
[5]Center for Language and Cognition, University of Groningen, Groningen, The Netherlands

(Received 16 September 2014; accepted 12 May 2015; first published online 26 June 2015)

In three eye-tracking experiments the influence of the Dutch causal connective "want" (*because*) and the working memory capacity of readers on the usage of verb-based implicit causality was examined. Experiments 1 and 2 showed that although a causal connective is not required to activate implicit causality information during reading, effects of implicit causality surfaced more rapidly and were more pronounced when a connective was present in the discourse than when it was absent. In addition, Experiment 3 revealed that—in contrast to previous claims—the activation of implicit causality is not a resource-consuming mental operation. Moreover, readers with higher and lower working memory capacities behaved differently in a dual-task situation. Higher span readers were more likely to use implicit causality when they had all their working memory resources at their disposal. Lower span readers showed the opposite pattern as they were more likely to use the implicit causality cue in the case of an additional working memory load. The results emphasize that both linguistic and cognitive factors mediate the impact of implicit causality on text comprehension. The implications of these results are discussed in terms of the ongoing controversies in the literature—that is, the focusing-integration debate and the debates on the source of implicit causality.

Keywords: Implicit causality; Eye tracking; Reading; Connectives; Working memory.

While processing a text, readers encounter various coherence cues that are used to guide comprehension. Some of these cues are coded in the discourse overtly and convey their processing instructions to the reader directly. Pronominals ("he", "her", etc.), for example, are linguistic devices that are used to maintain the focus on the most prominent entities (see e.g., Ariel, 2001; Arnold, 2010, for discussion). In addition to these referential markers, texts typically contain connectives (e.g., "because", "so", etc.), which signal more specific information about the relationship between clauses, sentences, or even larger segments of a text. For instance, the semantics of the backward causal connective "because"

Correspondence should be addressed to Arnout Koornneef, Department of Education and Child Studies, Leiden University, Pieter de la Court building, Wassenaarseweg 52, 2333 AK, Leiden, The Netherlands. E-mail: a.w.koornneef@fsw.leidenuniv.nl

We thank Iris Mulders, Alex Manus, Martijn van der Klis, Kate Backhouse, Renate van Zoest, Marja Oudega, and two anonymous reviewers for their involvement in this research.

Supported by an NWO (Netherlands Organization for Scientific Research) Veni grant [grant number 275-89-012] awarded to A.K.

signal that the second clause of the sentence "*John washed his shirt because it was stained*" will be the cause or explanation of the event described in the first clause of the sentence.

In addition to overt linguistic devices such as connectives and pronominals, discourses contain covert coherence cues. In the present study, the focus is on verb-based *implicit causality*, one of those more hidden cues. Implicit causality is a property of a group of interpersonal verbs in which one or the other of the verb's arguments is implicated as the underlying cause of the action or attitude (e.g., Au, 1986; Brown & Fish, 1983; Garvey & Caramazza, 1974; Garvey, Caramazza, & Yates, 1975; Greene & McKoon, 1995; Long & De Ley, 2000; Stewart, Pickering, & Sanford, 2000). For example, "apologize" can be categorized as a verb with a bias towards the subject. When people are asked to finish a sentence fragment such as "David apologized to Linda because . . . " they will typically provide information about the subject "David" as being the cause of the apologizing event (since in a canonical sentence order "David" is the first noun phrase, "apologize" is often referred to as an NP1 verb). On the other hand, an NP2 verb such as "praise" reveals an opposite bias. People are more inclined to continue the sentence "David praised Linda because …" about the object "Linda".

The influence of implicit causality on resolving pronominals and establishing coherence relations between consecutive clauses or sentences has been studied extensively. One key finding has been observed consistently across various languages and methodologies. That is, readers and listeners can activate and use the information provided by an implicit causality verb rapidly while building their mental representation of the discourse (Featherstone & Sturt, 2010; Koornneef & Van Berkum, 2006; Long & De Ley, 2000; McDonald & MacWhinney, 1995; Pyykkönen & Järvikivi, 2010). For example, in reading experiments, people slow down almost instantaneously after encountering a pronominal that is inconsistent with the bias of the verb (e.g., "David praised Linda because *he* …"; Featherstone & Sturt, 2010; Koornneef & Sanders, 2013; Koornneef & Van Berkum, 2006; Van Berkum, Koornneef, Otten, & Nieuwland, 2007).

The mechanisms underlying the activation and usage of implicit causality, however, are still poorly understood (Hartshorne & Snedeker, 2013). This is perhaps most clearly illustrated by the enduring debate between proponents of the focusing and integration accounts. Whereas in the former account implicit causality is thought to become active automatically, presumably due to the verb itself (e.g., McDonald & MacWhinney, 1995; Pyykkönen & Järvikivi, 2010), the latter states that implicit causality is only relevant as an integration cue later in the sentence, when readers or listeners relate the incoming information to the meaning of the verb (Garnham, Traxler, Oakhill, & Gernsbacher, 1996; Stewart et al., 2000). Another long-standing controversy revolves around the origins of implicit causality. On the one hand, a significant number of scholars claim that it should be treated as a lexical property encoded in the semantics of the verb. On the other hand, just as many authors suggest that there is nothing special about the lexicosemantics of implicit causality verbs, but rather that the biases originate from more general, nonlinguistic, cognitive preferences (for detailed discussions see Crinean & Garnham, 2006; Hartshorne & Snedeker, 2013; Kehler, Kertz, Rohde, & Elman, 2008; Pickering & Majid, 2007).

The persistence of these controversies is partly due to the fact that some basic aspects of implicit causality have not been addressed adequately. As a result, there are several unresolved issues pertaining to how both linguistic and cognitive factors impact the time course of implicit causality. For example, it is unclear to what extent implicit causality asserts its influence without a causal connective being present (Cozijn, Commandeur, Vonk, & Noordman, 2011)—a crucial piece of the linguistic puzzle to determine the source of implicit causality information. Furthermore, some findings have been taken to suggest that the activation of implicit causality depends on a complex causal inference and is, therefore, not activated and used by everyone in every situation: Only highly skilled readers have enough working memory resources to activate implicit causality online (Long & De Ley, 2000). The latter proposal emphasizes the importance of taking into account the cognitive differences between readers, most notably working memory capacity, while describing and explaining the impact of implicit causality on text comprehension. This, however, has been neglected in most of the studies conducted so far.

In all, there is a growing awareness that the time course of implicit causality is mediated by both linguistic and cognitive factors, yet relatively little is known about when and how these factors assert their influence. In the three reading experiments presented here, this void is addressed. In Experiments 1 and 2, the focus was on the influence of a linguistic factor. More specifically, we examined how the absence and presence of the Dutch causal connective "want" (*because*) mediates the activation of implicit causality. In Experiment 3 we focused on an important cognitive factor. Utilizing a dual-task design, we examined the impact of working memory on the time course of implicit causality. Together, the results of these experiments revealed important novel insights into the nature of implicit causality.

EXPERIMENT 1: IMPLICIT CAUSALITY IN THE ABSENCE OF A BACKWARD CAUSAL CONNECTIVE

Most of the studies on implicit causality have tested sentence structures in which a main clause containing the implicit causality verb is followed by a (subordinate) clause headed by the connective "because"—or its equivalent in other languages (Cozijn et al., 2011; Koornneef & Van Berkum, 2006; Pyykkönen & Järvikivi, 2010). These studies leave open the possibility that implicit causality is not truly a verb-based phenomenon, but rather emerges as a function of both the meaning of the verb and the connective. A number of studies made an attempt to disentangle these interactions between the verb and the connective, either by varying the connective (Ehrlich, 1980; Koornneef & Sanders, 2013; Rigalleau, Guerry, & Granjon, 2014), or by removing the connective altogether and using a full stop between the two clauses instead (Majid, Sanford, & Pickering, 2007; McDonald & MacWhinney, 1995). The studies revealed that replacing "because" with connectives such as "so", "and", and "but" drastically changes or reverts the bias of the verbs, indicating that the activation of implicit causality is closely tied to the coherence relation in which the verb appears (Ehrlich, 1980; Koornneef & Sanders, 2013; Stevenson, Crawley, & Kleinman, 1994).

Importantly, this should not be taken to suggest that the connective "because" is always *required* to trigger the activation of implicit causality. Offline completion studies and a probe verification study, for example, have shown that similar biases emerge for sentences in which no connective is present. However, although these studies presented some valuable insights into the characteristics of implicit causality verbs, they were not informative about the time course of implicit causality. For obvious reasons, the offline completion studies (Koornneef & Sanders, 2013; Majid et al., 2007) do not assess real-time comprehension processes. In that sense, the online probe study (McDonald & MacWhinney, 1995) is more informative about the time course of implicit causality, yet the results are difficult to interpret due to some methodological concerns. For example, probe studies tend to invite readers to memorize the content words of a sentence and to invest an unrepresentative amount of attention in assigning antecedents (Gordon, Hendrick, & Foster, 2000). Moreover, outside of the laboratory, people do not verify probes when they read or listen to linguistic input. These concerns related to the ecological validity of the probe verification methodology demonstrate the need to examine the time course of implicit causality with less obtrusive methodologies.

Addressing these concerns, Pyykkönen and Järvikivi (2010) conducted a visual-world eye-tracking study and showed that implicit causality information was activated *prior to* the connective. They concluded, therefore, that implicit causality was activated during real-time processing independently of the connective "because". As pointed out by Cozijn et al. (2011), however, this conclusion should be interpreted with care. Most importantly, Pyykkönen and Järvikivi included consistent continuations in their study only. Consequently, their results can be explained by a task strategy on the part of the participants to capitalize on the implicit causality bias information, since it always led to the interpretation that is favoured by the bias of the verb. Moreover, the visual array the participants were looking at during the visual-world experiment contained, as in many similar studies, only a couple of pictures, two of which highlight the subject or object of the implicit causality verb. This restricted visual world may invite listeners to make an educated guess about upcoming referents before they encounter the connective. In all, Pyykkönen and Järvikivi's visual world study may have overestimated the very early preconnective influence of implicit causality verbs.

Since the results of previous online studies were inconclusive with respect to the influence of backward causal connectives, the goal of our first experiment was straightforward: We examined whether implicit causality is used in real time, even in the absence of a backward causal connective. To complement previous visual world and probe studies, we opted for a reading paradigm in which participants processed short stories while their eye movements were recorded. The stories contained a pronoun that was either consistent or inconsistent with the bias of a preceding implicit causality verb. Examples that involve a verb with a bias towards NP1 ("apologize") are given in Table 1. We predicted that if the connective is not required to activate implicit causality information, the reading-time patterns should mirror the results of previous reading studies (Featherstone & Sturt, 2010; Koornneef & Sanders, 2013; Koornneef & Van Berkum, 2006). Readers should slow down rapidly after encountering a pronoun that is inconsistent with the bias of the verb. If, however, the connective is required for the online activation of implicit causality, there should be no early reading-time delay for the inconsistent pronouns.

Method

Participants

Participants were 46 undergraduate students (39 female, mean age 23 years, range 18–49 years) who received money for their participation. They were native speakers of Dutch, without a diagnosed reading or learning disability and with normal or corrected-to-normal vision.

Materials

The stimuli were created by adapting the stimuli of Koornneef and Sanders (2013). This stimulus set consisted of 24 stories containing implicit causality verbs with a strong NP1 bias and 24 stories containing verbs with a strong NP2 bias.[1] For each story, there is a consistent and inconsistent condition. Examples that involve a NP1 verb are given in Table 1, together with their approximate translation. In the first sentence, a situation was sketched in which a man and a woman were introduced by their proper names. In the second sentence, a pronominal (usually "they") was used to keep both characters in focus to an equal extent. The third sentence contained the implicit causality verb and repeated the names of the two characters. The fourth sentence started with the critical pronoun "he". Instead of contrasting the Dutch equivalents of "he" and "she", we manipulated whether the fixed pronoun "he" was consistent with the verb's bias by swapping the argument position of the man and the woman involved. The Dutch equivalent of "she" was avoided, because it is ambiguous between a singular and a plural third-person pronoun.

To be able to (a) get a fine-grained measure of the time course of implicit causality, (b) interpret

[1]The comparison of inconsistency effects elicited by NP1 verbs and NP2 verbs is confounded with the effects of distance between anaphor and antecedent, of first mention, and of the antecedent's structural position (see Garnham, 2001, for an overview of the relevance of these factors). The reason for including both NP1 and NP2 verbs was to be able to *control* for these factors, rather than studying potential differences between NP1 and NP2 verbs (see Koornneef & Sanders, 2013; Koornneef & Van Berkum, 2006).

Table 1. *NP1-biased example of stimuli used in Experiments 1–3*

Introduction section	David en Linda reden allebei behoorlijk hard. Bij een druk kruispunt botsten zij met hun auto's stevig op elkaar.
	(David and Linda were both driving pretty fast. At a busy intersection they crashed hard into each other.)
Critical sentences	
consistent pronoun no connective	*David* bood zijn excuses aan Linda aan. *Hij* was volgens de getuigen van het ongeluk de veroorzaker van alle ellende.
	(David apologized to Linda. He according to the witnesses was the one to blame.)
inconsistent pronoun no connective	Linda bood haar excuses aan *David* aan. *Hij* was volgens de getuigen van het ongeluk niet de veroorzaker van alle ellende.
	(Linda apologized to David. He according to the witnesses was not the one to blame.)
consistent pronoun connective	*David* bood zijn excuses aan Linda aan. Want *hij* was volgens de getuigen van het ongeluk de veroorzaker van alle ellende.
	(David apologized to Linda. Because he according to the witnesses was the one to blame.)
inconsistent pronoun connective	Linda bood haar excuses aan *David* aan. Want *hij* was volgens de getuigen van het ongeluk niet de veroorzaker van alle ellende.
	(Linda apologized to David. Because he according to the witnesses was not the one to blame.)

Note: The verb has a bias towards the subject—that is, the underlined story character. The story character in italics is the referent of the pronoun. The two no-connective conditions are used in Experiments 1 and 3. All four conditions are used in Experiment 2. NP1 = first noun phrase.

potential delayed effects, and (c) accommodate for general spillover, at least five words after "he" were held constant across conditions. After these five words, the consistent and inconsistent versions diverged and ended with explicit information that made the story coherent as a whole.

The stimuli were divided into two lists, with only one version of each story (i.e., consistent or inconsistent) in a particular list. Twenty-four stories of a different study examining the processing costs of various coherence relations were included as fillers (the typical filler item involved a mix of temporal and causal coherence relations: "Mary jumped on the table. The table broke. She apologized and took a seat on the couch."). The participants were assigned to one of the lists, and for each participant the list was presented in a unique pseudorandomized order. Half of the experimental and filler trials were followed by a statement about the story to encourage discourse comprehension. Participants had to indicate whether the statement about the story was correct or false (half were correct, and half were false). These statements never directly probed the interpretation of the pronoun. All participants scored above 80% correct (mean score 92%).

Procedure

Eye movements were recorded with a desktop-mounted EyeLink 1000 eye tracker, which sampled the eyes at 500 Hz. The system had an eye position tracking range of 32° horizontally and 25° vertically, with a gaze position accuracy of 0.5°. Viewing was binocular but the tracker monitored only the gaze location of the right eye. All participants were individually tested in a soundproof booth at Utrecht University. Each session started with a written and oral instruction, followed by the calibration procedure for the eye tracker. During this latter procedure, the participants had to fixate a random sequence of dots at various locations on screen. Upon successful calibration, the experiment started with five practice trials, three followed by a question. The stories were presented in their entirety at a viewing distance of approximately 60 cm. Before presentation, a fixation mark appeared on the screen at the position of the first word of the first sentence. Participants were instructed to fixate this mark, and after a successful fixation the story appeared automatically. Participants pressed a button to progress when they finished reading a story. The comprehension questions were answered using two buttons on the same response box. Each session was completed within 45 min (with a maximum time-on-task of 30 min).

Results

Dependent variables

An ongoing issue in eye-movement research is how to appropriately measure the processing time on an individual word (for reviews see e.g., Boland, 2004; Clifton, Staub, & Rayner, 2007; Rayner, Chace, Slattery, & Ashby, 2006). In particular because readers only fixate about two thirds of the words, researchers typically report a number of different, yet interrelated measures (Rayner, Reichle, Stroud, Williams, & Pollatsek, 2006; see Featherstone & Sturt, 2010; Koornneef & Van Berkum, 2006, for examples of this approach in the context of implicit causality). Often a distinction is made between so-called "early" measures and "late" measures (see Clifton et al., 2007, for a discussion of why this terminology may be misleading). In the current study, to be able to capture both early influences and somewhat delayed influences of implicit causality on pronoun resolution, we computed five commonly reported (first-pass) reading-time measures for the six regions of interest (i.e., the critical pronoun and the five words following the pronoun). The "early" measures consisted of first-fixation durations (the duration of the very first fixation on a word) and first-gaze durations (the total reading time of a word before the reader either moves on, or looks back into the text). These measures are considered to reflect early processes, such as word recognition and syntactic parsing operations (Boland, 2004). The "later" measures consisted of right-bound durations (the sum of all fixations on a word before moving on progressively) and regression path durations (the sum of fixation durations from the time when the reader fixates a word to the time when the reader moves on progressively). These reading-time measures capture the processing costs associated with the integration of a word into the preceding text (arguably an "early" effect), and, in addition, they capture the processing consequences of overcoming integration difficulties (arguably a "later" effect).[2] Finally, in addition to these continuous reading-time measures, we report the categorical measure probability of a regression (the likelihood of a regressive eye-movement after a word is fixated during first pass). Similar to the right-bound and regression path measures, the probability of a regression is often reported as a measure that is sensitive to word integration difficulties.

Analysis

Prior to all analyses, 1% of all trials were removed, because tracker losses or eye blinks made it impossible to determine the course of fixations. Regions that were skipped during first-pass reading were treated as missing data. Table 2 reports the means (and *SE* of the means) of the various dependent variables as a function of pronoun (two levels: implicit causality *consistent* pronoun versus implicit causality *inconsistent* pronoun) and sentence region. Mixed-effects linear regression models for the continuous reading time data (with the response variable log transformed to correct for right skewness) and mixed-effects logistic regression models for the categorical dependent measure probability of a regression were fitted on the data. We estimated the models with the the R package LME4.

It is common practice to analyse the regions of interest one by one for every reading-time measure (e.g., Featherstone & Sturt, 2010; Koornneef & Sanders, 2013; Koornneef & Van Berkum 2006). However, given that the current experiment contained six regions of interest (i.e., the critical pronoun and the five words following the pronoun), such a number of comparisons faced the risk of finding spurious effects.[3] To account for Type I errors, the word-by-word analyses were therefore preceded by a series of analyses in which we examined the various reading-time

[2]Since right-bound and regression path durations capture both early and later processes of reading, some researchers prefer to label them as early measures, whereas other researchers prefer to label them as late measures (Clifton et al., 2007).

[3]We are thankful to two anonymous reviewers for discussion on this issue.

Table 2. *Mean reading times and the probability of regressions at the critical pronoun and the five subsequent words in Experiment 1*

	Sentence region											
	Pronoun		*Pro + 1*		*Pro + 2*		*Pro + 3*		*Pro + 4*		*Pro + 5*	
Dependent variable and condition	*Mean*	SE	*Mean*	SE	*Mean*	SE	*Mean*	SE	*Mean*	SE	*Mean*	SE
First-fixation duration												
con	200	5	236	3	226	5	217	3	208	3	184	3
inc	204	6	240	4	228	4	227	4	206	3	199	4
First-gaze duration												
con	202	5	243	4	237	5	227	4	229	4	194	4
inc	206	6	246	5	239	5	238	4	228	4	208	4
Right-bound duration												
con	203	5	248	4	251	6	245	5	247	5	202	5
inc	210	6	256	6	250	5	257	5	248	5	223	5
Regression path duration												
con	257	22	268	7	299	11	304	11	329	14	244	9
inc	253	14	292	13	307	12	343	16	321	13	288	12
Probability of regression												
con	.06	.02	.05	.01	.17	.02	.17	.01	.16	.01	.10	.01
inc	.08	.02	.05	.01	.15	.02	.19	.02	.17	.01	.13	.01

Note: Reading times in milliseconds. The values reflect the means (with standard error) aggregated over all data points. Pro = pronoun; con = consistent pronoun; inc = inconsistent pronoun.

measures for all regions together. These initial analyses contained three fixed factors: region (six levels with the pronoun region as the reference level), pronoun (consistent and inconsistent condition; the former was the reference level), and the interaction of pronoun and region. In addition, the models included the intercept and pronoun random factors for subject and item random effects, which was the maximal converging structure. Likelihood-ratio tests were conducted to find the best fitting model among the models that included region as a factor (i.e., we did not study whether region as a main effect was a significant predictor, as this did not bear on our research question). The results revealed a significant effect of pronoun in right-bound and regression path durations [right-bound: $\chi^2(1) = 7.0$, $p < .01$; regression path: $\chi^2(1) = 8.3$, $p < .01$]. In first fixation, first gaze, and the probability of a regression, the results approached significance [first fixation: $\chi^2(1) = 3.6$, $p = .06$; first gaze: $\chi^2(1) = 2.9$, $p = .09$; probability of a regression: $\chi^2(1) = 3.5$, $p = .06$]. Since we considered all regions only once per reading measure, we limited the problem of multiple comparisons, yet at the same time this approach revealed that the main experimental manipulation of pronoun was a significant factor in the regions of interest.

To obtain a fine-grained measure of the time course of implicit causality, these initial analyses were followed up with mixed-effects linear regression analyses on a word-by-word basis. All models included the fixed effect of pronoun (with the consistent condition as reference level) and the maximal subject and item random-effect structure.[4] For the word-by-word analyses we report the fixed-effects estimates, *t*- and *z*-statistics, and the associated *p*-values.[5] The results revealed longer reading

[4]The regression formula of the models used in Experiment 1: dependent measure = 1 + pronoun + (1 + pronoun subject) + (1 + pronoun | item)

[5]It is not clear how to determine the degrees of freedom for the *t*-statistics estimated by the mixed models fitted on the continuous dependent measures (Baayen, Davidson, & Bates, 2008), so we follow the practice of Barr, Levy, Scheepers, and Tily (2013) and report *p*-values based on a *z*-statistic, which closely approximates the *t*-statistic for high degrees of freedom. The mixed-effects logistic

times in the inconsistent conditions than in the consistent conditions, five words after the critical pronoun (i.e., in the other regions we observed no significant effects). This was observed for all reading-time measures (first fixation: $\beta = 0.055$, $SE = 0.022$, $t = 2.5$, $p < .05$; first gaze: $\beta = 0.056$, $SE = 0.023$, $t = 2.5$, $p < .05$; right-bound: $\beta = 0.080$, $SE = 0.024$, $t = 3.3$, $p < .001$; regression path: $\beta = 0.10$, $SE = 0.033$, $t = 3.2$, $p < .01$). In addition, the mixed-effects logistic regression models showed that four words after the critical pronoun the probability of a regression was significantly higher in the inconsistent condition than in the consistent condition ($\beta = 0.44$, $SE = 0.19$, $z = 2.3$, $p < .05$).

Discussion

The results of Experiment 1 showed that a backward causal connective is not a prerequisite to activate implicit causality during real-time processing: Four to five words after the critical pronoun, but long before the end of the sentence, readers slowed down after encountering a bias-inconsistent pronoun. First of all, this finding is consistent with the results of a number of offline completion studies showing that similar referential biases emerge after a backward causal connective and a full stop. More importantly however, our results complement the findings of Pyykkönen and Järvikivi (2010) who observed in a visual-world study that implicit causality is activated before people encounter the connective. In contrast to their study, our stimuli contained as many consistent as inconsistent continuations. Hence, the objection of Cozijn et al. (2011) that the redundancy of the connective may result from a task strategy that builds on the absence of items that go against the bias of the verbs does not apply to our study. In that sense, our study is one of the first to show convincingly that implicit causality effects emerging during real-time processing can be traced back to the verb itself. No backward causal connective—or in fact no connective at all—is required to trigger the bias of the verb.

However, when we compare the results of Experiment 1 to those of previous reading studies on implicit causality that *did* include a backward causal connective, the reading-time delays after the inconsistent pronoun surfaced relatively late. More specifically, although both "early" eye-tracking measures (first-fixation and first-gaze durations) and "later" eye-tracking measures (right-bound and regression path durations) displayed the influence of implicit causality, they only did so five words after the critical pronoun. Using almost the same stimuli, Koornneef and Van Berkum (2006) observed a much earlier influence of implicit causality, namely at or immediately after the pronoun. Although these differences may be partly due to different statistical analyses in the present study—for example, Koornneef and Van Berkum reported by-subjects and by-items repeated measures analyses of variance (ANOVAs)—there may be a deeper reason for the discrepancy. That is, even though a backward causal connective is not required to activate implicit causality, its presence may very well *foreground* implicit causality as a reliable processing cue (cf. Kamalski, Lentz, Sanders, & Zwaan, 2008). Consequently, implicit causality may have a more pronounced influence on pronoun resolution when a connective is present than when it is absent. We examine this possibility further in a second eye-tracking experiment.

EXPERIMENT 2: FOREGROUNDING IMPLICIT CAUSALITY AS A PRONOUN RESOLUTION CUE

While resolving a pronoun, readers and listeners employ several cues to locate the most salient referent. For example, it is well known that people prefer to connect pronouns to entities that are mentioned first in a story. To give another example, people are more inclined to link a pronoun to a recently mentioned entity (note that these primacy and recency constraints do not always converge on the same referent). So, implicit causality is just one of

regression models fitted on the categorical dependent measure (probability of a regression) are directly provided with a z-statistic, so no issue arises there.

the many sources of information that readers and listeners recruit to interpret a pronoun, and although most accounts on implicit causality assume that its information is readily available at the pronoun, when and how this information appears as a significant factor in the service of pronoun resolution seems to depend on a range of other (linguistic) factors (e.g., McDonald & MacWhinney, 1995; Pyykkönen & Järvikivi, 2010).

In Experiment 2, we further examined one of these additional factors—namely, how the presence and absence of a backward causal connective affect the time course of implicit causality information. The same stimuli as those in Experiment 1 were used, with one important adaptation. In addition to comparing bias-consistent to bias-inconsistent pronouns, we added one factor to the design by varying whether the connective "want", a Dutch equivalent of the connective "because", preceded the critical pronouns (see Table 1).

As discussed, connectives rapidly—and often profoundly—constrain the way in which a sentence or discourse will proceed, including their referential continuations (e.g., Canestrelli, Mak, & Sanders, 2013; Koornneef & Sanders, 2013; Mak, Tribushinina, & Andreiushina, 2013). Furthermore, offline completion studies have shown that implicit causality biases tend to be stronger if a backward causal connective is present than in the situation in which a full stop is used between sentences (e.g., Koornneef & Sanders, 2013). Based on these results, the very general expectation follows that a causal connective will increase the influence of implicit causality during reading. We should, however, distinguish at least two ways in which a more prominent influence may surface in the eye-tracking data. On the one hand, it could primarily affect the intensity of the pronoun inconsistency effect. The connective further increases the preference for one of the arguments of the implicit causality verb, which in turn makes it even more difficult to interpret a pronoun that refers to the competing argument. In this case, one would expect that an inconsistent pronoun in both the connective and no-connective conditions slows down reading at the same moment in time, but that this inconsistency effect is larger for the connective conditions than for the no-connective conditions. The results of Experiment 1, however, suggest that it is also conceivable that the connective primarily affects *when* implicit causality is used during pronoun interpretation. If this is true, then an inconsistent pronoun in the connective conditions should slow down reading more quickly than an inconsistent pronoun in the no-connective conditions. Of course, it is also possible—and in fact plausible—that the eye-tracking data reveal a combination of these two patterns, meaning that a backward causal connective impacts both the intensity and the timing of the pronoun inconsistency effect.

Hence, the key prediction for Experiment 2 was that the factors pronoun (consistent vs. inconsistent) and connective (absent vs. present) should interact at or rapidly after the critical pronoun. In addition to this primary prediction, we anticipated that readers would speed up immediately following the connective, irrespective of whether the pronoun was consistent or inconsistent with the bias of the verb. This secondary prediction follows from the results of a number of studies showing that the sentence regions directly following a connective are read relatively quickly. It is often assumed that these shorter postconnective reading times reflect a more efficient (incremental) integration process, triggered by the connective (e.g., Canestrelli et al., 2013; cf. Cozijn, 2000; cf. Millis & Just, 1994; cf. Noordman & Vonk, 1998; see Sanders & Canestrelli, 2012, for an overview).

Method

Participants

Participants were 41 undergraduate students (32 female, mean age 23 years, range 19–30 years). None of them participated in Experiment 1.

Materials and procedure

The same stimuli as those in Experiment 1 were presented to the participants. As discussed above, for every item two additional conditions were created by varying whether the Dutch connective "want" preceded the pronominal (cf. Koornneef & Sanders, 2013). The stimuli were divided into four counterbalanced lists. Forty stories of a different study (the self-paced reading experiment in Van

Berkum, Brown, Zwitserlood, Kooijman, & Hagoort, 2005) examining processes of syntactic and semantic integration were included as fillers (the typical filler item involved a story with an anticipated noun or an unanticipated noun: "The burglar had no trouble locating the secret family safe. Of course, it was situated behind a big but unobtrusive painting/bookcase."). The participants were assigned to one of the lists, and for each participant the list was presented in a unique pseudorandomized order. Half of the experimental and filler trials were followed by a statement about the story to encourage discourse comprehension (half were correct, and half were false). All participants scored above 68% correct (mean score 85%). The procedure and apparatus were the same as those in Experiment 1.

Results

Due to tracker losses or eye blinks, 1% of all trials was removed from the analyses. We computed the same reading-time measures for the regions of interest (i.e., the critical pronoun and the five words following the critical pronoun). Table 3 reports the means of the various measures as a function of pronoun (consistent vs. inconsistent), connective (absent vs. present), and sentence region.

Mixed-effects linear and logistic regression models were fitted on the data (with the continuous reading measures log transformed to correct for right skewness). As in Experiment 1, we first considered the reading measures for all regions together, with the following factors: region (six levels; the pronoun region was the reference level), pronoun (consistent vs. inconsistent; the consistent condition was the reference level), and connective (no-connective vs. connective with the former as reference level), and the interaction of these factors (the models also included the maximal converging random structure for subjects and items).[6] The results of the likelihood-ratio tests revealed significant main effects of pronoun and connective in the regression path measure [pronoun: $\chi^2(1) = 3.9$, $p < .05$; connective: $\chi^2(1) = 7.4$, $p < .01$]. Moreover, the three-way interaction Pronoun × Connective × Region was significant for this measure, $\chi^2(5) = 11.7$, $p < .05$. In addition, in the probability of a regression measure the main effect of connective was significant, $\chi^2(1) = 7.2$, $p < .01$, and the main effect of pronoun approached significance, $\chi^2(1) = 3.3$, $p = .07$. The two-way interaction of Connective × Region and the three-way interaction of Pronoun × Connective × Region were significant, as well [Connective × Region: $\chi^2(5) = 45.6$, $p < .001$; Pronoun × Connective × Region: $\chi^2(5) = 15.1$, $p < .01$]. Together, these results showed that the experimental manipulations of pronoun and connective, and their interactions as a function of region, were significant factors in the regions of interest.

Subsequently, we conducted mixed-effects linear regression analyses on a word-by-word basis. All models included the interaction between the fixed factors pronoun (consistent vs. inconsistent) and connective (no-connective vs. connective) and the maximal subject and item random-effect structure (the consistent/no-connective condition was the reference level).[7]

The most relevant effects were observed two and four words after the critical pronoun. First of all, a significant Pronoun × Connective interaction emerged two words after the pronoun in regression path durations ($\beta = 0.20$, $SE = 0.079$, $t = 2.5$, $p < .05$). As illustrated in Figure 1, in conditions with a connective, readers displayed shorter regression path durations when the pronoun matched the bias of the verb than when it did not match the bias of the verb. Yet, there was no difference between the consistent and inconsistent conditions in the story versions without a connective. A similar interaction was observed for the probability of a regression measure ($\beta = 1.5$, $SE = 0.51$, $z = 2.9$, $p < .01$). In the connective conditions, readers were less likely to regress when

[6]As in Experiment 1, we did not run the likelihood-ratio tests to study the effect of region since this is irrelevant for our research.

[7]The regression formula of the models used in Experiment 2: dependent measure = 1 + Pronoun × Connective + (1 + Pronoun × Connective | participant) + (1 + Pronoun × Connective | item).

Table 3. *Mean reading times and the probability of regressions at the critical pronoun and the five subsequent words in Experiment 2*

		Sentence region											
		Pronoun		Pro + 1		Pro + 2		Pro + 3		Pro + 4		Pro + 5	
Dependent variable and condition		*Mean*	SE	*Mean*	SE	*Mean*	SE	*Mean*	SE	*Mean*	SE	*Mean*	SE
First-fixation duration													
no connective	con	208	9	238	7	224	6	210	6	213	5	185	5
	inc	200	10	239	6	214	5	223	6	206	4	198	6
connective	con	230	5	219	5	201	4	220	7	205	5	191	6
	inc	223	5	220	6	220	6	223	7	209	5	188	6
First-gaze duration													
no connective	con	208	9	263	11	238	8	228	9	234	7	196	6
	inc	202	10	253	8	226	7	237	7	228	6	206	6
connective	con	235	6	222	5	216	6	231	7	226	6	198	6
	inc	230	6	229	7	235	7	241	7	229	7	198	6
Right-bound duration													
no connective	con	213	9	269	12	247	8	246	11	257	8	212	8
	inc	204	10	263	9	239	8	253	8	254	7	220	7
connective	con	238	6	229	6	222	7	242	9	247	8	206	7
	inc	236	6	241	8	258	9	257	9	247	8	215	9
Regression path duration													
no connective	con	262	20	303	16	329	25	340	23	358	21	326	41
	inc	364	104	312	26	309	18	325	17	367	19	320	21
connective	con	270	13	287	14	287	36	300	20	328	18	268	20
	inc	327	42	298	15	343	27	328	21	341	31	317	35
Probability of regression													
no connective	con	.09	.03	.08	.02	.20	.03	.22	.02	.18	.02	.14	.02
	inc	.13	.04	.09	.02	.17	.02	.20	.02	.23	.02	.17	.02
connective	con	.08	.02	.19	.02	.08	.02	.12	.02	.16	.02	.12	.02
	inc	.11	.02	.18	.02	.17	.02	.16	.02	.15	.02	.12	.02

Note: Reading times in milliseconds. The values reflect the means (with standard error) aggregated over all data points. Pro = pronoun; con = consistent pronoun; inc = inconsistent pronoun.

the pronoun matched the bias of the verb than when it did not match the bias of the verb. This difference was absent in the no-connective conditions. Furthermore, four words after the pronoun a main effect of pronoun was observed in the probability of a regression measure. Readers were more likely to look back in the text in the pronoun-inconsistent conditions ($\beta = 0.45$, $SE = 0.22$, $z = 2.1$, $p < .05$) than in the pronoun-consistent conditions.[8]

In addition to these results, the analyses revealed main effects for the factor connective one, two, three, and five words after the pronoun. One word after the pronoun, the probability of a regression was larger if the story contained a connective than if it did not ($\beta = 1.1$, $SE = 0.42$, $z = 2.6$, $p < .05$). An opposite pattern was observed later in the sentence. Three and five words after the pronoun, the probability of a regression was smaller in the connective conditions than in the no-connective conditions (pronoun + 3: $\beta = -0.91$, $SE = 0.28$, $z = -3.3$, $p < .01$; pronoun + 5: $\beta = -0.76$, $SE = 0.31$, $z = -2.5$, $p < .05$). Moreover, various continuous dependent measures showed that two words after the pronoun, the reading

[8]Since we used the pronoun-consistent no-connective condition as reference level, this main effect provides a direct comparison between the pronoun-consistent no-connective condition and the pronoun-inconsistent no-connective condition.

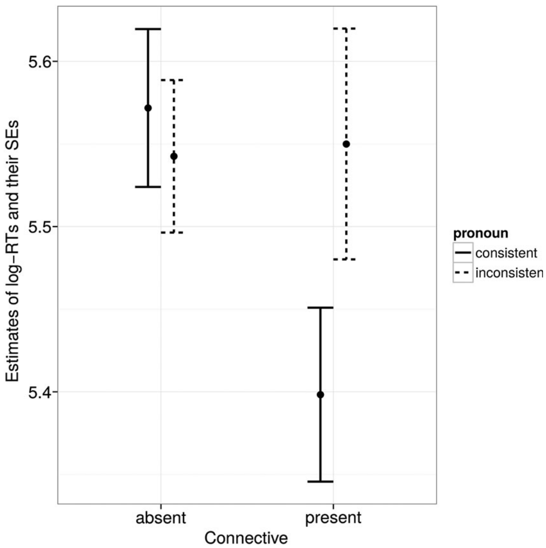

Figure 1. *Point estimates of regression path durations and corresponding standard errors two words after the pronoun in Experiment 2. RT = reading time.*

times in the connective conditions were shorter than those in the no-connective conditions (first fixation: $\beta = -0.080$, $SE = 0.030$, $t = -2.6$, $p < .01$; right-bound: $\beta = -0.10$, $SE = 0.052$, $t = -2.0$, $p < .05$; regression path: $\beta = -0.19$, $SE = 0.071$, $t = -2.6$, $p < .01$).

Discussion

First, the main effects for the factor connective are discussed. Initially the presence of a connective induced a processing disadvantage, as indexed by an increased probability to make a regressive eye movement out of the region immediately following the pronoun. However, this initial processing disadvantage was compensated by shorter reading times and a smaller number of regressive eye-movements somewhat later in the sentence. As such, the results partly confirm the results of previous studies examining the influence of connectives during reading. These studies showed that the presence of a connective leads to a more economical integration process, indicated by decreasing reading times right after the connective (see Canestrelli et al., 2013, for discussion; cf. Millis & Just, 1994; cf., Cozijn, 2000). Our results revealed a more complex processing signature of the integration phase. The anticipated processing advantage did occur, but only after an initial processing disadvantage. This shows that although connectives may boost integration, they do require—or induce—some additional processing at the very start.

Furthermore, the results of Experiment 2 confirmed and extended the general idea that the time course of implicit causality is mediated by a range of factors (e.g., Koornneef & Sanders, 2013; McDonald & MacWhinney, 1995; Pyykkönen & Järvikivi, 2010). More specifically, even though the connective "want" was not required to activate implicit causality information, its presence affected both the timing and the strength of implicit causality. First, implicit causality effects emerged more quickly in the sentence. Whereas the connective conditions revealed a robust bias-inconsistency effect as rapidly as two words after the critical pronoun, the first (and only) significant difference in the no-connective conditions appeared two words more downstream—that is, four words after the pronoun. Second, the impact of implicit causality was relatively weak in the no-connective conditions. The only measure revealing an inconsistency effect in the no-connective conditions was the probability of a regression. Hence, for these conditions, none of the continuous reading-time measures revealed a significant influence of implicit causality.

The observed main and interaction effects are consistent with the idea that at the moment readers encounter the connective "want", they infer that the sentence containing the implicit causality verb should be interpreted as the consequence or result of the adjacent sentence. Initially, the activation and incorporation of this specific knowledge induces additional processing costs. However, at the same time the semantic properties of the connective jump-start a more constrained integration phase. In this more restricted (or even causal-specific) integration phase, implicit causality becomes a prominent processing cue, impacting pronoun resolution more rapidly and more intensely than in the absence of a connective.

As a final point, the results for the no-connective conditions of Experiment 2 were consistent with the results of Experiment 1, which exclusively included stories without a connective: In both experiments, we observed a somewhat delayed influence of implicit causality for no-connective story versions. However, the influence of implicit causality seemed more pronounced in Experiment 1 as it impacted *all* measures, as opposed to Experiment 2 where only the probability of a regression measure was affected. Of course, many factors may have caused this discrepancy. Yet, it seems likely that it is related to the different experimental designs employed. In Experiment 1, none of the conditions contained a connective, whereas in Experiment 2 half of the test items contained a backward causal connective. It is conceivable that readers became less inclined to employ implicit causality as a pronoun resolution cue in the no-connective items of Experiment 2, since they encountered just as many stories with a connective. In other words, driven by the global experimental context, participants may have adopted slightly different processing strategies across experiments, inducing subtle differences in the strength of implicit causality. Although this explanation is somewhat speculative, it is consistent with proposals stating that the usage of implicit causality is easily attenuated (e.g., Ehrlich, 1980) and may depend on the processing strategy of the reader (Van Berkum, De Goede, Van Alphen, Mulder, & Kerstholt, 2013). Furthermore, it aligns with the perspective of the current study that not only linguistic factors but also cognitive factors play a vital role in when—or whether—implicit causality is used by readers during text comprehension.

As mentioned earlier, however, relatively little is known about the cognitive dimensions of the usage of implicit causality. Particularly striking in our opinion is the absence of studies examining the impact of working memory capacity, not only because working memory has been widely recognized as constraining numerous language-related computations, ranging from local syntactic and semantic parsing operations to more global pragmatic inferences (for reviews, see Daneman & Merikle, 1996; Gathercole & Baddeley, 1993), but in particular, because the claim has been made that implicit causality results from a resource-consuming causal inference (e.g., Long & De Ley, 2000)—as opposed to being the result of a more or less automatized process (McDonald & MacWhinney, 1995). This would imply that working memory capacity plays an important role in the activation and usage of implicit causality.

EXPERIMENT 3: THE INTERPLAY BETWEEN IMPLICIT CAUSALITY AND WORKING MEMORY

We are not aware of any studies examining the influence of working memory directly, yet two studies are relevant to the current issue. First of all, in a series of probe experiments, Long and De Ley (2000) examined how reading proficiency affected the time course of implicit causality. They reported that skilled readers (as indexed by their score on the Nelson–Denny Reading Test) were able to use implicit causality immediately during pronoun resolution. In contrast, less skilled readers only showed sensitivity to the cue near the end of a sentence. These results were explained by assuming that the activation of implicit causality requires a relatively complex causal inference. As a result, only skilled readers have the processing resources available to make the inference during online comprehension.

The second relevant study was conducted by Van Berkum et al. (2013). In an event-related potential (ERP) experiment, they investigated how people's mood affected the use of implicit causality. Although the main aim of their study was not to decipher the specifics of implicit causality, but rather, to explore how language processing is embodied into other systems of the human brain, the results were striking. Whereas people in a good mood displayed verb-bias inconsistency effects right at the pronoun, the influence of implicit causality was completely absent for people in a bad mood. According to Van Berkum et al., one of the explanations for this contrast could be related to the willingness of the participants to construct a rich mental model of a story. More specifically, readers in a bad or sad mood may resort to a superficial reading strategy and are more likely to avoid the processing costs of using implicit causality to integrate or anticipate upcoming referents. So, analogous to Long and De Ley, Van Berkum et al. construe a potential explanation around the assumption that the activation of implicit causality is a relatively resource-consuming mental operation and, as a result, is not used by everyone in every situation.

The findings reported so far are most easily reconciled with the idea that implicit causality demands a considerable amount of readers' working memory resources. But clearly, the evidence is scarce. Whereas Long and De Ley (2000) correlated implicit causality with reading proficiency, and not with working memory capacity directly, the evidence provided by Van Berkum et al.'s (2013) study is even more circumstantial. Moreover, other studies have shown that generating inferences—or even predictions—from the meaning of verbs may not be resource-consuming at all. For example, despite their smaller working memory capacity, children as young as three years old are able to use subtle differences in the transitive meaning of verbs to rapidly resolve ambiguous pronouns, and they may even be capable of using verb-semantics to estimate the likelihood that a particular story character will be mentioned in the upcoming discourse (Pyykkönen, Matthews, & Järvikivi, 2010). There is no a priori reason to assume that the opposite will hold for verb-based implicit causality, casting further doubt on the hypothesis that its activation and usage are cognitively demanding.

Given these indirect and mixed results, we conducted a third eye-tracking experiment to gain more insight into the involvement of working memory in the usage of implicit causality during reading. The issue was approached in two ways. First, we obtained measures of the working memory capacity of the participants (i.e., forward and backward digit span) to correlate individual differences to the usage of implicit causality. Analogously to the proposal of Long and De Ley (2000), we predicted that readers with a high working memory capacity should display an earlier use of implicit causality than readers with a lower memory capacity.

In addition (and in contrast to previous studies), we specifically aimed at obtaining a more direct measure of the interplay between working memory and implicit causality. The crucial manipulation in Experiment 3, therefore, consisted of adding a secondary task to the main task of reading. The same set of bias-consistent and inconsistent stimuli as that in Experiment 1 was presented to a new group of participants, yet we divided the reading session into two blocks. In one of the blocks, the participants simply read for comprehension—that is, replicating the task of Experiments 1 and 2—while in the other reading block the participants were also asked to mentally store and recall a random sequence of digits, presented before each story. Keeping the digits available during reading puts an additional strain on working memory and as such decreases the available resources to construct a rich situation model of the story. If the activation of implicit causality requires a critical amount of the reader's working memory capacity (e.g., because it relies on a complex causal inference), the influence of implicit causality should be attenuated, delayed, or even nonexistent in the case of an additional working memory load. Alternatively, no clear interaction between the factors working memory load and implicit causality is expected if the latter is activated and used without consuming too much of people's working memory resources.

Method

Participants

Participants were 40 undergraduate students (29 female, mean age 23 years, range 17–52 years) who received money for their participation. None of them participated in Experiment 1 or Experiment 2.

Materials and procedure

Before the eye-tracking session, participants completed computerized versions of the forward and backward digit span task. We included the digit span task in the present study as a quick yet reliable measure of working memory capacity. The freely available PEBL software was used to run standard visual versions of the tests (Mueller, 2013).

After a short 5-minute break, the eye-tracking session started. During this session, the same stimuli as those in Experiment 1 were presented —note that these stimuli did not include versions with a backward causal connective. In addition to manipulating whether the critical pronoun was consistent or inconsistent with the bias of the verb, we introduced a new factor to the design: dual task. As discussed briefly above, the reading session was divided into two blocks. In one of the blocks the participant read for comprehension only (i.e., no additional task). In the other block, the stories were preceded by a random string of five digits. The visually presented digits appeared on screen all at once for 3000 ms. A different randomized string was used before each trial. The participants were asked to memorize the five digits, and keep them active in memory while they were reading the story that appeared immediately afterwards. After they finished reading the story, another string of five digits was presented visually. The participants were asked to match these five digits to the ones they memorized, yet in a backward order. To illustrate this, they should indicate that it was a match if the string before the story was "1 2 3 4 5" and the string after the story was "5 4 3 2 1". In contrast, they should indicate that it was a mismatch if the latter string was "5 3 4 2 1" (note that the ordering of the integers 3 and 4 is incorrect). In half of the trials, the strings presented before and after the story matched, in the other half they did not—either because the strings contained different digits, or because the ordering of the digits did not match. The participants provided their answer by using two buttons on the response box, one for "match" and the other one for "mismatch".

In a Latin-square design, the stimuli were divided into eight lists. We counterbalanced the two levels of pronoun (consistent or inconsistent). In addition, we balanced whether the dual task was present during the first or, alternatively, during the second block of the reading session. Finally, we also made sure that across participants, each story was presented during both the dual-task block and the normal reading block. Thirty-two stories of a different study examining processes of negation and semantic integration were included as fillers (the typical filler item involved a story with a positive object, indicated by "a", or a negative object, indicated by "no": "John feels uncomfortable at work. He is the only one who holds *a/no* Bachelor's degree."). The participants were assigned to one of the lists, and for each participant the list was presented in a unique pseudorandomized order. As in Experiments 1 and 2, half of the experimental and filler trials were followed by a statement about the story to encourage discourse comprehension (half were correct, and half were false). In the dual-task block, the comprehension questions were presented immediately after the story—that is, before the participants had to recall the memorized digits. All participants scored above 80% correct (mean score 87%), and there was no difference between the dual-task block (88% correct) and the no-dual-task block (86% correct).

Results

Forward and backward digit span task

The forward and backward digit span scores of the participants revealed a normal pattern with slightly higher scores in the forward version (forward: mean = 7.5, range = 5–10; backward: mean = 6.8, range = 4–10). For further analyses, we computed a composite score (Wilde, Strauss, & Tulsky, 2004). First, for each task we multiplied the maximal span with the total number of correct responses. This made it possible to distinguish participants that displayed the same maximal span, yet differed in how many series of digits they remembered correctly (Kessels, Van Den Berg, Ruis, & Brands, 2008; Kessels, Van Zandvoort, Postma, Kappelle, & De Haan, 2000). Subsequently, we took the sum of these scores for the forward and backward digit span to obtain the final composite score.

Eye-tracking experiment

Most of the participants performed reasonably well on the dual task in the eye-tracking session. On average, 72% of the five-digit sequences were recalled correctly (the range of 42–96% correct indicated that some participants were at chance level).[9] Due to tracker losses or eye blinks, 1% of all trials were removed from the eye-tracking data. Table 4 reports the various dependent variables as a function of pronoun (consistent vs. inconsistent), dual task (absent vs. present), and sentence region.

As in Experiments 1 and 2, we estimated mixed-effects linear and logistic regression models, first on all regions and then on the words separately. The first series of models included the factors region, pronoun, dual task, and composite digit span (i.e., a continuous predictor), and their interactions.[10] Likelihood-ratio tests revealed a significant effect of dual task in first-gaze durations, $\chi^2(1) = 6.3$, $p < .01$, right-bound durations, $\chi^2(1) = 6.7$, $p < .01$, and regression path durations, $\chi^2(1) = 5.7$, $p < .05$. Main effects of pronoun were significant in right-bound durations, $\chi^2(1) = 6.7$, $p < .01$, regression path durations, $\chi^2(1) = 12.0$, $p < .001$, and the probability of a regression measure, $\chi^2(1) = 6.2$, $p < .01$. A borderline-significant main effect of pronoun was observed for first-gaze durations, $\chi^2(1) = 3.8$, $p = .05$. Furthermore, the four-way interaction between all factors was significant for first-gaze and right-bound durations [first gaze: $\chi^2(10) = 23.2$, $p = .01$; right-bound: $\chi^2(10) = 25.3$, $p < .01$] and fell just short of significance for the probability of a regression measure, $\chi^2(10) = 17.2$, $p = .07$.

The analyses reported above (in particular the four-way interactions) showed that the manipulations of Experiment 3 were successful. As in the previous experiments, we examined the effects of the factors on the individual word level in a second series of analyses. All models included the interaction between the factors pronoun, dual task, and composite digit span. The reference level was the consistent/no-dual-task condition for the fixed factors and the lowest digit span score for the continuous predictor. We could not include the continuous predictor composite digit span in the subject random slopes because the predictor did not have within-subject variation. In addition, the same predictor in the item random slopes stopped most models from converging. To avoid differences between mixed-effect models across regions and measures, we completely removed the predictor composite digit span from the random-effect structure.[11]

The analyses revealed various significant main and interaction effects at the critical pronoun itself and, in addition, three words later in the sentence. Detailed information of the relevant models is reported in Table 5. The analyses for first-fixation, first-gaze, and right-bound durations at the critical pronoun returned the exact same results. All three measures showed a significant main effect for the composite digit span and, more interestingly, a significant interaction between composite digit span and the fixed factor pronoun. To interpret these results, we plotted the coefficients of the mixed-effects models. As can be seen in the left graph of Figure 2, in the no-dual-task conditions readers with a higher digit span display signs of a typical bias-inconsistency effect (i.e., bias-inconsistent pronouns are read more slowly than bias-consistent pronouns), in contrast to readers with a lower digit span who display a tendency towards the reversed effect—they process the inconsistent pronouns more quickly than the consistent pronouns. Moreover, although the three-way interactions were not significant, the comparison between the left and right sections of Figure 2 suggests that this particular pattern only holds for the no-dual-task conditions.

A similar pattern emerged three words after the pronoun. As reported in Table 5, in addition to a main effect of composite digit span score (i.e., in first fixation and first gaze), we observed main effects for dual task (first fixation and first gaze) and two-way interactions between the factors pronoun and dual task (first gaze), and dual task

[9]A mixed-effects logistic regression analysis with a maximal subject and item random-effect structure and composite digit span as a continuous predictor revealed that participants with a higher working memory capacity performed better on the secondary task ($\beta = 0.0056$, $SE = 0.0025$, $z = 2.21$, $p < .05$) than participants with a lower working memory capacity.

[10]We did not run likelihood-ratio tests to study the effect of region since this is irrelevant for our research.

[11]The formula of the models used in Experiment 3: dependent measure = 1 + Pronoun × Dual Task × Composite Digit Span + (1 + Pronoun × Dual Task | subject) + (1 + Pronoun × Dual Task | item).

Table 4. *Mean reading times and the probability of regressions at the critical pronoun and the five subsequent words in Experiment 3*

		Sentence region											
		Pronoun		Pro + 1		Pro + 2		Pro + 3		Pro + 4		Pro + 5	
Dependent variable and condition		*Mean*	SE	*Mean*	SE	*Mean*	SE	*Mean*	SE	*Mean*	SE	*Mean*	SE
First-fixation duration													
no dual task	con	212	12	225	7	220	6	214	4	203	5	176	5
	inc	216	14	231	6	222	5	220	6	197	4	189	6
dual task	con	200	7	213	4	213	5	205	6	197	4	176	5
	inc	201	10	222	6	207	4	206	5	198	4	184	5
First-gaze duration													
no dual task	con	212	12	241	8	234	7	230	6	225	6	193	6
	inc	216	14	244	7	238	7	237	8	224	6	204	6
dual task	con	200	7	225	6	227	6	217	7	226	6	185	6
	inc	201	10	239	7	221	6	223	6	224	6	199	7
Right-bound duration													
no dual task	con	212	12	244	8	256	9	243	8	249	7	202	7
	inc	216	14	252	7	246	7	259	9	247	7	217	7
dual task	con	200	7	234	7	242	8	230	7	240	7	193	7
	inc	201	10	245	7	238	8	247	7	244	8	209	8
Regression path duration													
no dual task	con	353	61	287	34	362	26	352	40	346	22	245	13
	inc	325	64	288	14	299	13	363	22	352	21	282	16
dual task	con	229	20	267	15	318	21	304	18	322	24	241	16
	inc	306	42	297	18	345	25	337	21	334	17	253	16
Probability of regression													
no dual task	con	.12	.04	.04	.01	.26	.03	.19	.02	.20	.02	.12	.02
	inc	.06	.03	.07	.02	.19	.03	.23	.03	.23	.02	.15	.02
dual task	con	.07	.03	.06	.01	.17	.02	.17	.02	.17	.02	.11	.02
	inc	.13	.04	.08	.02	.23	.03	.24	.03	.22	.02	.10	.02

Note: Reading times in milliseconds. The values reflect the means (with standard error) aggregated over all data points. Pro = pronoun; con = consistent pronoun; inc = inconsistent pronoun.

and composite digit span (first fixation). Most importantly, the three-way interaction between the fixed factors pronoun and dual task and the continuous predictor composite digit span was reliable in first-fixation, first-gaze, and right-bound durations (see Figures 3–5). In the no-dual-task conditions, readers with a higher digit span showed a pattern that was consistent with a typical implicit causality effect: Words following a bias-consistent pronoun were read more quickly than those following a bias-inconsistent pronoun. On the other hand, for readers with a lower score the opposite seemed to hold. They spent less time processing the third word after bias-inconsistent pronouns than the third word after bias-consistent pronouns. Interestingly, this pattern completely reversed in the dual-task conditions. Whereas lower span readers processed the third word following a bias-consistent pronoun more quickly, higher span readers processed the third word following a bias-inconsistent pronoun more quickly.

For the probability of a regression measure, significant effects emerged three words after the pronoun. The analysis revealed a main effect for dual task ($\beta = 1.2$, $SE = 0.51$, $z = 2.3$, $p < .05$), meaning that readers tended to regress more often from this region when they were exposed to the working memory task than when they were not exposed to the working memory task. However, the significant interaction between dual task and composite digit span ($\beta = -0.016$, $SE = 0.0056$,

Table 5. *Fixed-effects estimates for models of eye-movement measures in Experiment 3*

Measure and fixed effect	*Estimate*	(SE)	t *value*
First-fixation, first-gaze, and right-bound durations on pronoun			
Intercept	5.37	(0.084)	63.8
Pronoun	−0.19	(0.13)	−1.5
Dual task	−0.21	(0.14)	−1.5
Digit span	−0.0020	(0.00098)	−2.1
Pronoun × Dual Task	0.23	(0.17)	1.3
Pronoun × Digit Span	0.0029	(0.0014)	2.0
Dual Task × Digit Span	0.0027	(0.0017)	1.6
Pronoun × Dual Task × Digit Span	−0.0034	(0.0019)	−1.7
First-fixation duration on pro + 3			
Intercept	5.40	(0.065)	83.4
Pronoun	−0.11	(0.075)	−1.4
Dual task	−0.15	(0.064)	−2.3
Digit span	−0.0014	(0.00071)	−2.0
Pronoun × Dual Task	0.16	(0.14)	1.2
Pronoun × Digit Span	0.0016	(0.00081)	1.9
Dual Task × Digit Span	0.0014	(0.00067)	2.0
Pronoun × Dual Task × Digit Span	−0.0023	(0.0011)	−2.1
First-gaze duration on pro + 3			
Intercept	5.47	(0.078)	70.4
Pronoun	−0.15	(0.12)	−1.2
Dual task	−0.18	(0.086)	−2.1
Digit span	−0.0017	(0.00084)	−2.0
Pronoun × Dual Task	0.27	(0.13)	2.1
Pronoun × Digit Span	0.0020	(0.0013)	1.5
Dual Task × Digit Span	0.0017	(0.00095)	1.8
Pronoun × Dual Task × Digit Span	−0.0034	(0.0014)	−2.3
Right-bound duration on pro + 3			
Intercept	5.47	(0.083)	66.3
Pronoun	−0.11	(0.14)	−0.8
Dual task	−0.15	(0.093)	−1.6
Digit span	−0.0013	(0.00091)	−1.4
Pronoun × Dual Task	0.29	(0.16)	1.8
Pronoun × Digit Span	0.0018	(0.0016)	1.2
Dual Task × Digit Span	0.0013	(0.0010)	1.2
Pronoun × Dual Task × Digit Span	−0.0036	(0.0018)	−2.0

Note: All t-values below −2.0 and above 2.0 were significant at $p < .05$.

$z = -2.8$, $p < .01$) showed that this primarily held for readers with a lower digit span score. Since these results do not bear on our main question, we do not further address them in the Discussion section below. However, in our opinion, the finding suggests that lower span readers, while faced with a concurrent task, more often lack the required working memory resources to resolve the pronoun on the fly and instead have to relocate their visual attention to find the proper antecedent in the text.

Discussion

The key finding of Experiment 3 was that the time course of implicit causality varied as a function of the dual task and the working memory capacity of the reader. However, the directions of these interactions were not always as anticipated. Based on previous studies (Long & De Ley, 2000; Van Berkum et al., 2013), the most straightforward expectation would be that readers with a lower working memory capacity are less likely to employ

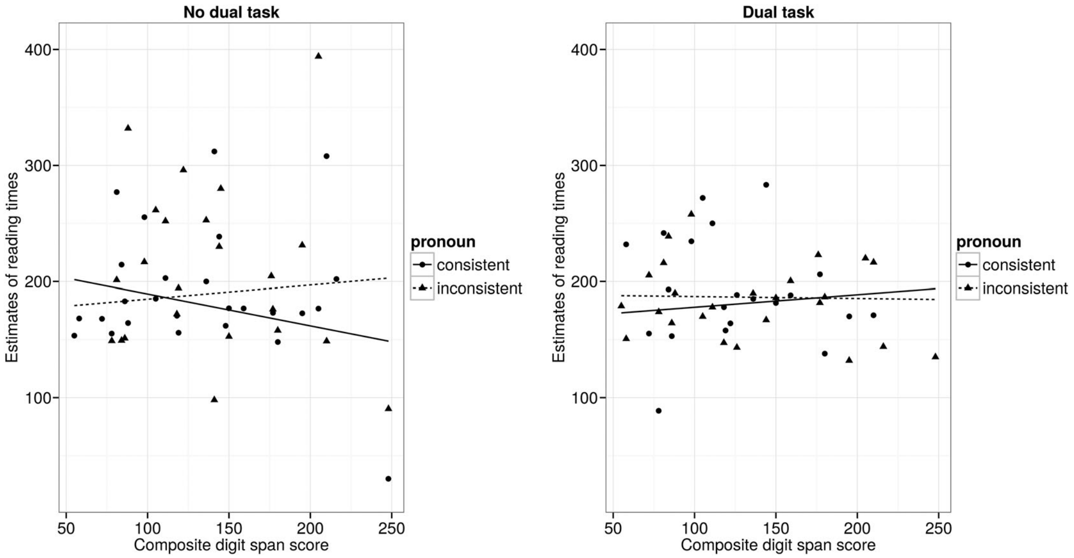

Figure 2. *Reading durations at the critical pronoun in Experiment 3 (the estimates were identical on the pronoun for first-fixation, first-gaze, and right-bound durations). The lines indicate estimates of the mixed-effects model; the points are exponentiated mean values of log-reading times in each composite digit span score. To increase the readability, we removed one data point from the graph in the inconsistent/dual-task condition (composite digit span score = 145, exponentiated mean log-reading time = 782).*

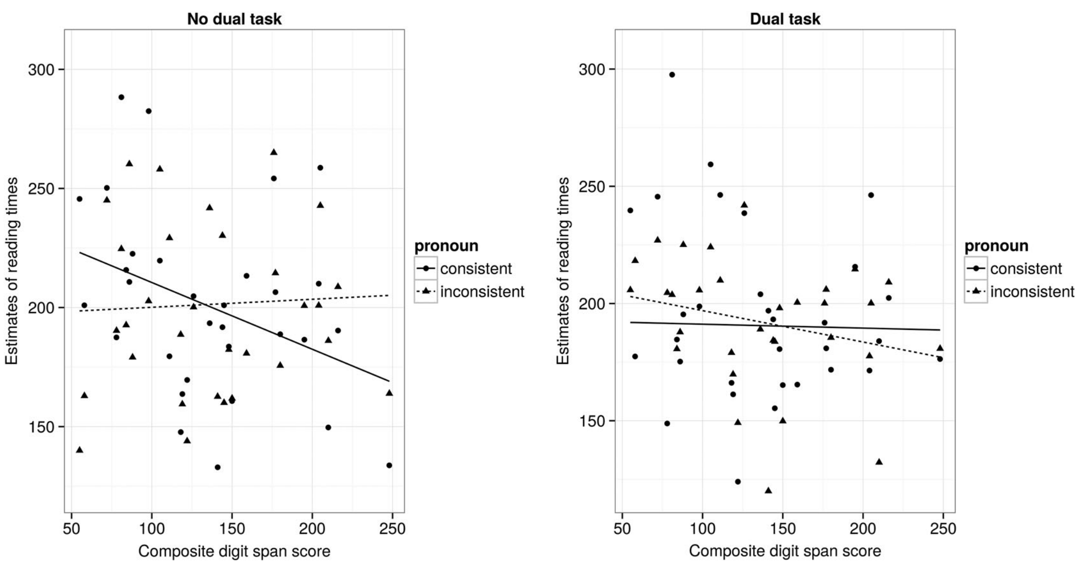

Figure 3. *Estimated first-fixation durations three words after the critical pronoun in Experiment 3. The lines indicate estimates of the mixed-effects model; the points are exponentiated mean values of log-reading times in each composite digit span score.*

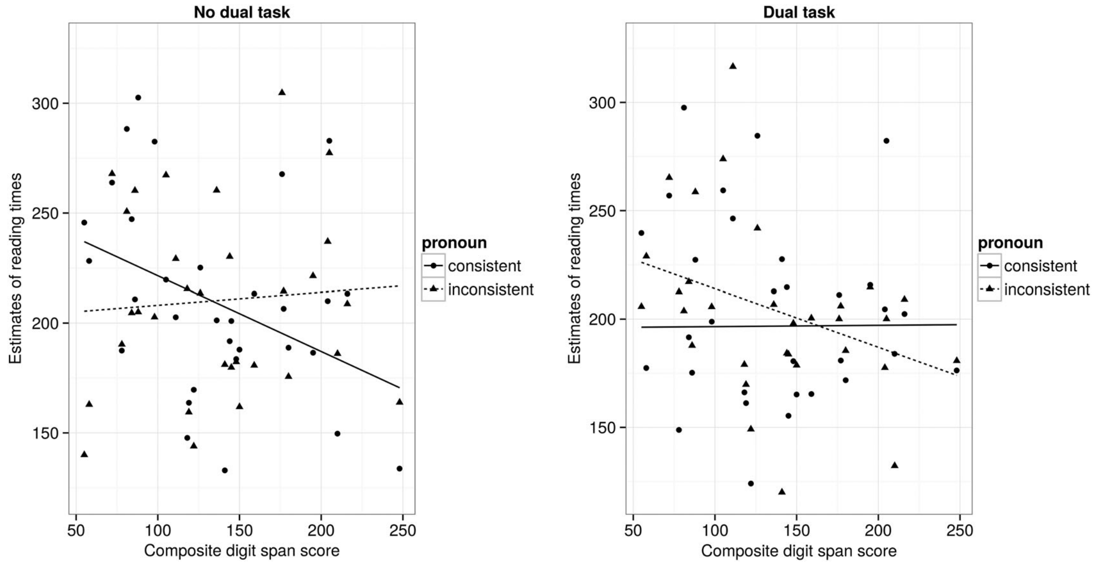

Figure 4. *Estimated first-gaze durations three words after the critical pronoun in Experiment 3. The lines indicate estimates of the mixed-effects model; the points are exponentiated mean values of log-reading times in each composite digit span score.*

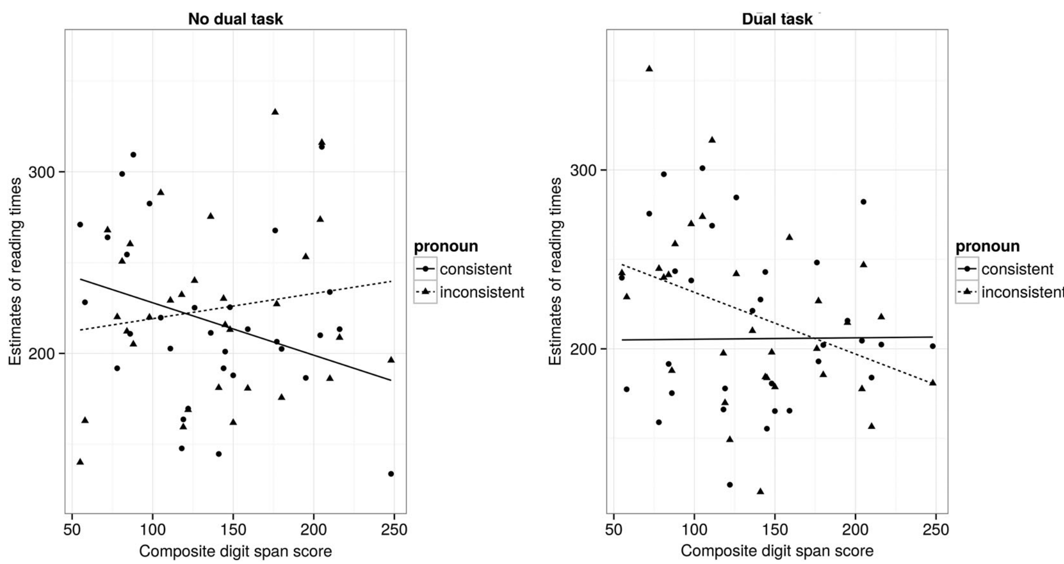

Figure 5. *Estimated right-bound durations three words after the critical pronoun in Experiment 3. The lines indicate estimates of the mixed-effects model; the points are exponentiated mean values of log-reading times in each composite digit span score.*

implicit causality information during online processing. This prediction was only partly borne out by the data. When participants were reading for comprehension without an additional working memory task, the predicted contrast between low- and high-memory-span readers surfaced in the eye-tracking data, namely at the critical pronoun and three words after the pronoun. In addition, we observed a contrast between low- and high-memory-span readers three words after the pronoun if the participants were engaged in a concurrent working memory task. This effect, however, was in the opposite direction. Thus, somewhat surprisingly, in the case of a pressured working memory system, lower span readers displayed a more pronounced usage of implicit causality than higher span readers.

These results speak against accounts in which implicit causality information is (always) activated by a relatively resource-consuming inference (e.g., Long & De Ley, 2000). After all, in that case you would not expect that lower span readers utilize the cue at times when their working memory resources are compromised. In addition, the results show that the usage of implicit causality depends on a delicate interplay of individual and situational (or task-depended) variations and as such raise some intriguing questions on the nature of implicit causality. Specifically, why does the status of implicit causality as a pronoun resolution cue vary in the opposite direction for higher and lower span readers as a function of an additional working memory load?

There are at least two fundamentally different ways of approaching this somewhat puzzling interaction. A possible explanation would be that even though implicit causality information is activated by different types of readers, the underlying mechanisms differ substantially from reader to reader. Although this possibility cannot be refuted definitively, we are somewhat uncomfortable with some of its implications. For example, in order to be able to explain the results of Experiment 3, it seems we need to assume that for higher span readers the influence of implicit causality originates from a costly (causal) inference. For lower span readers, on the other hand, it should constitute an easily activated cue, more likely to surface at times of reduced or even insufficient processing resources. In addition to these somewhat counterintuitive implications, the explanation would also be unsatisfying from a theoretical point of view because it would imply abandoning previous proposals that specifically aim at providing a unified (and parsimonious) account of implicit causality (e.g., Crinean & Garnham, 2006; Hartshorne & Snedeker, 2013; Kehler et al., 2008).

An alternative approach builds on the idea (described in several studies) that implicit causality can be used by readers to generate referential expectations of an unfolding sentence (e.g., Featherstone & Sturt, 2010; Koornneef & Van Berkum, 2006; Van Berkum et al., 2007). Another crucial element is derived from the study of Van Berkum et al. (2013). In their ERP study, they not only showed that a proactive usage of implicit causality is closely tied to the mood of readers but, in addition, proposed that this influence is related to differences in the processing styles of readers. More specifically, readers in a good mood are more likely to anticipate upcoming story characters, as a good mood promotes a broadly oriented, exploratory, and top-down processing strategy. In contrast, readers in a bad mood are more likely to resort to a narrow, conservative, and bottom-up focus (see Van Berkum et al., 2013, and the references therein).

An analogous contrast can be applied to the results of the present study. The main difference between the higher and lower span readers is not how they compute implicit causality information but, rather, in which reading situations it becomes a relevant factor. Under normal circumstances, higher span readers may be relying on proactive top-down processing strategies, but to avoid costly reanalysis in demanding reading situations, they resort to more conservative yet reliable bottom-up strategies. This explains why for higher span readers the influence of implicit causality is more pronounced in the standard, no-dual-task conditions. Lower span readers are displaying the exact opposite behaviour. Hence, their default processing strategy may be more bottom-up in nature, and they will only start to anticipate or guess while facing reduced working memory resources. This perhaps paradoxical reading behaviour has been described previously for older adult readers, who employ a "risky reading" strategy to compensate for their declining cognitive abilities and to maintain an appropriate reading pace (Rayner, Castelhano, & Yang, 2009; Rayner, Reichle et al., 2006).

Of course, this hypothesis based on differences in bottom-up and top-down reading modes requires further testing. Most notably, it warrants a satisfying answer as to why default and coping strategies of higher and lower span readers are different. Nonetheless, the proposal has some clear advantages. It aligns nicely with fairly well-established ideas on implicit causality, and, furthermore, it does not require different viewpoints on the nature of implicit causality as a function of individual differences of the reader (e.g., it being the result of a complex causal inference for high-span readers vs. the output of a quick and dirty heuristic mechanism for low-span readers).

In conclusion, whatever may be the outcome of future studies examining the influence of working memory (and individual differences in general) on the time course of implicit causality, the proposal put forward by Long and De Ley (2000) seems too simplistic. It is not necessarily the case that implicit causality always requires a costly causal inference and consequently can be used in real time by highly proficient readers only. The relationship between implicit causality and the available processing resources of a reader is intrinsically more complex.

GENERAL DISCUSSION

The main purpose of the present study was to provide new and more broadly oriented empirical data on the cognitive and linguistic factors pertaining to the time course of implicit causality. The first main question was whether implicit causality impacts pronoun resolution during online comprehension in the absence of a backward causal connective. In Experiment 1 we showed that this indeed was the case, and, importantly, the results of Experiments 2 and 3 further substantiated this conclusion. Another important general observation was that the time course of implicit causality is easily altered by both linguistic and cognitive factors. With respect to the former, Experiment 2 revealed that although a linguistic cue such as a backward causal connective is not required to trigger its influence, connectives may strengthen implicit causality as a pronoun resolution cue, both in its timing and in its intensity. Furthermore, the results of Experiment 3 highlighted the importance of taking individual differences into account. That is, even though the activation of implicit causality may not be a particularly resource-consuming mental operation, readers with higher and lower working memory capacities reacted differently to a situation where, on top of reading, their working memory resources were burdened with a dual task. Whereas higher span readers seemed more inclined to use implicit causality when they had all their working memory resources at their disposal, lower span readers showed the opposite pattern: They were more sensitive to the cue in the case of an additional working memory load.

Overall, the results of our eye-tracking experiments provided a clear picture of the time course of implicit causality. At the same time, we also observed some variations across experiments that were perhaps more difficult to explain. For example, the influence of implicit causality did not always emerge at the exact same sentence regions. Whereas in Experiment 1 reading-time delays were observed five words after a bias-inconsistent pronoun, the influence of implicit causality in Experiment 3 emerged somewhat earlier, namely three words after the critical pronoun. In addition, the various "early" (first-fixation and first-gaze durations) and "later" eye-tracking measures (right-bound and regression path durations, probability of a regression) did not reveal the exact same pattern across experiments. In Experiment 2, only the regression path and probability of a regression measures (i.e., "later" measures) captured the reading differences between the pronoun-consistent and pronoun-inconsistent conditions. In contrast, in Experiments 1 and 3, both "early" and "later" eye-tracking measures captured these differences.

It is difficult to determine what exactly caused these subtle discrepancies across experiments. One possibility is that, although the participants in the experiments were recruited from the same homogeneous population, the three groups of participants differed significantly in a relevant dimension still. If that was the case, it seems plausible that the time course of implicit causality was not identical across experiments—in particular since Experiment 3 showed that relatively small individual differences may have a strong impact on the usage of implicit causality. Furthermore, whereas the experimental stimuli were held constant in the three experiments, the experiments themselves were not designed as full replication studies. Consequently, they contained different manipulations (e.g., dual task vs. no dual task, connective vs. no connective) and different filler items, all of which may have caused the observed variations in the time course of implicit causality across experiments.

From a more general perspective, it is important to note that variations in the time course of implicit causality not only apply to the series of eye-tracking experiments presented here. In other reading-time studies, a similar variability has been observed. For instance, using self-paced reading and eye-tracking, Koornneef and Van Berkum (2006) reported pronoun-inconsistency effects at the pronoun itself and at the three words following the critical pronoun. In a more recent follow-up study, however, Koornneef and Sanders (2013) observed a more localized effect one to two words after the pronoun. Similarly, Featherstone and Sturt (2010) reported a localized pronoun inconsistency effect at the word immediately following a critical pronoun, but, in addition, they observed a more global inconsistency effect for a pooled region consisting of the pronoun plus the following five words. Together these studies illustrate that, although the influence of implicit causality on pronoun resolution is a robust phenomenon, finer grained characteristics of its time course vary from situation to situation. We further explore potential causes for this fluctuating nature of implicit causality below, where an overview is provided of the implications of our study for various ongoing debates in the implicit causality literature.

Implications for focusing, integration, and the origins of implicit causality

As mentioned in the introduction, a particularly enduring controversy in the literature has been the discussion between proponents of the focusing account and its contender, the integration account. In the focusing account, implicit causality verbs are thought to highlight one of the arguments of the verb at the expense of the other. At the moment readers or listeners encounter a pronoun, it will be most easily connected to the highlighted, more accessible, antecedent. Hence, the focusing properties of the verb are thought to underlie the robust pronoun inconsistency effects reported in so many studies. The integration account, on the other hand, states that the implicit causality information of verbs remains dormant until it is required for the proper interpretation of a sentence or discourse. This is the case, for example, when readers encounter a disambiguating pronoun. So in a sense, in the integration account unambiguous pronouns function as catalysts: They initiate the activation of implicit causality information (cf. Cozijn et al., 2011; Garnham et al., 1996).

Some current proposals seem to move away from the idea that either focusing or integration is the correct approach, but, rather, both notions are required to explain the full spectrum of implicit causality. Whereas the focusing account aims at explaining when implicit causality information and its associated referential bias are activated, the integration account describes how various factors codetermine the persistence and usage of implicit causality information throughout an unfolding sentence (Koornneef & Sanders, 2013; Pyykkönen & Järvikivi, 2010). In the context of this more recent theoretical framework, the current study presents new insights into the specifics of the integration phase. First, backward causal connectives speed up processing when implicit causality becomes a relevant cue during pronoun comprehension. Second, and more speculatively, the processing strategy of readers determines whether implicit causality is used at all.

The implications for the focusing phase are less clear-cut. The stimuli in our study were designed to examine the influence of implicit causality at, and immediately after, the critical pronoun. As a result, we are unable to dissociate the possibility that implicit causality is activated before the pronoun from the possibility that the pronoun itself functions as an implicit causality trigger. Given these limitations, we avoid making strong claims on the focusing properties of implicit causality verbs, other than that a backward causal connective is not required to trigger implicit causality—and hence, in that sense its activation can be traced back to the verb itself.

Another enduring controversy in the literature has been whether implicit causality information is derived from the lexicosemantic properties of a verb or, alternatively, arises as a side-effect of more general cognitive tendencies. To further illustrate the latter possibility, note that in contrast to being a purely lexicosemantic (i.e., linguistic) phenomenon, implicit causality could be the result of a nonspecific pattern recognition system that keeps track of how often events and/or objects tend to co-occur in discourses—or real life for that matter. Another plausible, extralinguistic, explanation is that implicit causality biases are guided by more elaborate long-term memory systems responsible for the storage of our knowledge and beliefs about the world. And, more controversially, it may even be rooted in the "hard-wired" human tendency to always look for causal relations between events (cf. Crinean & Garnham, 2006; Hartshorne & Snedeker, 2013; Kehler et al., 2008; Pickering & Majid, 2007).

Our findings do not necessarily speak in favour of nor against any of these approaches. In that sense we agree with Hartshorne and Snedeker (2013), who in an eloquent defence of a more fine-grained lexicosemantic account concluded that both lexicosemantic and world-knowledge approaches can account for most of the findings on implicit causality. Yet, as noted by these authors, the opposing viewpoints make quite different claims about the underlying processes—and the means by which implicit causality biases are acquired throughout life. Moreover, they also recognized a potential theoretical advantage of the lexicosemantic account. Once one has fully mastered and stored what a verb means, the implicit causality bias comes essentially for free. In more cognitively oriented, extralinguistic approaches, a verb's definition is insufficient, and one has to assume additional mechanisms that, together with the meaning of the verb, give rise to implicit causality. Hence, utilizing the argument of parsimony, one could state that the lexicosemantic account reflects the preferred alternative (but see Koornneef & Sanders, 2013, for a different point of view).

In a similar vein, the empirical results of the current study do not single out one specific account as the correct one, yet some of the results are more naturally explained with a lexicosemantic approach. Most notably, the observation that the activation of implicit causality does not appear to be very demanding in terms of working memory resources is consistent with the idea that implicit causality biases come more or less for free, an assumption most clearly articulated within the lexicosemantic framework. Furthermore, the robust finding that implicit causality effects emerge even in the absence of a connective is more easily explained in a lexicosemantic framework that by default assumes a particular bias for each verb. World knowledge and/or probabilistic accounts need to posit a more complex chain of mental operations. By way of illustration, consider the probabilistic framework of Kehler et al. (2008) who solve the issue by taking a two-step approach. They proposed that, regardless of the connective, implicit causality verbs initially trigger an expectation of an explanation of the implicit causality event, and this expectation in turn creates a causal bias towards one of the referents.

On the other hand, any account that identifies lexicosemantic information as the primary source of implicit causality has to provide a satisfying explanation for the sometimes subtle variations in the time course of implicit causality, as observed in the present study. This appears to be more of a challenge. For instance, the observation that implicit causality is used more rapidly in the presence of a backward causal connective is somewhat problematic for an approach assuming a fully verb-based activation of implicit causality. Although a lexicosemantic account can ultimately explain these results, for instance, by defining multiple potential (thematic) roles for the arguments of a verb, which become more or less prominent as a consequence of connectives (see Hartshorne & Snedeker, 2013), these type of accounts clearly require some stipulations. A probabilistic framework along the lines of Kehler et al. (2008) by its very nature incorporates a graded, discourse-dependent, implicit causality bias and hence is more flexible in its description and explanation of subtle differences in the timing and strength of implicit causality.

A similar line of reasoning holds for the hypothesis that the time course of implicit causality may change as a function of the processing strategy of readers. More specifically, if the semantics of verbs offer the bias for free, why would readers use it differently in high and low working memory load situations, or when they are happy or sad (Van Berkum et al., 2013)? Of course, a way out is to assume that although the activation of the bias comes for free, the way people eventually use it during pronoun comprehension varies as a function of their processing strategy (cf. our discussion of Experiment 3). However, again note that a fluctuating use of implicit causality is at the very heart of more cognitively oriented, probabilistic approaches, and, hence, these frameworks will provide a more parsimonious explanation.

In all, the current study offers some interesting new empirical data that any account on implicit causality needs to explain. At this point in time, however, we are not in the position to single out a unique framework as the correct approach. In particular, questions pertaining to the processing strategy of readers and the origins of implicit causality require a definite answer still. With respect to these open issues, we agree with Hartshorne and Snedeker (2013) who emphasize that important steps can be taken by studying how implicit causality is acquired throughout life. In addition, a thorough examination of the influence of individual differences on the usage of implicit causality in text comprehension will be another indispensable line of research to solve the issues. Combined with the very promising theoretical advances currently made in (formal) semantics (Bott & Solstad, 2014; Hartshorne & Snedeker, 2013), we should soon arrive at a more grounded decision on whether implicit causality is rooted in the stored meaning of a verb, or arises as a side-effect of more general cognitive biases. In the long run, this will also help us to decipher why readers differ in their usage of the (overt and covert) coherence cues offered by a text.

REFERENCES

Ariel, M. (2001). Accessibility theory: An overview. In T. Sanders, J. Schilperoord, & W. Spooren (Eds.), *Text representation, linguistic and psycholinguistic aspects* (pp. 29–87). Amsterdam: John Benjamins.

Arnold, J. E. (2010). How speakers refer: The role of accessibility. *Language and Linguistics Compass*, *4*(4), 187–203.

Au, T. K. F. (1986). A verb is worth a thousand words: The causes and consequences of interpersonal events implicit in language. *Journal of Memory and Language*, *25*(1), 104–122.

Baayen, R. H., Davidson, D. H., & 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., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*, 255–278.

Boland, J. E. (2004). Linking eye movements to sentence comprehension in reading and listening. In M. Carreiras, C. Clifton Jr. (Eds.), *The On-Line Study of Sentence Comprehension: Eye-tracking, ERPs and Beyond* (pp. 51–76). New York: Psychology Press.

Bott, O., & Solstad, T. (2014). From verbs to discourse: A novel account of implicit causality. In B. Hemforth, B. M. Mertins, & C. Fabricius-Hansen (Eds.), *Psycholinguistic approaches to meaning and understanding across languages* (pp. 219–251). Cham, Switzerland: Springer.

Brown, R., & Fish, D. (1983). The psychological causality implicit in language. *Cognition*, *14*(3), 237–273.

Canestrelli, A. R., Mak, W. M., & Sanders, T. J. (2013). Causal connectives in discourse processing: How differences in subjectivity are reflected in eye movements. *Language and Cognitive Processes*, *28*(9), 1394–1413.

Clifton, C., Jr., Staub, A., & Rayner, K. (2007). Eye movements in reading words and sentences. In R. van Gompel (Ed.), *Eye movements: A window on mind and brain* (pp. 341–372). Amsterdam: Elsevier.

Cozijn, R. (2000). *Integration and inference in understanding causal sentences* (Unpublished doctoral dissertation). Tilburg University, Tilburg, Netherlands.

Cozijn, R., Commandeur, E., Vonk, W., & Noordman, L. G. (2011). The time course of the use of implicit causality information in the processing of pronouns: A visual world paradigm study. *Journal of Memory and Language*, *64*(4), 381–403.

Crinean, M., & Garnham, A. (2006). Implicit causality, implicit consequentiality and semantic roles. *Language and Cognitive Processes*, *21*(5), 636–648.

Daneman, M., & Merikle, P. M. (1996). Working memory and language comprehension: A meta-analysis. *Psychonomic Bulletin & Review*, *3*(4), 422–433.

Ehrlich, K. (1980). Comprehension of pronouns. *The Quarterly Journal of Experimental Psychology*, *32*(2), 247–255.

Featherstone, C. R., & Sturt, P. (2010). Because there was a cause for concern: An investigation into a word-specific prediction account of the implicit-causality effect. *The Quarterly Journal of Experimental Psychology*, *63*(1), 3–15.

Garnham, A. (2001). *Mental models and the interpretation of anaphora*. Hove: Psychology Press.

Garnham, A., Traxler, M., Oakhill, J., & Gernsbacher, M. A. (1996). The locus of implicit causality effects in comprehension. *Journal of Memory and Language*, *35*(4), 517–543.

Garvey, C., & Caramazza, A. (1974). Implicit causality in verbs. *Linguistic Inquiry*, 5, 459–464.

Garvey, C., Caramazza, A., & Yates, J. (1975). Factors influencing assignment of pronoun antecedents. *Cognition*, *3*(3), 227–243.

Gathercole, S. E., & Baddeley, A. D. (1993). *Working memory and language processing*. New York: Psychology Press.

Gordon, P. E., Hendrick, R., & Foster, K. L. (2000). Language comprehension and probe list memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 766–775.

Greene, S. B., & McKoon, G. (1995). Telling something we can't know: Experimental approaches to verbs exhibiting implicit causality. *Psychological Science*, *6*(5), 262–270.

Hartshorne, J. K., & Snedeker, J. (2013). Verb argument structure predicts implicit causality: The advantages of finer-grained semantics. *Language and Cognitive Processes*, *28*(10), 1474–1508.

Kamalski, J., Lentz, L., Sanders, T., & Zwaan, R. A. (2008). The forewarning effect of coherence markers in persuasive discourse: Evidence from persuasion and processing. *Discourse Processes*, *45*(6), 545–579.

Kehler, A., Kertz, L., Rohde, H., & Elman, J. L. (2008). Coherence and coreference revisited. *Journal of Semantics*, *25*(1), 1–44.

Kessels, R. P., van den Berg, E., Ruis, C., & Brands, A. M. (2008). The backward span of the Corsi Block-Tapping Task and its association with the WAIS-III Digit Span. *Assessment*, *15*(4), 426–434.

Kessels, R. P., van Zandvoort, M. J., Postma, A., Kappelle, L. J., & de Haan, E. H. (2000). The Corsi block-tapping task: Standardization and Normative data. *Applied Neuropsychology*, *7*(4), 252–258.

Koornneef, A. W., & Sanders, T. J. (2013). Establishing coherence relations in discourse: The influence of implicit causality and connectives on pronoun resolution. *Language and Cognitive Processes*, *28*(8), 1169–1206.

Koornneef, A. W., & Van Berkum, J. J. (2006). On the use of verb-based implicit causality in sentence comprehension: Evidence from self-paced reading and eye tracking. *Journal of Memory and Language*, *54* (4), 445–465.

Long, D. L., & De Ley, L. (2000). Implicit causality and discourse focus: The interaction of text and reader characteristics in pronoun resolution. *Journal of Memory and Language*, *42*(4), 545–570.

Majid, A., Sanford, A. J., & Pickering, M. J. (2007). The linguistic description of minimal social scenarios affects the extent of causal inference making. *Journal of Experimental Social Psychology*, *43*(6), 918–932.

Mak, W. M., Tribushinina, E., & Andreiushina, E. (2013). Semantics of connectives guides referential expectations in discourse: An eye-tracking study of Dutch and Russian. *Discourse Processes*, *50*(8), 557–576.

McDonald, J. L., & MacWhinney, B. (1995). The time course of anaphor resolution: Effects of implicit verb causality and gender. *Journal of Memory and Language*, *34*(4), 543–566.

Millis, K. K., & Just, M. A. (1994). The influence of connectives on sentence comprehension. *Journal of Memory and Language*, *33*(1), 128–147.

Mueller, S. T. (2013). The psychology experiment building language (Version 0.13) [Software]. Retrieved from http://pebl.sourceforge.net

Noordman, L. G., & Vonk, W. (1998). Memory-based processing in understanding causal information. *Discourse Processes*, *26*(2–3), 191–212.

Pickering, M. J., & Majid, A. (2007). What are implicit causality and consequentiality? *Language and Cognitive Processes*, *22*(5), 780–788.

Pyykkönen, P., & Järvikivi, J. (2010). Activation and persistence of implicit causality information in spoken language comprehension. *Experimental Psychology*, *57*(1), 5–16.

Pyykkönen, P., Matthews, D., & Järvikivi, J. (2010). Three-year-olds are sensitive to semantic prominence during online language comprehension: A visual world study of pronoun resolution. *Language and Cognitive Processes*, *25*(1), 115–129.

Rayner, K., Castelhano, M. S., & Yang, J. (2009). Eye movements and the perceptual span in older and younger readers. *Psychology and Aging*, *24*(3), 755–760.

Rayner, K., Chace, K. H., Slattery, T. J., & Ashby, J. (2006). Eye movements as reflections of comprehension processes in reading. *Scientific Studies of Reading*, *10*(3), 241–255.

Rayner, K., Reichle, E. D., Stroud, M. J., Williams, C. C., & Pollatsek, A. (2006). The effect of word frequency, word predictability, and font difficulty on the eye movements of young and older readers. *Psychology and Aging*, *21*(3), 448–465.

Rigalleau, F., Guerry, M., & Granjon, L. (2014). The effect of connectives on the selection of arguments: Implicit consequentiality bias for the connective "but". *Journal of Psycholinguistic Research*, *43*(4), 465–485.

Sanders, T. J. M., & Canestrelli, A. R. (2012). The processing of pragmatic information in discourse. In H. Schmid (Ed.), *Cognitive pragmatics*, Handbook of pragmatics (Vol. 4, pp. 201–232). Berlin: Mouton de Gruyter.

Stevenson, R. J., Crawley, R. A., & Kleinman, D. (1994). Thematic roles, focus and the representation of events. *Language and Cognitive Processes*, *9*(4), 519–548.

Stewart, A. J., Pickering, M. J., & Sanford, A. J. (2000). The time course of the influence of implicit causality information: Focusing versus integration accounts. *Journal of Memory and Language*, *42*(3), 423–443.

Van Berkum, J. J., Brown, C. M., Zwitserlood, P., Kooijman, V., & Hagoort, P. (2005). Anticipating upcoming words in discourse: Evidence from ERPs and reading times. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(3), 443–467.

Van Berkum, J. J., De Goede, D., Van Alphen, P. M., Mulder, E. R., & Kerstholt, J. H. (2013). How robust is the language architecture? The case of mood. *Frontiers in Psychology*, 1–19. doi:10.3389/fpsyg.2013.00505

Van Berkum, J. J., Koornneef, A. W., Otten, M., & Nieuwland, M. S. (2007). Establishing reference in language comprehension: An electrophysiological perspective. *Brain Research*, *1146*, 158–171.

Wilde, N. J., Strauss, E., & Tulsky, D. S. (2004). Memory span on the Wechsler scales. *Journal of Clinical and Experimental Neuropsychology*, *26*(4), 539–549.