

Implicit causality bias in English: a corpus of 300 verbs

Evelyn C. Ferstl · Alan Garnham ·
Christina Manouilidou

Published online: 24 November 2010
© Psychonomic Society, Inc. 2010

Abstract This study provides implicit verb causality norms for a corpus of 305 English verbs. A web-based sentence completion study was conducted, with 96 respondents completing fragments such as “John liked Mary because...” The resulting bias scores are provided as supplementary material in the Psychonomic Society Archive, where we also present lexical and semantic verb features, such as the frequency, semantic class and emotional valence. Our results replicate those of previous studies with much smaller numbers of verbs and respondents. Novel effects of gender and its interaction with verb valence illustrate the type of issues that can be investigated using stable norms for a large number of verbs. The corpus will facilitate future studies in a range of areas, including psycholinguistics and social psychology.

Keywords Psycholinguistics · Verbs · Thematic roles · Causality · Corpus studies

Language comprehension involves the interaction of a large number of factors. Texts are usually about interrelated sequences of *events* or *states*, which are expressed linguistically by specific *verbs*. Typically, both events and states have causes. To understand a narrative properly it is necessary to compute the causal relations between the events and the states described in it (Graesser, Singer, & Trabasso, 1994), and narrativization has been seen as a

mechanism to make sense of the causal structure of the world around us (e.g., Bruner & Feldman, 1993).

Both causes and causal relationships can be stated explicitly. In (1) below the cause is explicit, but the relationship needs to be inferred. In addition, the causal relationship is signalled linguistically in (2), by a connective such as ‘because.’

- (1) Kate quit her job. She never liked her boss.
- (2) Kate quit her job, because she never liked her boss.

Alternatively, causes may remain *implicit*, particularly when they are not important for the development of the narrative (Garnham, 2001). An event or state is said to have an *implicit* cause when the way it is described suggests, but does not explicitly state, how it was caused. For example, if *John frightened Mary*, it is unlikely that one can guess the exact reason for this (e.g., ‘because he jumped out the bushes’); what is more likely to guess is that it is *John* who did something to cause Mary to be in the state of fear.

Implicit causality has usually been associated with the causal directionality contained in the meanings of interpersonal verbs (Garvey & Caramazza, 1974). Verbs that give rise to inferences that would assign the cause to the subject of a simple active sentence of the form *NP1 verb NP2*, and thus to the first noun phrase, are usually called NP1-biased. When the cause is assigned to the object, the verbs are referred to as NP2-biased. The term “bias” is used because when implicit causality is measured by asking people to add explicit causes to statements containing interpersonal verbs or to make judgments about causality, the results are not completely consistent, but show a preponderance of responses favoring either the NP1 or the NP2.

The effects of implicit causality in sentence comprehension and production have been manifested with great regularity across different research paradigms, across

Electronic supplementary material The online version of this article (doi:10.3758/s13428-010-0023-2) contains supplementary material, which is available to authorized users.

E. C. Ferstl (✉) · A. Garnham · C. Manouilidou
University of Sussex, School of Psychology,
Falmer, Brighton BN1 9QH, UK
e-mail: e.c.ferstl@sussex.ac.uk

different languages and cultures, and for children as well as adults (for a review, see Rudolph & Försterling, 1997). For instance, in psycholinguistics, implicit causality is known to play a role in the comprehension of discourse, since the causal inferences reflect part of the general knowledge one must have access to in order to grasp the meaning of the text.

Implicit causality has been shown to have an effect in on-line comprehension tasks, for example in timed reading tasks or plausibility judgments (Caramazza, Grober, Garvey, & Yates, 1977; McKoon, Greene, & Ratcliff, 1993). The results of these studies suggest that when the second clause is consistent with the verb implicit causality bias, as in (3), then comprehension is faster than when the second clause is inconsistent with the verb implicit causality bias, as in (4).

- (3) Kate praised Liam because he had done well in his exams.
 (4) Kate praised Liam because she felt obliged to do so.

This effect is known in the literature as the *congruency* effect (e.g., Carreiras, Garnham, & Oakhill, 1996; Garnham & Oakhill, 1985; Garnham, Oakhill, & Cruttenden, 1992).

The nature of implicit causality and its role in language comprehension have been widely investigated within diverse areas of psychology. Apart from the speculative formulation of the concept of implicit causality at an intuitive level, there have been attempts to attribute the specific implicit causality verb bias to certain verb properties, including *semantic class* and *verb thematic roles*.

Initially, the idea of attributing implicit causality to verb thematic roles was rejected (Garvey, Caramazza, & Yates, 1974–1975). However, Brown and Fish (1983) later attempted to explain verb implicit causality bias by dividing verbs into *actions* and *states* and by stating different predictions about the implicit causality bias for these two classes of verbs. This distinction, although useful, is still very broad. Thus, Brown and Fish (1983) attempted a verb-implicit causality taxonomy based on verb thematic roles. More specifically, they suggested that action verbs attribute causes to the *Agent*, while mental state verbs impute causes to the *Stimulus* rather than the *Experiencer*. This line of research has been followed by other researchers working on implicit causality, and it has led to a finer grained taxonomy, called *revised action-state taxonomy*. For instance, Rudolph (1997) and Rudolph and Försterling (1997) divided interpersonal verbs in four categories depending on their thematic roles: *Agent-Patient* (AgPat) as in (5), *Agent-Evocator* (AgEvo) as in (6), *Stimulus-Experiencer* (StimExp) as in (7) and *Experiencer-Stimulus* (ExpStim) as in (8).

- (5) Paul_{Ag} kissed Alexia_P because he liked her.
 (6) Philippa_{Ag} criticized Frank_E because he had behaved inappropriately.

- (7) George_{Stim} charmed Pauline_{Exp} because he was so polite.
 (8) Peter_{Exp} adored Zoe_{Stim} because she was so sweet.

This taxonomy, which is effectively the same as that suggested by Au (1986), who noted that there is a subset of action verbs that has *Patient/theme* causality, has been widely adopted. There are numerous studies showing that there is indeed a relation between the implicit causes of an event and the thematic roles of the verb that encode the specific event (cf. Crinean & Garnham, 2006). Using 100 Spanish verbs, Goikoetxea, Pascual, and Acha (2008) confirmed in both children and adults that AgPat and StimExp were biased towards NP1, and AgEvo and ExpStim verbs towards NP2. This difference was more pronounced for the state verbs than for the action verbs.

In psycholinguistic research, implicit causality has often been defined as a lexical property of the verb. However, there are also suggestions that the bias reflects general world knowledge about typical causes and relationships in the world (see Rudolph & Försterling, 1997, for a more detailed review of theories of implicit causality). These theoretical approaches can be distinguished in studies taking into account contextual factors and individual differences. The hypothesis is that a lexical property is rather stable, whereas a script-based, general knowledge feature can be more easily influenced by the context of an event. And indeed, it has been shown that perceived causality takes into account the semantic connotations not only of the verbs denoting states and actions, but also of the nouns denoting the participants (Corrigan, 2001; Garvey & Caramazza, 1974). In particular, Corrigan (2001) showed that some nouns are rated more agentive or potent than others (e.g., *warrior* vs. *grandmother*), and thus, that they are more likely to be considered instigators of events.

Applying this finding to the minimal contexts used in sentence completion studies, LaFrance, Brownell, and Hahn (1997) argued that the gender of noun phrases should influence implicit causality attribution. And indeed, in a series of experiments they confirmed that men were more likely to be considered to be causing actions in mixed gender events. Similarly, in the recipient role, women were more likely to be seen to elicit the actions of others. These strong effects were independent of the social status, and they were obtained using explicit ratings. However, in sentence completion tasks, a corresponding gender effect has not yet been documented (see Goikoetxea et al., 2008).

In addition to the influence of the properties of the participants in an interpersonal event, previous studies have also paid attention to emotional *valence* (negative vs. positive) as a factor that can possibly affect implicit causality bias (e.g. Franco & Arcuri, 1990; Rudolph, 2008; Rudolph

& Försterling, 1997). Franco and Arcuri (1990) have shown that NP1 bias is more likely for negative verbs than positive verbs, at least for a subset of their materials. In contrast, semantic valence did not prove to be systematically associated with implicit causality biases in other studies (see Rudolph & Försterling, 1997: 203).

Finally, individual differences in implicit causality attribution have been studied. Although not often investigated, LaFrance et al. (1997) found complex interactions between protagonist gender and participant gender (see also Mannetti & de Grada, 1991).

There are many questions about implicit causality that are currently under investigation. One such question is whether implicit causality has an early focusing effect (e.g., Koornneef & van Berkum, 2006; Long & De Ley, 2000; McDonald & MacWhinney, 1995; or a later effect on clausal integration (Garnham, Traxler, Oakhill, & Gernsbacher, 1996; Stewart, Pickering, & Sanford, 2000). Recent evidence from comprehension tasks using event-related potentials (van Berkum, Koornneef, Otten, & Nieuwland, 2007) and the visual world paradigm (Pyykkönen & Järvikivi, 2010) seems to favor an early effect, either due to focusing or immediate integration. Note, however, that in production, implicit causality biases do not readily yield focusing effects, at least not as reflected in a particular choice of referential expression (Fukumura & van Gompel, 2010).

Another issue is whether the basic phenomenon of implicit causality is best characterized as all-or-none (e.g., for ExpStim verbs, Stim IS the implicit cause, Crinean & Garnham, 2006), with biases arising through additional factors when explicit causes are generated, or as graded (Pickering & Majid, 2007), so that a bias is directly associated with the verb in addition to its thematic structure.

In addition, implicit causality can be used as the basis of studies on the role of pragmatics and world knowledge in inference making. Furthermore, because of the interpersonal nature of the verbs studied and the link with causal attribution, implicit causality can be used as the basis of social psychology studies, for example to shed light on gender relations, group processes and cultural stereotypes (see Holtgraves & Kashima, 2008). These and related questions will be easier to address with a large set of verb norms.

The present study

In order to carry out studies of the effect of implicit causality in sentence comprehension and production, it is necessary to have normative data on the implicit causality of specific verbs. However, these data sometimes comprise

the experimenter's intuitions, and in other cases are based on small numbers of verbs and small numbers of observations per verb. As noted in Rudolph and Försterling (1997), many studies tend to use the same verbs as in Brown and Fish (1983) or other early studies. In order to carry out better, replicable studies of verb causality effects, norms are needed for a larger set of verbs, based on responses by more people. Such a set of normative data has recently been provided for Spanish (Goikoetxea et al., 2008). However, most work on implicit causality has been carried out in English, and it is likely that a substantial amount of future work will also be carried out in English. A stable set of implicit causality norms for a large set of verbs is, therefore, required.

Thus, given the small number of verbs in previous English norming studies, the main purpose of the present study is to offer normative data to researchers investigating implicit causality based on a larger corpus and a broader range of verbs. A variety of verbs was collected from various sources (see Materials section). Factors that have been shown to affect implicit causality were carefully taken into account as well as additional semantic features and verb frequency and length. To this end, a sentence completion experiment was carried out using more than 300 verbs.

The method most frequently used to investigate implicit causality, and the one we used to collect our own normative data, is the sentence completion task, a language production task. Participants are asked to provide an explicit cause for an event for which the cause is at this point implicit. The sentence to be completed usually looks like example (9), where the linguistic signal 'because' is added to the end of the description of an interpersonal event.

(9) Kate accused Mary because she ...

It is well known that the use of different connectives influences the choice of continuations (e.g., Ehrlich, 1980; Stevenson, Crawley, & Kleinman, 1994; Stevenson, Knott, Oberlander, & McDonald, 2000). For example, *so* tends to elicit consequences rather than causes. However, including *because* creates a context that ensures that a high proportion of continuations explicitly mention causes.

In (9) the two protagonists in the event are of the same sex. The use of the pronoun is intended to force the participants to refer back to one of the protagonists and, hopefully, to choose one entity as the main cause of the event. One difficulty with this method is that the resulting continuations are often ambiguous. A method to reduce ambiguity is to ask the participants to circle the person their sentence completion referred to (e.g., Goikoetxea et al., 2008).

In an alternative version of the task, which we employed for the present study, sentence fragments such as (10) were presented:

(10) Kate accused Bill because...

In these fragments a mixed gender pair participated in the interpersonal event, and the pronoun is omitted. If the sentence completion begins with a pronoun, it clearly signals which entity has been selected as the cause.

We used a web-based version of the sentence completion task to assess the implicit verb causality (IVC) bias. The sentence completions were scored as NP1 or NP2 completions using a semi-automatic procedure. First, all completions containing a pronoun (*he* or *she*) were categorized as referring to the female or male protagonist, and thus to NP1 or NP2. The few remaining completions were scored by two independent raters.

To evaluate context effects and response strategies we included the gender of the protagonist, as well as the gender of the participants in our analyses. The questions of interest were (1) whether male protagonists would be chosen more often as the causers of events than female protagonists, (2) whether such a difference would be modulated by the valence of the event, and (3) whether men and women would use different strategies for attributing causation.

In addition to these novel aspects, several reliability analyses were conducted to ensure comparability of our results with previously published verb bias data. Furthermore, a replication of the well-described differences in implicit verb causality for the four semantic categories was attempted.

The complete corpus is available as supplemental material in the Psychonomic Society Archive. The numbers of completions are presented separately for male and female subjects, and as a function of whether the first noun phrase was male or female. In addition, bias scores expressing the strength of the implicit verb causality are provided (see Results section). To facilitate on-line studies, we also list lexical and semantic features, including the verb frequency (CELEX, Baayen, Piepenbrock, & Gulikers, 1995), the word length, valence ratings, and the categorization into the four thematic/semantic classes according to the revised state-action taxonomy.

Methods

Selection and characterization of the corpus

The initial corpus consisted of 318 verbs. One hundred nine verbs were taken from previous studies on verb bias: 77 from Au (1986), 49 from Crinean and Garnham (2006; collected by Stewart, Pickering, & Sanford, 1998; Stewart et al., 2000)

and 16 verbs from Rudolph (2008), with some verbs used in more than one of the previous studies. These 109 verbs were classified according to Levin's (1993) semantic classes. Based on Levin's (1993) classification, the original set of verbs was expanded using verbs with similar properties. For instance, the verb *charm*, which was included in Au's (1986) study, belongs to Levin's (1993) 'Amuse-type psychological verbs.' Subsequently, we selected as many Amuse-type psychological verbs as possible and included them in the corpus. This procedure was applied to all semantic classes.

Before proceeding to the actual experiment an initial pre-selection was carried out. This resulted in the exclusion of 13 verbs that were either very low frequency or produced odd interpersonal or colloquial connotations when used in sentences (e.g., *lull*, *disgruntle*, *boggle*, *elate*, *awe*, *afflict*, etc.). Following this pre-selection, 305 transitive verbs were used in the sentence completion task.

The 305 English interpersonal transitive verbs were further characterized using the features length, frequency, emotional valence, semantic class and thematic roles. Descriptive statistics for these features are displayed in Table 1. They were defined as follows:

For each verb, the word frequency was determined using the CELEX database (Baayen et al., 1995). For all calculations, the log frequency was used. The word length was defined as the number of characters. For 17 verbs, this included the length of the preposition (e.g., *apologize to*). As expected, word length and frequency were negatively correlated ($r = -.17$, $n = 305$, $p < .01$), indicating that longer words were less frequent.

The emotional valence of the verbs was determined using the ratings of 12 native speakers of British English. The participants were instructed to rate for each verb whether they associated with it negative or positive feelings. They were given a 7-point rating scale (ranging from -3: extremely negative, to +3: extremely positive). The valence ratings ($M = -.35$, $SD = 1.6$) were not correlated with either length or frequency.

To encode the different *semantic classes* the verbs were divided into *activity verbs* (e.g., *kiss*) and *psychological verbs* (e.g., *love*), including a small number of mental state and perception verbs. The two categories of verbs were further classified with respect to their thematic roles. Following the psycholinguistic literature on implicit causality (Au, 1986; Crinean & Garnham, 2006; Rudolph, 1997; Rudolph & Försterling, 1997) we divided *activity verbs* into two distinct categories, < Agent < Patient >> (AgPat) verbs (e.g., *carry*) and < Agent < Evocator >> (AgEvo) verbs (e.g., *praise*). Similarly, *psychological verbs* were divided into < Experiencer < Stimulus >> (ExpStim) verbs (e.g., *love*) and into < Stimulus < Experiencer >> (StimExp) verbs (e.g., *upset*). Table 1 presents descriptive statistics for all these features.

Table 1 Descriptive statistics of the verb corpus, and for each of the four linguistic categories

		Activity verbs		Psychological verbs		All verbs
		AgPat	AgEvo	StimExp	ExpStim	
Word length (no. of letters)	N	96	55	110	44	305
	M	7.7	8.4	8.2	7.8	8.0
	SD	2.1	1.8	1.8	1.9	1.9
	Range	3 -14	4 - 13	4-13	5-13	3 - 14
Log frequency (CELEX)	M	1.4	1.2	1.2	1.7	1.3
	SD	.9	.8	.8	.7	.8
	Range	0 – 3.0	0 – 2.8	0 - 3.0	0 – 3.0	0 - 3.0
Valence ratings	M	-.21	-.55	-.46	-.13	-.35
	SD	1.4	1.7	1.6	1.9	1.6
	Range	-2.8 - +2.5	-2.7 - +2.7	-2.9 - + 2.5	-2.8 - + 2.5	-2.9 - +2.7
Bias score	M	2.3	-33.4	46.1	-63.5	2.2
	SD	47.7	37.4	38.2	29.8	56.2
	Range	-96 - +82	-94 - +93	-77 - +87	-100 - +30	-98 - +93

One-way ANOVAs were conducted for each of the three features to compare the linguistic classes. The verb categories were well matched with respect to valence ($F(3, 301) = 1.0$), but differed in frequency ($F(3, 301) = 4.0$, $p < .01$). There was also a tendency for a length difference ($F(3, 301) = 2.5$, $p = .06$). AgPat and ExpStim verbs were slightly shorter and more frequent than AgEvo and StimExp verbs. Because of this difference, length and frequency will be included in subsequent analyses as covariates.

Experimental materials

For creating sentence fragments, 305 female and 305 male common British English names were used. The names were chosen from the “British names” section of the website “Baby names world.” Two native speakers of British English confirmed that all names were clearly unambiguous in terms of gender and that they did not sound old-fashioned or bizarre.

One male and one female proper name was randomly assigned to each verb. For each verb we created two sentence fragments, one with the male name in NP1 position (“*M* *verbed* *F* *because* ...”), and one with the female name in NP1 position (“*F* *verbed* *M* *because*...”). For counterbalancing, one list was created with half of the sentences having a male NP1 and half a female NP1, and a second list was created by switching the proper names in each sentence fragment. Each of the two lists was presented in two different orders; thus, there were four different versions of the experiment. Due to a technical error two words were presented twice in each

list. Only the first occurrence was used for these two verbs.

Participants

Ninety-six participants (52 women, 44 men) completed the entire questionnaire and were included in the study. Their ages ranged from 18 to 57 years ($M = 20.7$, median = 19, $SD = 4.9$). They were all first or second year undergraduate students at the University of Sussex, and they received course credits or a small monetary reimbursement for their participation.

Procedure

The data collection was conducted using a web-based questionnaire programmed with the software *Dreamweaver 8*. Participants were contacted by e-mail and provided with a website where they could download the questionnaire, which was stored on the University of Sussex webserver. Each participant was randomly assigned one of the four versions of the experiment. The participants were instructed to type a sensible completion of these sentences, similar to the short examples provided to them (e.g., “The lion ate the zebra, because ... it was hungry”). They were also instructed to answer spontaneously and complete each sentence at once without going back and revising. Halfway through the questionnaire, the participants had the option to log off and continue at some later point or to proceed with the completion of the second half. There was no time limit. After the completion of the questionnaire, the participants were notified that their task was over and they had to press

the “Submit” button in order to send their data to the server. The completion of the entire questionnaire lasted between 45 and 70 min, depending on the participant’s response speed.

Coding

The participants left 1.5% of the continuations blank. A large majority of the responses (91.2%) started with either of the pronouns *he* or *she*, so that they could be automatically scored as either NP1 or NP2. The remaining 2,139 (7.3%) responses were examined manually in order to detect cases where causation was clearly attributed to either NP1 or NP2, but did not start with a pronoun (e.g., *John admired Mary because nothing seemed to faze her*). Two raters classified each of the responses as either NP1 or NP2, or as “other.” Examples for “other” responses were continuations starting with the pronoun ‘they’ (e.g., *John liked Mary because they had known each other for years*) or general continuations of the type ‘*because it was fun.*’ A small number of responses for which the raters disagreed (ca. 1%) were resolved by two of the authors. In the final classification, 94% of the continuations were either NP1 or NP2, 1.5% were blank, and 4.5% were “other” responses. For each verb, bias scores were defined as the difference between the number of NP1 and NP2 responses, weighted by the total number of valid responses [i.e., bias = $100 * (\text{noNP1} - \text{noNP2}) / (\text{noNP1} + \text{noNP2})$, with noNP1 being the number of NP1 continuations, and noNP2 being the number of NP2 continuations]. Using this definition, the bias score varies between 100 (all relevant continuations attributed the cause to NP1) and -100 (all relevant continuations were NP2 causes). A bias score of 0 reflects an equal number of NP1 and NP2 continuations. Null responses and non-classifiable responses were not taken into account for the bias score.

Results

Across subjects, 5.8% of the responses were blank or non-classifiable (SD = 5.0, range: .33–29%), 47.8% were NP1 continuations (SD = 6.5, range: 27–75) and 46.3% NP2 continuations (SD = 6.0, range: 24–66), indicating that all subjects used a variety of responses. NP1 and NP2 continuations were chosen equally often (see section “Gender” for the full statistical analysis by subjects).

Across verbs, the bias scores were widely and rather evenly distributed ($M = 2.2$, $SD = 56.2$, range: -98 to +93). The slight preference towards NP1 continuations was not significant in the analysis by items ($t(304) = 1.5$, n.s.).

Assuming a random binomial distribution of NP1 and NP2 continuations with 96 observations, more than 99% of the bias scores would fall between -26 and +26. According to this criterion, a large number of verbs in the corpus show

a significant bias towards either NP1 ($n = 127$) or NP2 ($n = 112$). Eighty-eight of the NP1 verbs and 73 of the NP2 verbs even met the very strict criterion of a bias score above 50 or below -50.

Reliability

To confirm that the continuations collected using a web-based questionnaire replicated previous results, the bias scores were compared to published verb classifications.

For all 77 verbs included in one of the three completion studies by Au (1986), we calculated a combined bias score, weighted by the number of participants who had provided a continuation to the particular verb (ranging between 20 and 160, depending on the experiments in which the verb was included). The correlation with the bias scores collected in the present study was $r = .94$, $n = 77$, $p < .0001$. Despite the high correlation, there were a number of qualitative differences. Some verbs were neutral in Au, but showed a rather strong NP1 bias in the present study (e.g., *cheat*, *betray*, *disobey*, *play with*, *surprise*). Conversely, some of Au’s biased verbs were neutral in the current study (e.g., NP1: *greet*; NP2: *reproach*).

Stewart et al. (1998, see Crinean & Garnham, 2006) conducted a sentence completion study using 48 verbs and 32 participants. For these 48 verbs the correlation between their data (using the same bias score as defined above) and the present web-based questionnaire was almost perfect ($r = .98$, $n = 49$, $p < .0001$). Note, however, that the verbs had been selected to be clearly biased toward NP1 or NP2 (Stewart et al., 2000). There were a few quantitative differences. For example, *dumbfound* had a weaker NP1 bias in the present study (bias scores of 42 vs. 89). None of the verbs showed a qualitative difference, however.

An inspection of the verbs used by Long and De Ley (2000) yields similar results. Of their 48 verbs, 46 are included in the present corpus. For all of their NP1 verbs and 21 of 23 NP2 verbs the very strong preferences were replicated (with bias scores larger than .50). NP1 verbs had a mean bias score of +74 (SD = 8.7), NP2 verbs of -78 (SD = 17.8; $t(44) = 36.8$, $p < .0001$). Two exceptions were the NP2 verbs *blame* and *deplore*, which in the current study had slightly weaker, but still highly significant, bias scores of -30 and -34, respectively.

A recent study by Goikoetxea et al. (2008) evaluated verb bias in Spanish. Based on the English translations given in their article, the correlation for the 42 verbs shared with the present corpus was 0.69. Despite slight variations in meaning between Spanish and English (e.g., Goikoetxea et al., 2008, translated both “admirar” and “estimar” as “admire;” “worry” in English is ambiguous, whereas its Spanish equivalent “preocupar” is not), this correlation shows good comparability of the two corpora. A few verbs

showed qualitatively different patterns. For instance, *bother* and *worry* were NP1 verbs in the present study, but NP2 in the Spanish version (*molestar* and *preocupar*). *Forget* and *move* were neutral in the present study, but NP1 biased in Spanish (*olvidar* and *conmover*).

Finally, we evaluated the verbs in the seminal study by Brown and Fish (1983). The correlation for 32 verbs was 0.91. Once more, most verbs showed the same bias in both studies. Exceptions were *disobey*, which had a stronger NP1 bias in the present study, and *protect*, which in the present study was biased towards NP2 rather than being neutral.

Length and frequency

The large number of items allows us to evaluate the influence of lexical features. The bias scores were correlated with the word frequency. Low frequency verbs elicited more NP1 continuations than verbs higher in frequency ($r = -.22$, $p < .0001$). There was a corresponding but weaker relationship between word length and bias ($r = .12$, $p < .05$).

Thematic roles and semantic class

Descriptive statistics of the bias scores for the four linguistic categories are presented in Table 1, and Fig. 1 shows the relationship between word frequency and bias scores for the four verb categories. As expected, the bias scores differed considerably for the categories. AgEvo and ExpStim verbs elicited more NP2 continuations, and StimExp verbs more NP1 continuations. Interestingly, for the AgPat verbs, no preference was found. An ANCOVA was conducted using the factors activity verb vs. psychological verb (i.e., AgPat/AgEvo vs. ExpStim/StimExp), and expected NP1 causality vs. expected NP2 causality (i.e.,

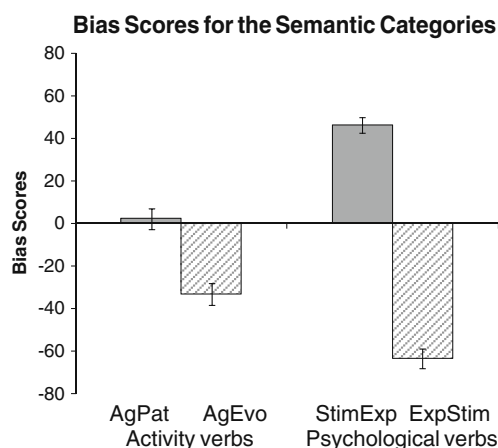


Fig. 1 Mean bias scores for the four linguistic categories (AgPat = Agent-Patient, AgEvo = Agent-Evocator, ExpStim = Experiencer-Stimulus, StimExp = Stimulus-Experiencer). The error bars show one standard error above and below the mean

AgPat/StimExp vs. AgEvo/ExpStim), controlling for length, frequency and valence. The results confirmed the pattern of correlations between the bias scores and the three features. The covariates frequency ($F(1, 298) = 5.5$, $p < .05$) and valence ($F(1, 298) = 15.8$, $p < .001$) were both significant, but not word length ($F(1, 298) = 1.9$, n.s.).

Controlling for these factors, there was no reliable difference between psychological and activity verbs ($F(1, 298) = 2.8$, $p = .10$). The expected causality influenced bias scores overall ($F(1, 298) = 222.0$, $p < .0001$), and this effect was larger for the psychological verbs than for the activity verbs ($F(1, 298) = 38.5$, $p < .0001$). For both verb types separately, the semantic categories yielded the predicted preference differences [StimExp vs. ExpStim verbs: $t(152) = 17.1$, $p < .0001$; AgEvo vs. AgPat verbs: $t(149) = 4.8$, $p < .0001$].

Gender

To evaluate the effects of gender, an analysis by participants was conducted. The mixed ANOVA included the factors Referent Position (NP1 vs. NP2), Referent Gender (M vs. F) and Participant Sex (women vs. men). There was an overall preference for continuations attributing the cause to the male character in the sentence fragment ($F(1, 94) = 25.4$, $p < .0001$). The participants produced a male continuation in 48.8% of the cases ($SD = 4.3$), compared to 45.4% female continuations ($SD = 4.2$). This preference was more pronounced for men than for women ($F(1, 94) = 4.7$, $p < .05$, for the interaction Gender \times Sex). Across subjects NP1 continuations were chosen as often as NP2 continuations ($F(1, 94) < 1.5$). However, women produced NP1 continuations more often than NP2 continuations, whereas men were more likely to continue with a continuation towards the male protagonist ($F(1, 94) = 5.6$, $p < .05$, for the interaction Position \times Sex). Neither the interaction between gender of the continuation and position in the sentence ($F(1, 94) < 1$) nor the three-way interaction with participant sex were significant ($F(1, 94) = 2.5$, n.s.).

For an item analysis of these gender effects, four bias scores were calculated, separately based on the women's and men's responses to sentence fragments with the male vs. female noun in sentence initial position. The means of these bias scores are shown in Fig. 2, and the scores for the individual verbs are included in the supplemental material (Psychonomic Society Archive). A 2×2 within-item ANCOVA, controlling for the factors valence, frequency and length, confirmed the analysis by subjects. First, the tendency for women to choose more NP1 continuations was confirmed ($F(1, 301) = 21.3$, $p < .0001$). Second, an NP1 continuation was more likely when the NP1 was male than when it was female ($F(1, 301) = 32.5$, $p < .0001$). The interaction between NP1 gender and participant sex

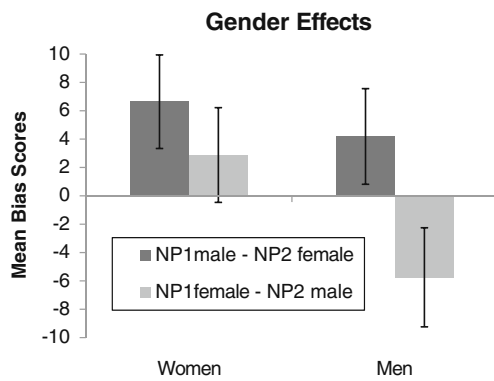


Fig. 2 Differential effects of the gender of the first noun phrase on the continuations chosen by men and women. The diagram shows the mean bias scores across all verbs as a function of the gender of the first noun phrase, separately for female and male participants. Women tend to follow an NP1 strategy. Men are more likely to choose a causal attribution towards the male protagonist. The error bars indicate the standard error of the mean

($F(1, 301) = 8.2, p < .01$) showed that women tended to follow an NP1 strategy independent of the order of the two NPs, whereas men were more likely to attribute the cause to the male protagonist, independent of whether he was mentioned in subject or object position.

Emotional valence

The ANCOVA also yielded significant effects of valence (see also the analysis of semantic categories). The main effect of valence indicated that negatively valenced verbs had more positive bias values, i.e., they were more likely to be attributed to NP1. There was also an interaction between the gender of the NP1 and valence ($F(1, 301) = 19.4, p < .0001$). Causes for negative events were more likely to be attributed to male protagonists, whereas positive events were more often attributed to the second NP, independent of gender. To illustrate this finding, the verbs were divided into three groups of approximately equal size according to their valence ratings. The means for the bias scores for the resulting positive, neutral and negative sets are shown in Fig. 3.

Table 2 displays a number of examples of individual verbs that were particularly sensitive to gender differences, i.e., those verbs for which the bias scores differed greatly, depending on whether NP1 was male or female. As can be seen, the verbs eliciting more male continuations tend to be negative in valence, whereas verbs that are more likely to elicit a female continuation have more positive valence ratings.

Discussion

The study provides normative data on implicit verb causality in English. To elicit causes we asked participants

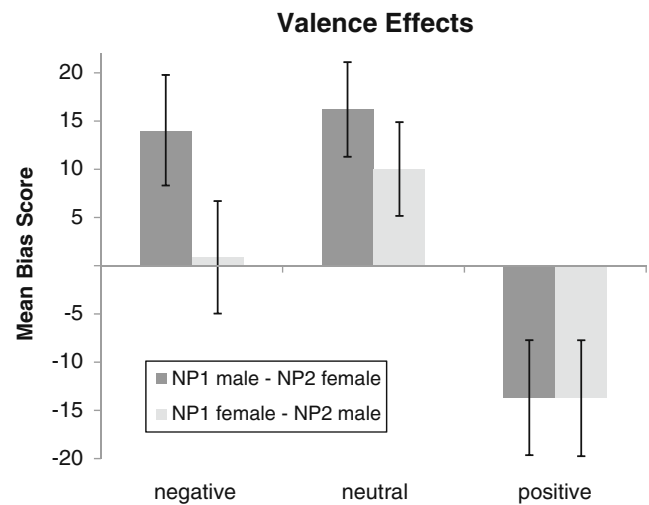


Fig. 3 The verb’s valence ratings influence the implicit causality bias. Verbs with more positive ratings are more likely to be biased towards an NP2 continuation than negative and neutral verbs. For the most negative verbs, the cause of the event is more likely to be attributed to the male protagonist. The categories shown are three categories of approximately equal size ($n = 104, 99, 102$). The error bars indicate the standard error of the mean

Table 2 Examples of verbs that showed exceptionally large gender effects. The table contains valence ratings, the bias scores (with negative values indicating NP2 bias, positive values NP1 bias; see text for formula). The gender effect is the difference in bias scores when NP1 was male (positive scores) and when NP1 was female (negative scores). Note that both subgroups of verbs contained NP1 bias, NP2 bias and neutral verbs

Verb	Valence rating	Bias score	Gender effect
Female bias:			
answered	0.55	-64	-41
rushed to	-0.55	-32	-44
welcomed	1.18	-17	-70
cuddled	2.54	-10	-56
relaxed	2.09	19	-41
met	0.82	53	-35
aggravated	-2.09	59	-41
enthralled	0.91	72	-31
Male bias:			
escorted	0	-36	54
hit	-1.82	-14	78
jolted	-0.91	-3	60
killed	-2.82	5	65
avoided	-1.09	14	53
defied	-1.27	27	60
tormented	-2.73	45	61
pained	-2.45	61	49

to complete sentence fragments ending with the connective *because*. The results replicate and extend previous findings. With over 300 verbs, we included a much larger number of interpersonal verbs than have previously been studied and showed that a majority of these verbs exhibit a clear bias in a standard sentence completion test to either NP1 or NP2 causality. These results are based on a larger group of respondents than most previous studies and should, therefore, provide more accurate estimates of the biases of individual verbs.

The collection of such a large data set was facilitated by the use of a web-based questionnaire and by the automation of some aspects of the scoring of the responses. The automated procedure was successful because a vast majority of the responses started with a pronoun. The proportion of over 90% is comparable to that reported by Fukumura and van Gompel (2010).

When the same verbs were used, our data largely replicate the results of previous normative studies (Au, 1986; Brown & Fish, 1983; Long & De Ley, 2000; Stewart et al., 1998). In addition, our results are broadly comparable to those of Goikoetxea et al. (2008) for Spanish verbs. The fact that this methodology produced similar results to previous “pencil and paper” studies is encouraging.

Nevertheless, care must be taken in using automated procedures. While the continuations beginning with a pronoun were easily identified, a number of cases with unambiguous NP1 or NP2 attribution were initially missed and had to be identified by inspection. In addition, we noted one minor complication in scoring the completions from the main study. As one of us (AG) has noted informally before, continuations of the following form occasionally occur.

- (11) Mary confided in John because *she* thought *he* might feature her in his new book.

Our automatic scoring procedure would classify such a continuation as NP1 (“she” = Mary). However, the *she thought* can be analyzed as merely a subjectification of, or as a hedge on, the underlying cause, which is a property of John (and hence, arguably, an NP2 attribution). This issue is not typically discussed in the literature, and we do not know how these cases were scored in previous studies. To estimate the number of such cases, we determined how many completions contained both a male and female pronoun (e.g., *he* and *she*). Only a very small proportion of continuations had this form (3.5%), and only a subset was of the described type. Thus, we can be confident that sentences of this type did not unduly influence outcome of the bias counts. Further research on this type of construction would be desirable.

We have followed much of the recent psycholinguistic literature in using the four-way classification of verbs into

the classes AgPat, AgEvo, ExpStim and StimExp. The findings largely replicate previous studies (Goikoetxea et al., 2008). For the psychological verbs, there was a clear dissociation between ExpStim and StimExp verbs. As inherent in the definition of these classes, the former verb type showed a clear NP2 bias, whereas the latter verbs were almost exclusively NP1. For the activity verbs, such a clear-cut dissociation was not found. Whereas the AgEvo verbs exhibited the predicted, but weaker, NP2 bias, there was no general pattern in the AgPat category.

The four-way semantic classification is not without its critics, however. In particular, as Crinean and Garnham (2006) argue, AgEvo verbs often have a psychological component to their meaning. Thus they effectively have an ExpStim component, though the *Experiencer* also performs an (evoked) action, and so has the properties of an agent. Furthermore, some StimExp verbs can have both a purely psychological (stative) interpretation and one in which an action also occurs (eventive). For example John can frighten Bill just by being a scary person, or by, for example, jumping out and shouting “boo.” In the eventive reading, these verbs have an AgPat component as well as a StimExp component to their meaning. In addition, some verbs carry ambiguities between a literal and figurative sense. For example, *move* is clearly an activity verb, but in the context of interpersonal relations it is more likely to be interpreted as referring to an emotional consequence. Thus, it was classified as an ExpStim verb. Despite the somewhat fuzzy boundaries between categories, we include the semantic classes in the present corpus.

The corpus also contains lexical features of the verbs that are likely to influence processing (e.g., lexical access or reading times). Even in the sentence completion study presented here, length and frequency (CELEX, Baayen et al., 1995) influenced the direction of implicit causality. Lower frequency words elicited more NP1 continuations. This result is not readily interpretable and might depend on the particular selection of verbs. However, lexical factors are undoubtedly important in on-line studies on verb causality. Shorter words and more frequent words are read faster, they are accessed more quickly, and they are subjectively more familiar. Thus, it is crucial to control for these factors. Given that the present corpus contains more than 70 verbs with very strong biases towards either NP1 or NP2, and more than 120 when applying a weaker criterion, it becomes possible to select subsets to match or manipulate these lexical features.

In addition to the lexical features, we also present ratings of the verbs’ emotional valence. This factor has been studied particularly in social psychological studies of causality (Corrigan, 2001; Semin & Marsman, 1994). Our own sentence completion data also show some interesting patterns of the effect of valence and its interaction with

gender of the protagonists in the event portrayed. Positive events were more likely to be attributed to the object noun, independent of its gender. In contrast, negative events were more likely to be attributed to the male protagonist. These effects deserve more systematic study, including investigations of how they are manifested in on-line comprehension. Our normative data will be helpful in selecting appropriate verbs for such studies.

The most interesting observations were the clear-cut gender effects. These effects occurred on two separate, but linked levels. First, there were effects caused by the order in which the mixed gender pair appeared in the sentence. For some verbs, it made a difference whether the male or the female protagonist was mentioned first, and this was independent of the specific direction of the bias. Thus, some interpersonal events are more likely to be attributed to a woman, and others to a man. Correlations with the valence ratings of the verbs suggested that these effects are modulated by the emotional connotations. Negative events (e.g., *hit*, *kill*) are more likely to be caused by men, whereas positive events (e.g., *cuddle*, *welcome*) are more likely to be caused by women. These findings are unsurprising from a social psychology point of view, in that they clearly reflect cultural stereotypes and knowledge about typical gender relations.

From a psycholinguistic point of view, on the other hand, these observations raise the question of whether implicit causality can be described as a lexical property of verbs (e.g., Crinean & Garnham, 2006). Lexical properties are seen as largely fixed. For example, the gender of a noun such as *aunt* is not affected by the context in which it occurs. Similarly, if implicit causality is a lexical property of the verb, it should not be greatly influenced by characteristics of the protagonists. However, even in the original psycholinguistic paper on implicit causality, Garvey and Caramazza (1974) noted that casual attribution is likely to vary between sentences such as “the son punished the father” and “the father punished the son.” In order to avoid such semantic asymmetries, we used very simple sentence fragments made up of arbitrary proper names providing only minimal context. Nevertheless, the gender of these names had a strong effect, at least for a number of verbs. Thus, the question arises whether verb biases assessed in sentence completion tasks are measures of a lexically represented bias. It is more likely that they are measures of a more complex phenomenon to which implicit causality is only one contributor.

The second set of gender effects was found when comparing the sentence completions of female and male participants. Although gender differences have been investigated in some studies, the results have been inconsistent (see Rudolph & Försterling, 1997, for review). LaFrance, Brownell, and Hahn (1997) showed that in mixed gender

pairs, the man is more often considered to be causing the event than the woman, and that these effects are modulated by verb valence, semantic class and participant gender. Goikoetxea et al. (2008), in contrast, could not replicate this result. One difference between these studies is that LaFrance et al. (1997) used explicit judgments of the respective roles, whereas Goikoetxea et al. (2008) based their analyses on a sentence completion task. Furthermore, stable gender differences can only be documented with a sufficiently large sample size.

In the present study, we tested almost 100 participants in roughly equally sized subgroups of women and men and obtained a set of rather complex results. Across all subjects, there was a main effect of gender, with more continuations attributing the cause to the male protagonist. However, an interaction with participant gender showed that this replication of LaFrance et al. (1997) held for male subjects only. As hypothesized, men exhibited an *own-gender* effect. They produced more male continuations, independent of the position of the male proper name in the main clause. In contrast, women chose more NP1 continuations, independent of gender, and did not show a preference for either the male or female as the event’s cause. At this point, an interpretation of these effects is not apparent. One explanation could be strategy differences during the completion of the questionnaires, or an asymmetry in the strength of gender stereotypes in the male and female groups. Further research is needed to investigate these findings, including on-line studies to document corresponding comprehension differences, and task manipulations to alter possible response strategies.

Our corpus of normative data should be useful in a range of studies in psycholinguistics and social psychology and, no doubt, other areas of psychology. It provides data on many more verbs than have previously been available and more reliable data for each verb, as the norm for that verb is based on almost 100 respondents. Studies that require a large number of different items, such as ERP and fMRI work, will benefit particularly, as will experiments requiring correlational analysis. Such analysis often requires relatively large numbers of items, and good estimates of individual verb biases will eliminate some noise from the data collected in such studies.

In addition, the corpus can be useful in a variety of applications beyond psycholinguistics. In particular, studies of pragmatic knowledge, social interactions, and interpersonal relations can benefit from a corpus that allows control of lexical properties of stimuli. Besides the intentional manipulation of implicit verb causality in such studies, the corpus can also help to avoid unwanted or confounding biases by selecting neutral verbs. For example, we recently conducted a study on the processing of gender stereotype information, as it is present in culturally defined nouns (e.g., *kindergarten teacher* is more likely to be interpreted

as a woman). The availability of a large number of neutral verbs facilitated this study.

Implicit causality remains an interesting research area with many open questions. The present corpus could facilitate studies of lexical and semantic representation in psycholinguistics, as well as studies of interpersonal relations and cultural norms in social psychology.

Author Notes This work was funded by the Economic and Social Research Council (ESRC, grant no. RES-000-22-2929 to Evelyn Ferstl and Alan Garnham). We thank Emily Scott for her help with the data collection. Christina Manouilidou is now at the University of Patras, Greece.

References

- Au, T. K. (1986). A verb is worth a thousand words: the causes and consequences of interpersonal events implicit in language. *Journal of Memory and Language*, *25*, 104–122.
- Baby names world. Retrieved October 26, 2008 from <http://babynamesworld.parentsconnect.com/category-british-names.html>
- Baayen, H. R., Piepenbrock, R., & Gulikers, L. (1995). *The CELEX lexical database. Release 2 (CD-ROM). Linguistic data consortium*. Philadelphia: University of Pennsylvania.
- Brown, R., & Fish, D. (1983). The psychological causality implicit in language. *Cognition*, *14*, 237–273.
- Bruner, J., & Feldman, C. (1993). Theories of mind and the problem of autism. In S. Baron-Cohen, H. Tager-Flusberg, & D. J. Cohen (Eds.), *Understanding other minds: perspectives from autism*. Oxford: Oxford University Press.
- Caramazza, A., Grober, E., Garvey, C., & Yates, J. (1977). Comprehension of anaphoric pronouns. *Journal of Verbal Learning & Verbal Behavior*, *16*, 601–609.
- Carreiras, M., Garnham, A., & Oakhill, J. (1996). Understanding anaphora: The role of superficial and conceptual information. In M. Carreiras, J. E. García-Albea, & N. Sebastián-Gallés (Eds.), *Language processing in Spanish* (pp. 241–274). Mahwah: Erlbaum.
- Corrigan, R. (2001). Implicit causality in language: Event participants and their interactions. *Journal of Language and Social Psychology*, *20*, 285–320.
- Crinean, M., & Garnham, A. (2006). Implicit causality, implicit consequentiality and semantic roles. *Language and Cognitive Processes*, *21*, 636–648. doi:10.1080/01690960500199763
- Ehrlich, K. (1980). Comprehension of pronouns. *The Quarterly Journal of Experimental Psychology*, *32*, 247–255.
- Franco, F., & Arcuri, L. (1990). Effect of semantic valence on implicit causality of verbs. *The British Journal of Social Psychology*, *29*, 161–170.
- Fukumura, K., & Van Gompel, R. P. G. (2010). Choosing anaphoric expressions: do people take into account likelihood of reference? *Journal of Memory and Language*, *62*, 52–66. doi:10.1016/j.jml.2009.09.001
- Garnham, A. (2001). *Mental models and the interpretation of anaphora*. New York: Psychology Press.
- Garnham, A., & Oakhill, J. (1985). On-line resolution of anaphoric pronouns: effects of inference making and verb semantics. *British Journal of Psychology*, *76*, 385–393.
- Garnham, A., Oakhill, J., & Cruttenden, H. (1992). The role of implicit causality and gender cue in the interpretation of pronouns. *Language & Cognitive Processes*, *7*, 231–255.
- Garnham, A., Traxler, M., Oakhill, J. V., & Gernsbacher, M. A. (1996). The locus of implicit causality effects in comprehension. *Journal of Memory and Language*, *35*, 517–543.
- Garvey, C., & Caramazza, A. (1974). Implicit causality in verbs. *Linguistic inquiry*, *5*, 459–464.
- Garvey, C., Caramazza, A., & Yates, J. (1974-1975). Factors influencing assignment of pronoun antecedents. *Cognition*, *3*, 227–243.
- Goikoetxea, E., Pascual, G., & Acha, J. (2008). Normative study of the implicit causality of 100 interpersonal verbs in Spanish. *Behavior Research Methods*, *40*, 760–772. doi:10.3758/BRM.40.3.760
- Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review*, *101*, 371–395.
- Holtgraves, T. M., & Kashima, Y. (2008). Language, meaning, and social cognition. *Personality and Social Psychology Review*, *12*, 73–94. doi:10.1177/1088868307309605
- Koornneef, A. W., & van Berkum, J. J. A. (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*, 445–465. doi:10.1016/j.jml.2005.12.003
- LaFrance, M., Brownell, H., & Hahn, E. (1997). Interpersonal verbs, gender, and implicit causality. *Social Psychology Quarterly*, *60*, 138–152.
- Levin, B. (1993). *English verb classes and alternations*. Chicago, IL: The University of Chicago Press.
- 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*, 545–570. doi:10.1006/jmla.1999.2695
- Mannetti, L., & De Grada, E. (1991). Interpersonal verbs: implicit causality of action verbs and contextual factors. *European Journal of Social Psychology*, *21*, 429–443.
- 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*, 543–566. doi:10.1006/jmla.1995.1025
- McKoon, G., Greene, S. B., & Ratcliff, R. (1993). Discourse models, pronoun resolution, and the implicit causality of verbs. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *19*, 1040–1052. doi:10.1037//0278-7393.19.5.1040
- Pickering, M. J., & Majid, A. (2007). What are implicit causality and consequentiality? *Language and Cognitive Processes*, *22*, 780–788. doi:10.1080/01690960601119876
- Pyykkönen, P., & Järvikivi, J. (2010). Activation and persistence of implicit causality information in spoken language comprehension. *Experimental Psychology*, *57*, 5–16. doi:10.1027/1618-3169/a000002
- Rudolph, U. (1997). Implicit verb causality: verbal schemas and covariation information. *Journal of Language & Social Psychology*, *16*, 132–158.
- Rudolph, U. (2008). Covariation, causality, and language: developing a causal structure of the social world. *Social Psychology*, *39*, 174–181. doi:10.1027/1864-9335.39.3.174
- Rudolph, U., & Försterling, F. (1997). The psychological causality implicit in verbs: a review. *Psychological Bulletin*, *121*, 192–218.
- Semin, G. R., & Marsman, J. G. (1994). “Multiple inference-inviting properties” of interpersonal verbs: event instigation, dispositional inference and implicit causality. *Journal of Personality and Social Psychology*, *67*, 836–849.
- Stevenson, R. J., Crawley, R. A., & Kleinman, D. (1994). Thematic roles, focus and the representation of events. *Language and Cognitive Processes*, *9*, 519–548.
- Stevenson, R. J., Knott, A., Oberlander, J., & McDonald, S. (2000). Interpreting pronouns and connectives: interactions among focusing, thematic roles and coherence relations. *Language and Cognitive Processes*, *15*, 225–262.

- Stewart, A. J., Pickering, M. J., & Sanford, A. J. (1998). Implicit consequentiality. In *Proceedings of the Twentieth Annual Conference of the Cognitive Science Society* (pp. 1031–1036). Mahwah: Lawrence Erlbaum.
- 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, 423–443. doi:10.1006/jmla.1999.2691
- Van Berkum, J. J. A., Koornneef, A. W., Otten, M., & Nieuwland, M. S. (2007). Establishing reference in language comprehension: an electrophysiological perspective. *Brain Research*, 1146, 158–171. doi:10.1016/j.brainres.2006.06.091