Language, Cognition and Neuroscience, 2013
http://dx.doi.org/10.1080/01690965.2013.796396

Routledge
Taylor & Francis Group

What is implicit causality?

Joshua K. Hartshorne*

Massachusetts Institute of Technology, Harvard University, 77 Massachusetts Avenue, 46-4053H, Cambridge, MA, USA

(*Received 4 May 2012; final version received 24 March 2013*)

In causal dependent clauses, the preferred referent of a pronoun varies systematically with the verb in the main clause (contrast *Sally frightens Mary because she*... with *Sally loves Mary because she*...). This "implicit causality" phenomenon is understood to reflect intuitions about who caused the event. Researchers have debated whether these intuitions are based on linguistic structure or instead a function of high-level, non-linguistic cognition. Two lines of evidence support the latter conclusion: implicit causality is related to a broad social judgement task, and it is affected by general knowledge about the participants in the event. On closer inspection, neither of these claims have been established. Eight new experiments find that (a) the relationship between implicit causality and the social judgement task is tenuous, and (b) previously employed event-participant manipulations have minimal to no effect on implicit causality. These findings support an account on which implicit causality is driven primarily by linguistic structure and only minimally by general knowledge and non-linguistic cognition.

Keywords: pronoun resolution; implicit causality; thematic roles; inference; pragmatics.

In 1974, Catherine Garvey and Alfonso Caramazza introduced a puzzling phenomenon, illustrated below:

(1) a. Sally frightened Mary because she...
b. Sally loved Mary because she...

In principle, the third-person pronoun *she* is ambiguous and could refer to anyone. However, most English speakers resolve the pronoun to Sally in (1a) but to Mary in (1b). This bias appears in production as well, with people more likely to continue (2a) with a reference to Sally and (2b) with a reference to Mary, whether they do so with a pronoun or not (see esp. Kehler, Kertz, Rohde, & Elman, 2008):

(2) a. Sally frightened Mary because...
b. Sally loved Mary because...

In this paper, I will use the term "re-mention bias" to refer collectively to both phenomena (1–2). The facts that (a) the connective *because* introduces an explanation, and (b) explanations usually refer to causes, suggested to Garvey and Caramazza that somehow *frighten* implicitly marks its subject (e.g., Sally) as the cause of the fright, whereas *love* marks its object (e.g., Mary) as the cause of the love. This, they suggested, is the function of an "implicit cause" feature carried by verbs. Many verbs like *frighten* are "subject-biased" in that they lead speakers to explain the event in terms of the verb's subject (*cf* 2a) and comprehenders to interpret explanations as referring to the verb's subject (*cf* 1a). Many others, such as *love*, are "object-biased" (*cf* 1b, 2b).

Despite the somewhat grammatical flavour of this explanation, it was clear from the beginning that this re-mention phenomenon[1] is not due to a rule of syntax, as the preferred pronoun interpretation can be overturned by later content (though not without a processing cost: Caramazza, Grober, Garvey, & Yates, 1977; Koornneef & van Berkum, 2006; Stewart, Pickering, & Sanford, 2000):

(3) a. Sally frightened Mary because she is very timid.
b. Sally loved Mary because she loves everyone.

As such, re-mention biases appear to be pragmatic inferences. Many researchers have argued that many pragmatic inferences are more properly part of higher-level, domain-general cognition (Clark, 1996; Goodman & Stuhlmuller, 2013; Grice, 1989; Noveck & Reboul, 2008; Sperber & Wilson, 1986; see also Lee & Pinker, 2010). Theories of pragmatics thus require operations and representations beyond those employed in more squarely linguistic phenomena (i.e., syntax, semantics and phonology).[2]

Much work on implicit causality subsequent to Garvey and Caramazza (1974) fits squarely in this framework, motivated in part by reports that implicit

*E-mail: jkhartshorne@gmail.com

causality could be identified in less obviously linguistic tasks (Au, 1986; Blankenship & Craig, 2012; Brown & Fish, 1983) and can be affected by essentially general knowledge (Corrigan, 1988, 1992, 2001, 2002, 2003; Ferstl, Garnham, & Manouilidou, 2011; Garvey, Caramazza, & Yates, 1974; LaFrance, Brownell, & Hahn, 1997; van Kleeck, Hillger, & Brown, 1988). These will be discussed in detail below ("Implicit Causality as a Non-Linguistic Inference: Re-Evaluating the Evidence").

The most succinct statement of this developing consensus comes from Pickering and Majid (2007):

> How then do people compute the inference of implicit causality? Various components of the verb's meaning are of course important [including] how enduring the event is, how concrete it is, whether it is telic or not (Semin & Fiedler, 1988; Rudolph & Forsterling, 1997), and how negative its connotative meaning is (Semin & Marsman, 1994). In addition, properties of the participants affect implicit causality. Changing the gender (Lafrance, Brownell, & Hahn, 1997), animacy (Corrigan, 1988, 1992), or typicality (Corrigan, 1992; Garvey et al., 1976 [sic]) of the participants changes the implicit-causality bias, as do contextual factors that affect focus (Majid, Sanford, & Pickering, 2006) . . . All of these factors affect the construction of the event representation, and *it is this event representation that is used to infer the cause.* (pp. 785–786, emphasis added).

Implicit causality and linguistic structure

As described above, while the actual *tasks* in (1–3) – interpreting pronouns and completing sentences – are linguistic, many researchers have asserted that the relevant computations and representations are not. As such, implicit causality has been used to probe the development of causal schemas in children (Au, 1986; Corrigan & Stevenson, 1994), the stability of these schemas across cultures (Brown & Fish, 1983) and the conceptualisation of social relationships and dominance hierarchies (Corrigan, 2001; LaFrance et al., 1997).

Implicit causality and argument structure

Nonetheless, recent findings have suggested a tighter relationship between re-mention biases and core linguistic phenomena. The first line of work implicates verb argument structure and the syntax–semantics interface. Linguists have long noted that when verbs are categorised according to the syntactic frames in which they can appear, the verbs in each class exhibit systematic correspondences in meaning (Levin & Rappaport Hovav, 2005). Competent speakers are sensitive to these correspondences and children make use of them during acquisition (Ambridge, Pine, Rowland, Chang, & Bidgood, 2013; Gleitman, 1990; Kako, 2006; Pinker, 1989). While theories of argument structure vary, nearly all assert that how a verb encodes causation is a core feature of verb meaning that drives verb argument structure and its syntactic realisation (Levin & Rappaport Hovav, 2005).

The suggestion that argument structure is implicated implicit causality goes back to at least Brown and Fish (1983) and is described in more detail in Appendix 1. Recently, Hartshorne and Snedeker (in press) showed for a large corpus of verbs that re-mention bias varies systematically with the above-mentioned syntactic verb classes. Moreover, the direction of bias tracked how the verbs encode causality according to one well-known class of verb argument structure theories. Hartshorne (in press) replicated these findings for a smaller set of verbs across seven additional languages.

Thus, on this account, no inference is necessary to determine who caused the event in *Sally frightened Mary*, since to comprehend the sentence is to know that Sally caused the frightening (Pesetsky, 1995). Schemes for inferring causality, such as overarching cognitive schema for conceptualising of interpersonal interactions (Brown & Fish, 1983; Semin & Fiedler, 1991) and non-linguistic event representations (Pickering & Majid, 2007), become superfluous.

Implicit causality and discourse structure

The second line of work invokes discourse structure. Discourse structure theory attempts to explain how information from different sentences in a text, dialog or other discourse relates to one another. Researchers have identified a short set of relations which govern discourse, such as EXPLANATION, RESULT and PARALLELISM, though the exact list and their definitions remain an area of active research (Kehler, 2002; Wolf & Gibson, 2006):

(4) a. Sally frightened Mary because Sally is scary. (EXPLANATION)
b. Sally frightened Mary so Mary ran away. (RESULT)
c. Sally frightened Mary, and John terrified Bill. (PARALLELISM)

These relations have been implicated in a wide range of syntactic and semantic phenomena including VP ellipsis, subjacency violations, gapping and – importantly for the present discussion – reference and anaphor. Kehler and colleagues (Kehler, 2002; Kehler et al., 2008) have shown that re-mention biases reliably follow the discourse structure (see also Crinean & Garnham, 2006; Stewart & Pickering, 1998; Stevenson, Crawley, & Kleinman, 1994). When a pronoun appears in an explanatory clause, it preferentially refers to the previous clause's cause (5a); when a pronoun appears in a result clause, it preferentially refers to the previous clause's affected entity (5b); when a pronoun refers in a parallelism clause, it preferentially refers to the entity playing the parallel role in the previous clause (5c–d):

(5) a. Sally frightened Mary because she [= Sally] . . .
b. Sally frightened Mary so she [= Mary] . . .
c. Sally frightened Mary, and she [= Sally] terrified Bill.
d. Sally frightened Mary, and Bill terrified her [= Mary].

Analogous findings hold for production and are independent of whether or not the referring phrase is a pronoun (see especially Kehler et al., 2008). As such, implicit causality becomes one sub-case of re-mention biases, which are themselves a by-product of discourse structure.

Implicit causality as a non-linguistic inference: re-evaluating the evidence

The close relationship between implicit causality and linguistic structures responsible for core linguistic phenomena motivates re-evaluating the proposition that implicit causality is an inference based on high-level, non-linguistic representations of events (Brown & Fish, 1983; Corrigan, 2001, 2002, 2003; Pickering & Majid, 2007; Semin & Fiedler, 1991). What facts about implicit causality, if any, cannot be explained by argument structure and discourse structure? The literature provides two such phenomena, which I review below: (a) implicit causality appears in non-linguistic tasks, and (b) implicit causality is modulated by non-linguistic, general knowledge.

Causal attribution

Garvey and Caramazza (1974) coined the term "implicit causality" to refer to the effect of verbs on pronoun interpretation in comprehension (1) and also on the likelihood of re-mention (2). The equation of these two phenomena is both theoretically justified (see previous section) and experimentally confirmed, with tight correspondences between the results of both types of tasks (see especially Hartshorne, in press; Hartshorne & Snedeker, in press; and also below). Subsequent to Garvey and Caramazza (1974), researchers have invoked implicit causality to explain a variety of other phenomena. The most prominent of these is the Brown and Fish (1983) causal attribution task:

(6) Sally frightened Mary. How likely is it that this was because:
a) Sally is the kind of person who frightens people.
Not likely 1 2 3 4 5 6 7 8 9 Definitely likely
b) Mary is the kind of person people frighten.
Not likely 1 2 3 4 5 6 7 8 9 Definitely likely
c) Some other reason.
Not likely 1 2 3 4 5 6 7 8 9 Definitely likely

Results are typically analysed by subtracting the answer for (b) from the answer for (a), so that positive numbers reflect greater causal attribution to the subject (Sally) whereas negative numbers reflect greater causal attribution to the object (Mary). Although the stimuli are linguistic, Brown and Fish argued that its results reflect high-level social cognition (their work grew out of Attribution Theory: Kelley, 1967; Kelley & Michela, 1979; McArthur, 1972). In the absence of additional information, we have biases about which participant in an interpersonal interaction is likely to have caused the interaction, biases which are tapped explicitly in (6). These same biases can also provide top-down influences on language production and interpretation.[3]

In fact, whether the appearance of implicit causality in the Brown and Fish causal attribution task provides evidence that high-level cognition governs the re-mention effect as opposed to evidence that linguistic processes govern the causal attribution task is unclear. In any case, while a close relationship between implicit causality re-mention biases and causal attribution is widely assumed in the literature – researchers frequently justify claims about one by citing results of the other (Au, 1986; Blankenship & Craig, 2012; Corrigan, 1988, 2001, 2002; Ferstl, Garnham, & Manouilidou, 2011; Goikoetxea, Pascual, & Acha, 2008; Greene & McKoon, 1995; Holtgraves & Raymond, 1995; Kasof & Lee, 1993; McKoon, Greene, & Ratcliff, 1993; Pickering & Majid, 2007; Pynte, Kennedy, Murray, & Courrieu, 1988; Rudolph, 2008; Rudolph & Forsterling, 1997; Semin & Fiedler, 1991; Van Kleeck et al., 1988; Vorster, 1985) – there is little evidence that such a relationship exists.

For instance, though verbs which are subject- or object-biased in re-mention tasks are widely assumed to be similarly subject- and object-biased, respectively, in causal attribution tasks (e.g., Hartshorne & Snedeker, in press; Rudolph & Forsterling, 1997), empirical tests are limited and the evidence is mixed. Brown and Fish (1983) compared 36 verbs – mostly mental state verbs – in causal attribution and sentence completion and found an $r = 0.88$ correlation in verb bias, a finding replicated by Ferstl et al. (2011) for 32 of these verbs ($r = 0.80$). Brown and van Kleeck (1989) tested 24 mental state verbs – 7 of which were tested in Brown and Fish – and likewise found a strong correlation ($r = 0.86$). Vorster (1985), in an Afrikaans study based on Brown and Fish's (1983) verbs, reported a somewhat weaker relationship. Corrigan (1988) investigated 32 verbs – none of which were mental state verbs – and found no correlation ($r = 0.00$).[4] Rudolph and Forsterling (1997) compared previously reported studies of implicit causality, with conflicting findings: while there was a significant correlation in biases, an ANOVA also revealed a main effect of methodology (they distinguished two method types, corresponding roughly to causal attribution tasks and sentence continuation tasks; pronoun comprehension tasks were excluded). Thus, while re-mention and causal attribution verb biases correlate for at least some mental state verbs, there is little evidence that this correlation generalises.

Additional indirect evidence comes from studies of verb taxonomies. Starting with Brown and Fish (1983), researchers have attempted to create taxonomies of verbs that predict whether a verb will be subject- or object-biased in causal attribution tasks. If such a taxonomy also predicted re-mention biases, that would be evidence for the two tasks' equivalence. Some early studies with small numbers of verbs were promising (e.g., Van Kleeck et al., 1988). However, Au (1986) and later researchers identified many exceptions to the taxonomy for both causal attribution and re-mention biases. While other taxonomies were proposed in the implicit causality literature, Hartshorne and Snedeker (in press) recently demonstrated that none of these perform appreciably better than chance at predicting pronoun processing on a large, representative set of verbs. Instead, Hartshorne and Snedeker showed that syntactic verb classes proposed in the linguistic literature predict re-mention biases extremely well. However, these verb classes have never been applied to causal attribution biases, so whether they predict causal attribution biases is unknown. Thus, the evidence from verb taxonomies is inconclusive at best.

In conclusion, there is little evidence that re-mention and causal attribution involve the same verb biases, though this is as much absence of evidence as evidence of absence.[5] In the experiments below, I investigate this question systematically.

Event-participant manipulations

A second line of argumentation placing re-mention biases in high-level cognition rather than linguistic processing is that they are affected by "world knowledge" (see Pickering & Majid quote above). The most widely studied evidence is that the relative social status of the event participants interacts with verb bias such that the high-status individuals are viewed as more likely to cause interpersonal events, mediating the effect of the verb. By hypothesis, the pronoun is more likely to refer to the manager in (7a) than in (7b):

(7) a. The manager frightened the employee because he...
b. The manager frightened the CEO because he...

Following the same reasoning, some researchers have suggested that, to the extent men are viewed as higher-status and/or more likely to cause interpersonal events than women, (8a) is more likely than (8b) (Ferstl et al., 2011; Goikoetxea et al., 2008; Mannetti & de Grada, 1991; and esp. LaFrance et al., 1997):

(8) a. John frightened Sally because he...
b. Sally frightened John because she...

Since any expectation that managers are more likely to cause employees to do things than cause CEOs to do things and that men are more likely to causally affect women than *vice versa* likely resides in general cognition, not the linguistic system, such effects would suggest that implicit causality itself resides in general cognition.

However, the evidence supporting these claims is limited and contradictory, and much of the evidence actually comes from causal attribution studies that may not generalise to re-mention biases. The effect of social hierarchy on re-mention biases has been addressed in a pilot study by Garvey and colleagues (1974), who manipulated the "typicality" of the event for five verbs, finding significant effects on three. The one example given involves a social hierarchy (*The prisoner confessed to the guard because he*... vs. *The guard confessed to the prisoner because he*...), though it was not one of the ones to show a significant effect of the manipulation. The text gives another example (*The son praised the father because he*... vs. *The father praised the son because he*...), but it is not clear whether this was used as a stimulus in the experiment.

Several studies have considered the effect of social hierarchy on causal attribution. The most extensive come from Corrigan (2001, 2002, 2003). While many experiments revealed significant effects and/or interactions of potency, the designs involved are complex (the 2003 study employs a $2 \times 2 \times 2 \times 2 \times 2 \times 3$ design), and results are not necessarily in the same direction from experiment to experiment. For instance, while the 2001 study found that event participants who are independently rated to be more potent are judged in causal attribution tasks to be more causal, the 2002 study found that sentential subjects were judged to be more causal when *either* the subject or object was potent, at least in Exp. 1. In Exp. 2, sentential subjects were rated more causal if the object was potent, but there was no effect of subject potency. The 2003 study failed to find any effect of potency.[6] It is conceivable that differences in the overall composition of the stimuli combined with higher-order interactions mask the actual consistency in the results. For instance, Corrigan argues that more potent individuals are only judged to be more causal when the role they play in the event is typically played by a potent individual (e.g., *The king condemned the butler*), whereas if there is a mismatch between the verb and the event participant, the opposite pattern should be found (that is, participants should attribute more causality to the situation that the event participants find themselves in). A similar argument is made for the valence (positive/negative) of the event and the event participants. While these predictions are generally born out in the 2002 study, the 2003 study shows no evidence of such interactions, though again this is perhaps due to higher-order interactions.[7]

LaFrance and colleagues (1997; Exp. 3) report that when the event participants are in a clear social hierarchy (e.g., employer/employee), the sentential subject is rated more causal when it is higher in the hierarchy. This result is difficult to interpret because attributions to the sentential object were not analysed; as such it is not clear whether this represents the sentences becoming more subject-biased or whether it represents stronger causal attributions to *both* the (high-status) subject and (low-status) object.

In terms of gender, the most comprehensive re-mention study comes from Ferstl and colleagues (2011), who found that continuations were more likely to mention the male character in a sentence-continuation study of mixed-gender implicit causality sentences with over 300 English verbs, a finding particularly true of negatively valenced verbs. However, in a very similar study of 100 verbs in Spanish, Goikoetxea and colleagues (2008) failed to find any effect of gender, though they did not consider negatively valenced verbs separately. Mannetti and de Grada (1991), who did, also failed to find an effect of gender.[8] One study of causal attribution (LaFrance et al., 1997) is often cited as showing that men are judged to be more causal than women (Corrigan, 2001; Ferstl et al., 2011; Goikoetxea et al., 2008; Pickering & Majid, 2007; Rudolph, 2008), though in fact the results are complex: Although in Exp. 2, male sentential subjects are assigned more causality than female sentential subjects – 6.2 vs. 6.0 on a 9-point Likert scale – Exp. 1 shows an effect in the opposite direction and over twice as large – 4.4 vs. 4.9.[9] Strangely, when verbal bias is calculated by subtracting the object rating from the subject rating (the analysis used below and in most other studies of causal attribution), the results are negatively correlated between the two experiments ($r = –0.8$), an unexplained result motivating caution in interpreting the results of this study.[10]

In conclusion, while the gender and social status of the event participants may play a causal role in implicit causality, that they do so has not been definitively established. Because these are the most widely studied and frequently cited lines of evidence supporting the claim that places re-mention biases in high-level cognition, I focus on them in the studies below. I return to other related findings in the General Discussion.

Overview of experiments

Exps. 1–4 investigate (a) whether Brown and Fish causal attribution biases reliably predict implicit causality re-mention biases, and (b) whether either phenomenon is affected by social hierarchy and gender manipulations. I find little relationship between causal attribution and implicit causality re-mention biases; Thus, in Exps. 5–6, I test whether this lack of relationship could be due to superficial aspects of the task designs. It is not. Thus, Exps. 7–8 attempt to elucidate what the Brown and Fish causal attribution task measures.

Experiments 1–2

In Exps. 1–2, event-participant manipulations are explored in paired re-mention and causal attribution tasks. Exp. 1 employs a social hierarchy manipulation, and Exp. 2, a gender manipulation. The methods and results for the two experiments are presented and discussed jointly.

Method

Experiment 1

Participants. Participants in Exps. 1–7 were tested via Amazon Mechanical Turk and included only if they responded to every trial and had not participated in any condition of any other of the experiments reported in this paper. Exps. 1–6 were restricted to native English speakers. Forty-eight individuals participated in the causal attribution task (32 female; 18–67 years old with 1 no-report, M = 37, SD = 13), and 48 in the re-mention bias task (32 female; 19–75 years old, M = 36, SD = 14). An additional five participants were excluded for giving the same response to every question ($N = 4$) or for experimenter error ($N = 1$).

Materials. Stimuli are shown in Appendix 2. Twenty-four verbs were chosen so as to provide a broad range of verbs with implicit causality biases. These verbs were chosen from among those tested by Hartshorne and Snedeker (in press), who investigated implicit causality pronoun processing for a large number of verbs classified according to the syntactic frames in which they can appear (see Levin, 1993, for a review), showing that verbs in these classes tend to have the same re-mention bias. In order to have broad coverage, six verbs were chosen from each of two subject-biased classes (causal verbs and experiencer–object emotion verbs) and six from each of two object-biased classes (judgement verbs and experiencer–subject emotion verbs). These four classes represented the largest classes with consistent re-mention biases identified by Hartshorne and Snedeker.[11] Six pairs of characters with defined social hierarchies (e.g., *duke* and *butler*, *king* and *knight*) were chosen. Each was assigned randomly to one verb from each of the four verb classes such that across each class the same role pairs were used. Four lists were created, counterbalancing which character was the sentential subject (within and between lists) as well as trial order (between lists).

Procedure. The procedure for the re-mention task was:

(9) The butler blamed the duke because he is a froom.
Who is a froom? the butler/the duke

Each sentence ended with a unique novel word such as *froom*. This procedure, introduced by Hartshorne and Snedeker (in press), mitigates the fact that the material following the pronoun can override the pronoun bias and force particular interpretations (e.g., *Sally frightened Mary because she was easily scared*). Results of this task correlate strongly with production tasks (Hartshorne, in press; Hartshorne, in press; see also below).

The causal attribution task was adapted from Exp. 1 of Brown and Fish (1983) (see 6). As responses to the third part of the question (*c some other reason*) are not typically analysed, it was dropped in order to shorten the experiment. Whether dropping (c) affected the results is investigated in Exp. 6.

Experiment 2

Participants. Ninety-six completed the re-mention task (65 female, 1 no-report; 18–67 years old, M = 35, SD = 12) and 96 completed the causal attribution task (65 female, 18–81 years old, M = 36, SD = 15). Five additional participants were excluded for giving the same response to every question.

Materials. The same verbs and lists from Exp. 1 were used. Subjects and objects of the verbs were chosen from common male and female names. Four stimulus lists were constructed as in Exp. 1.

Procedure. The re-mention task was a forced-choice variant of the sentence continuation paradigm frequently used to assess re-mention biases (e.g., Arnold, 2001; Ferstl et al., 2011; Goikoetxea et al., 2008; Kehler et al., 2008; Stevenson et al., 1994):

(10) Which word is the most likely continuation for the following sentence?
Christopher affected Ashley because
a. he b. she

The causal attribution task was identical to that in Exp. 1.

Results

Event-participant manipulations

At issue was whether the event-participant manipulations affected the re-mention and/or causal attribution biases – e.g., whether there was an overall high-status bias in Exp. 1 (greater subject bias when the sentential subject was high status than when the sentential subject was low status) or male bias in Exp. 2 (greater subject bias when the sentential subject was male than when the sentential subject was female). For purposes of discussion and presentation in the figures, high-status and male biases were calculated by item for both re-mention and causal attribution, and then converted to a (−100, +100) scale to facilitate comparison across the tasks.[12]

Results are shown in Figures 1 and 2. Differences between the verb classes are not the key focus here (representatives of four classes were used in order to ensure that a variety of verbs were represented), and there was no evidence that the manipulations of interest affected the classes differently in any of the experiments. However, the results are graphed by class for the interested reader throughout. While there was an overall high-status bias for causal attribution (high-status bias = +9, SE = 3; $t = 3.2$, $p = 0.001$), there was no effect on re-mention biases (high-status bias = +1,

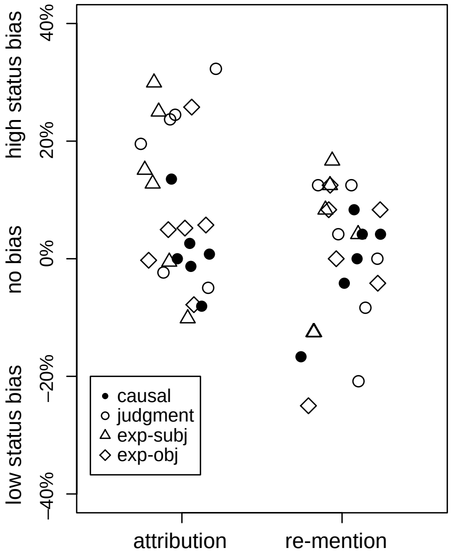

Figure 1. Effects of the social hierarchy manipulation in Exp. 1, displayed in terms of high-status bias (possible range: [−100, +100]) and plotted by verb. The four verb classes are indicated.

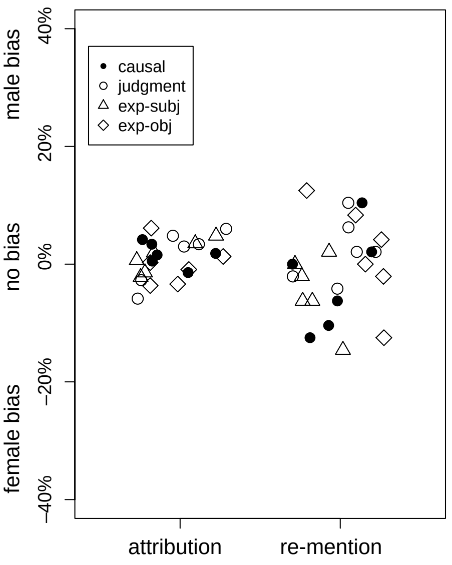

Figure 2. Effects of the gender manipulation in Exp. 2, displayed in terms of male bias (possible range: [−100, +100]) and plotted by verb. The four verb classes are indicated.

SE = 2; *Wald's z* = 0.5, *p* = 0.6). This difference in high-status biases elicited by the tasks was significant ($t1(94) = 2.97$, $p = 0.004$; $t2(46) = 2.28$, $p = 0.03$) (analysed in by-subjects and by-items *t*-tests; the differences in outcome measures prevented analysis by mixed effects models). The gender manipulation did not affect either causal attribution (male bias = 1, SE = 1; $t = 1.6$, $p = 0.12$) or re-mention (male bias = –1, SE = 2; *Wald's z* = 1.1, *p* = 0.29).

Comparison of re-mention and causal attribution biases

Were the results for individual items consistent across tasks (e.g., were items that were subject-biased on re-mention subject-biased on causal attribution)? Collapsing across experiment and event-participant, re-mention and causal attribution verb biases correlated significantly ($r = 0.54$, $p = 0.007$; Figure 3A). However, if one made a binary distinction between subject- and object-biased items – as is typical in the literature – and attempted to generalise the causal attribution results to the re-mention results, one would be correct only 58% of the time, a rate not significantly greater than chance in a binomial test ($p = 0.3$, one-tailed). By comparison, the by-verb correlation between the two re-mention experiments (Exp. 1 and 2) was near ceiling ($r = 0.96$, $p < 0.001$; Figure 3C), with 96% of verbs showing the same bias in both tasks ($p < 0.001$). The correlation

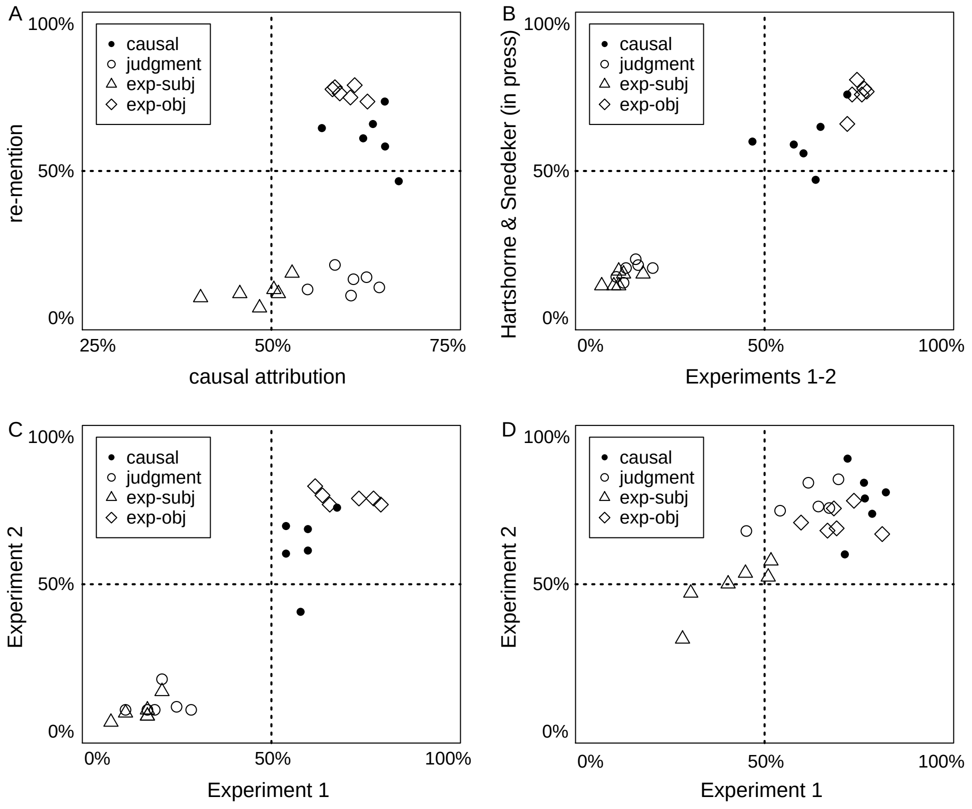

Figure 3. Note: Axes in all panels are percentage of maximum possible subject bias (0% = maximum possible object bias). Panel A: Correlation between re-mention and causal attribution biases, collapsing across event-participant and across Exps. 1–2. Panel B: Correlation between re-mention biases in Hartshorne & Snedeker (in press) and the re-mention biases in Exps. 1–2, collapsing across experiment and event-participant. Panel C: Correlation between re-mention biases in Exps. 1 and 2. Panel D: Correlation between causal attribution biases in Exps. 1 and 2.

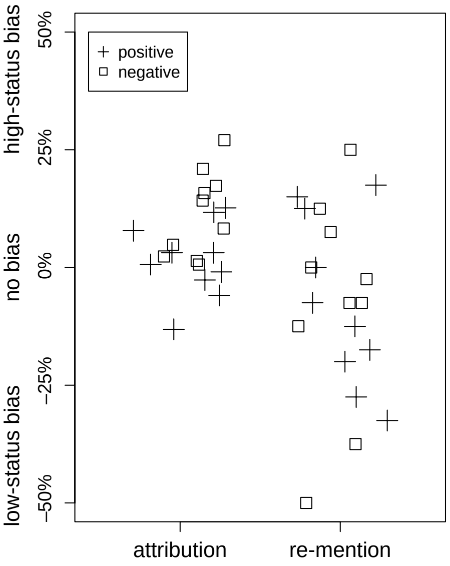

Figure 4. Effects of the social hierarchy manipulation in Exp. 3, displayed in terms of high-status bias (possible range: [−100, +100]) and plotted by verb. Positively valenced and negatively valenced items indicated.

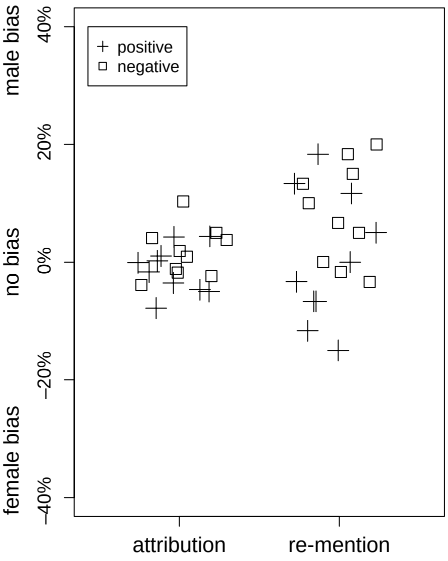

Figure 5. Effects of the gender manipulation in Exp. 4, displayed in terms of male bias (possible range: [−100, +100]) and plotted by verb. Positively valenced and negatively valenced items indicated.

between the two causal attribution experiments was somewhat lower ($r = 0.78$, $p < 0.001$; Figure 3D), perhaps because of the significant event-participant manipulation in Exp. 1. Nonetheless, this correlation was much higher than the correlation between causal attribution and re-mention biases, with 88% of verbs revealing same bias on both tasks, a rate significantly greater than chance ($p < 0.001$).

Discussion

Exps. 1–2 showed no evidence that knowledge about social and gender roles affects re-mention biases. This is not because the event-participant manipulations are too subtle, as the social hierarchy manipulation did significantly affect Brown and Fish causal attribution. It is also unlikely due to insufficient sensitivity in the re-mention tasks, as these were exquisitely sensitive to verb bias and showed ceiling levels of test-retest reliability (Figure 3 and surrounding discussion) despite the fact that the re-mention tasks were not identical (pronoun interpretation vs. sentence continuation). In fact, the re-mention biases collapsed across event participant and experiment correlated extremely well with what was reported for the same verbs in Hartshorne and Snedeker (in press) ($r = 0.97$, $p < 0.001$; Figure 3B), with 92% of verbs showing the same bias (subject- or object-). [13]

Exps. 1–2 showed striking differences between re-mention and causal attribution. Not only were they differently affected by the social hierarchy manipulation, the by-item (and by-verb) correspondence was poor. Particularly striking is that the judgement verbs were subject-biased in causal attribution but object-biased in re-mention.

Experiments 3–4

Although the results of Exps. 1–2 are striking and consistent, one concern is that they are based on only 24 verbs; however, carefully chosen for representativeness. Ferstl et al. (2011) tested over 300 verbs and reported that a subset – particularly those with negative valence – was sensitive to a gender manipulation. This is the only study – beyond the five-verb study of Garvey et al. (1974) to report a significant effect of event-participant manipulation on re-mention biases. Thus, I repeated Exps. 1–2 using verbs chosen from Ferstl et al.'s set.

One methodological improvement was made relative to Exps. 1–2. Those experiments did not include any catch trials that can be used to ensure that participants were paying attention. The fact that the re-mention biases correlated very strongly with the results of Hartshorne and Snedeker (in press) suggests that the controls used were sufficient. Nonetheless, unambiguous filler trials were employed in Exps. 3–4 in order to test for – and remove – participants who were not paying attention or did not understand the task.

Method

Experiment 3

Participants. Eighty participants completed the causal attribution task (47 female; 18–60 years old, M = 33, SD = 12) and 80 participants completed the pronoun interpretation task [40 female (1 no-response);

18–64 years old (1 no-response), M = 34, SD = 12]. An additional 52 participants (18 in causal attribution) were excluded for low accuracy on filler items (see below).

Materials and procedure. The materials and procedure were identical to those of Exp. 1 except as follows. Twenty verbs were chosen from Ferstl et al. (2011) and are listed in Appendix 3: the ten most negatively valenced transitive verbs (according to Ferstl et al.'s ratings) which also showed a numeric male bias (i.e., more attributions to the male character) and the ten most positively valenced transitive verbs which also showed a numeric female bias (i.e., more attributions to the female character).[14] Four lists were made, counterbalanced as in Exps. 1–2.

Ten social hierarchy pairs were used. Each was used for one positively and one negatively valenced verb. Four filler sentences were created so as to be unambiguous, adapted for re-mention (11a) and causal attribution (11b):

(11) a. The reporter believed the actor because he is a gullible person.
Who does 'he' refer to? The reporter The actor

b. The reporter believed the actor because the reporter was gullible.
How likely is it that this happened because:
The reporter is the kind of person who believes people?
Not likely 1 2 3 4 5 6 7 8 9 Definitely likely
The actor is the kind of person who people believe?
Not likely 1 2 3 4 5 6 7 8 9 Definitely likely

The fillers were always the first two or last two items. Participants were excluded from the re-mention task for missing any filler items and from the causal attribution task for not reporting a mean subject-bias greater for the subject-correct fillers than for the object-correct fillers.

Experiment 4

Participants. One hundred and twenty participants completed the causal attribution task (68 female, 18–67 years old (2 no-responses), M = 35, SD = 12) and 120 participants completed the re-mention task (82 female, 18–68 years old (1 no-response), M = 36, SD = 12). An additional 27 participants (16 in causal attribution) were excluded for low accuracy on filler items (see below).

Materials. The materials and procedure were identical to those of Exp. 2 except as follows. The verbs were the 20 verbs used in Exp. 3. The four fillers were created such that the two event participants were of the same gender (2 male, 2 female trials), rendering the re-mention bias task unambiguous (*Christina believed Melissa because.... she/he*?). For the causal attribution task, these sentences were adjusted as to render causality unambiguous (*Christina believed Melissa because Christina was very gullible*), with the correct answer being the subject twice and the object twice. The fillers were always the first two or last two items. Exclusion criteria were as in Exp. 3.

Results

Event-participant manipulations

Results were analysed analogously in Exps. 1–2 and are shown in Figures 4–5. There was a significant effect of social hierarchy on causal attribution (high status bias = + 6, SE = 2; $t = 3.9$, $p < 0.001$), confirming the result of Exp. 1, which was qualified by an interaction with valence ($t = 2.52$, $p = 0.01$), reflecting the fact that the effect was specific to negatively valenced items (high status bias = + 11, SE = 3; $t = 3.9$, $p < 0.001$) and absent in positively valenced items (high status bias = + 2, SE = 3; $t < 1$). In contrast, there was no main effect of social hierarchy on re-mention bias (high status bias = –7, SE = 4; *Wald's* $z < 1$) nor an interaction with valence (*Wald's* $z < 1$). This difference across tasks in the high-status bias for negatively valenced verbs was significant ($t1(158) = 4.42$, $p = 0.00001$; $t2(18) = 2.43$, $p = 0.03$).

Once again, there was no effect of gender on causal attribution (male bias = 0, SE = 1; $t = 1.4$, $p = 0.15$). While there was a marginal interaction of gender and valence ($t = 1.72$, $p = 0.09$), follow-up analyses revealed that there was no significant effect for gender for either positively (M = –1, SE = 1, $t = 1.04$, $p = 0.30$) or negatively valenced verbs (M = + 2, SE = 1, $t = 1.43$, $p =$ 0.15). In contrast, the re-mention task did reveal an effect of gender. The interaction of gender and valence was significant (*Wald's* z = 2.22, p = 0.03). Analysed separately, negatively valenced verbs showed a significant male biases (M = + 8, SE = 3, *Wald's* $z = 3.90$, $p < 0.001$) while positively valenced verbs did not (M = + 1, SE = 4, *Wald's* $z = 0.64$, $p = 0.52$). Three of the ten negatively valenced verbs showed significant male biases ($0.01 \leq ps \leq 0.05$): *deceived*, *loathed* and *dreaded*. There was a significant correlation between the male biases in Exp. 3's pronoun task and male biases in Ferstl et al. ($r = 0.46$, $p = 0.04$). This difference across tasks in the male bias for negatively valenced verbs was significant ($t1(238) = 2.51$, $p = 0.01$; $t2(18) =$ 2.25, $p = 0.04$).

Comparison of re-mention and causal attribution biases

In Exps. 1–2, the correspondence between re-mention and causal attribution biases was limited. What about

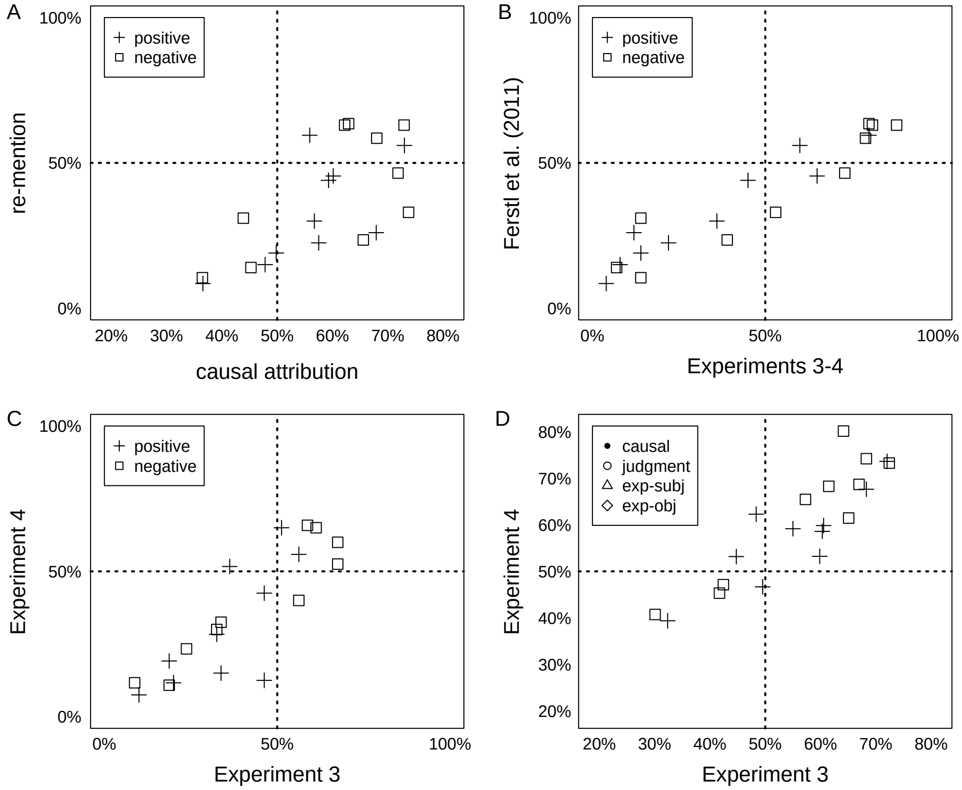

Figure 6. Note: Axes in all panels are percentage of maximum possible subject bias (0% = maximum possible object bias). Panel A: Correlation between re-mention and causal attribution biases, collapsing across event-participant and across Exps. 3–4. Panel B: Correlation between re-mention biases in Ferstl et al. (2011) and in Exps. 3–4 (collapsing across experiment and event-participant). Panel C: Correlation between re-mention biases in Exps. 3–4. Panel D: Correlation between causal attribution biases in Exps. 3–4.

Exps. 3–4? Collapsing across experiment and event participants, the correlation in verb biases was significant ($r = 0.67$, $p = 0.001$; Figure 6A), but only 60% of the verbs had the same bias (subject- or object-) in both tasks, a rate not significantly different from chance ($p = 0.25$, one-tailed). In contrast, the re-mention biases correlated strongly with those reported in Ferstl et al. (2011) ($r = 0.94$, $p < 0.001$; Figure 6B), with 85% of the verbs showing the same bias in both ($p = 0.001$, one-tailed). Similarly, the re-mention biases in Exps. 3–4 correlated strongly with one another ($r = 0.84$, $p < 0.001$; Figure 6C), with 90% of verbs having the same bias in both ($p < 0.001$). The causal attribution biases in Exps. 3–4 likewise correlated strongly ($r = 0.89$, $p < 0.001$; Figure 6D), with 90% of verbs having the same bias in both ($p < 0.001$).

Discussion

Like Exp. 1, Exp. 2 indicated that the social status of the event participant does not affect re-mention bias. This cannot be attributed to the manipulation being too weak, as it did affect causal attribution biases – at least for negatively valenced verbs – or to the data being too noisy, as the verb-specific re-mention biases correlated at near-ceiling rates between experiments and with the data of Ferstl et al. (2011).

In the case of the gender manipulation, the situation is cup-half-full/cup-half-empty. Ferstl et al. (2011) tested 305 verbs and reported that 24 verbs (8%) exhibit significant gender biases. If correcting for multiple comparisons (Sidak method), this drops to 1 verb (0.3%). Indeed, most likely some of the apparently gender-biased verbs in the Ferstl et al. data are false positives, since of the 10 that I retested, only 3 showed significant gender bias (not correcting for multiple comparisons), despite using a significantly larger sample than did Ferstl and colleagues. Thus, while it does appear that gender can modulate re-mention bias, it does so in only a vanishingly small number of cases.

I return to these findings in the General Discussion and discuss them in terms of theories that place the re-mention bias in linguistic structure or high-level cognition.

In terms of the relationship between causal attribution and re-mention biases, Exps. 3–4 – like Exps. 1–2 – indicate that it is much weaker than has been assumed in the literature. The social hierarchy and gender manipulations affect the two phenomena differently. Additionally, while verb-specific causal attribution and re-mention biases do correlate modestly, if one were to predict a verb's re-mention bias based on causal attribution bias or *vice versa* – as is widely done in the literature (e.g., Brown & Fish, 1983; Hartshorne & Snedeker, in press; Rudolph & Forsterling, 1997) – one would be wrong nearly as often as right. This is not due merely to a shift in the distribution – though indeed causal attribution biases are far more subject-biased than re-mention biases (see also Corrigan, 2001) – as there are systematic differences in the biases (see above discussion of judgement verbs in Exps. 1–2). The low correlation is not merely a function of using a different task, as the correlations between pronoun interpretation and sentence continuation tasks was quite strong (compare Figures 3A and 6A with Figures 3C and 6C).

One possible concern is that the poor correlations between causal attribution and re-mention biases are due to irrelevant surface features of the task. We consider two such possibilities in Exps. 5–6.

Experiments 5–6

Exps. 5–6 consider two methodological issues that may have affected the results of Exps. 1–4. First, the above experiments employ different outcome measures (independent Likert scales vs. binary forced choice). In order to better equate the methodologies, in Exp. 5 I replicate the re-mention condition from Exp. 1 using two 9-point Likert scales. This additionally may make the task more sensitive. Second, the original Brown and Fish causal attribution task consists of three questions. I omitted the third (judging whether the event happened for "some other reason") in order to shorten the experiment, as it is never analysed in the literature. However, it may be that the presence of the third question changes how individuals approach the first two. Therefore, Exp. 6 repeats the causal attribution condition of Exp. 4 – the one experiment to find any effect of event participant manipulation on pronoun processing – with the original Brown and Fish design in order to see whether this difference in methodology affects the results – for instance, by making causal attribution more sensitive to the gender manipulation.

Experiment 5

Method

Participants. Sixty-four participants were included (39 female, 1 no-response; 19–74 years old, M = 35, SD = 13). An additional two participants were excluded for giving the same answer to all questions (see Exps. 1–2).

Materials and procedure. Materials and procedure were identical to the re-mention task from Exp. 1 except that the questions looked like the following example:

(12) The butler blamed the duke because he is a froom.
 a. How likely is it that "he" refers to the butler?
  not likely 1 2 3 4 5 6 7 8 9 definitely likely
 b. How likely is it that "he" refers to the duke?
  not likely 1 2 3 4 5 6 7 8 9 definitely likely

Results and discussion

The verb-specific re-mention biases elicited in Exp. 5 correlated extremely well with the re-mention biases from Exp. 2 ($r = 0.97$, $p < 0.001$). The correlation with the causal attribution biases of Exp. 2 ($r = 0.71$, $p < 0.001$) was significantly weaker ($t = 6.0$, $p < 0.001$) and not statistically distinguishable from the correlation between Exp. 2's re-mention and causal attribution biases ($t < 1$). Again, there was no effect of the social hierarchy manipulation (high-status bias = 0%, SE = 2%; $t = 0.3$, $p = 0.74$). Thus, the change in methodology had no appreciable effect on the results.

Experiment 6

Exp. 6 was identical to the causal attribution condition of Exp. 3 except the task was exactly as originally put forth by Brown and Fish (1983).

Method

Participants. Eighty participants were included (37 females; 18–68 years old, M = 34, SD = 13). An additional five participants were excluded for poor performance on the filler trials, following the method outlined in Exps. 3–4.

Materials and procedure. Materials and procedure were identical to the causal attribution task in Exp. 4, except that for each trial, participants were additionally asked to rate on a 1–9 scale how likely it was that the event happened because of "some other reason".

Results and discussion

The verb-specific causal attribution biases elicited in Exp. 6 correlated extremely well with those of Exp. 4 ($r = 0.98$, $p < 0.001$). The correlation with the re-mention biases of Exp. 4 ($r = 0.42$, $p = 0.06$) was much weaker ($t = 6.5$, $p < 0.001$) and was if anything weaker ($t = 2.0$, $p = 0.06$) than the correlation between the causal attribution and re-mention biases elicited in Exp. 4 ($r = 0.50$, $p = 0.03$). As in Exp. 4, there was no effect of the gender manipulation on causal attribution (male bias = 0, SE = 0.2; $t = 1.1$, $p = 0.26$) nor did gender interact with valence ($t < 1$). Thus, the original

Brown and Fish task is no better at predicting the results of re-mention tasks than was the modified version used in Exps. 1–4.

What does the Brown and Fish causal attribution task measure?

This paper set out to evaluate the evidence that implicit causality re-mention biases are a function of high-level, non-linguistic cognition. Along the way, I show that the Brown and Fish causal attribution task – despite expectations to the contrary – is only weakly related to re-mention: Manipulations which affect one do not necessarily affect the other (e.g., event-participant manipulations) and verb-specific biases differ systematically across the phenomena. This removes one crucial argument in favour of the argument that re-mention biases are a function of high-level cognition.

From the narrow purposes of the present paper, there is little more that needs to be said. However, given the outsized role that the Brown and Fish task has played in the study of pronoun interpretation and re-mention biases in general, the reader may reasonably want to know, if the Brown and Fish task does *not* measure the same intuitions about causality driving implicit causality re-mention biases, what *does* is measure? An answer to this question would moreover assure us that the above findings are not due to some yet-undiscovered, uninteresting, superficial difference between the tasks: *Initial* causal judgements are identical, but subsequent processes specific to re-mention or to the Brown and Fish task affect the ultimate judgement and thus muddy the empirical situation.

Fully determining what the Brown and Fish task in fact measures is a large project and beyond the scope of the present paper, which is primarily concerned with other matters. The final two experiments represent an initial step in that direction.

Experiment 7

The Brown and Fish task asks a very specific question, e.g., is it more likely that Sally frightened Mary because Sally is the kind of person who frightens people or because Mary is the kind of person people frighten? To answer this question, one need not think either is particularly likely. Moreover, one may assert that Sally caused Mary to be afraid without agreeing that Sally is the kind of person who frightens people (it was a one-time fluke), and although Mighty Casey is not the kind of person who strikes out, we would not want to absolve him of all causal responsibility when he does strike out. Likewise, consider:

(13) John kicked the ball. How likely is it that this was because:
  a. John is the kind of person who kicks balls.
     Not likely 1 2 3 4 5 6 7 8 9 Definitely likely
  b. Balls are the kind of thing that people kick.
     Not likely 1 2 3 4 5 6 7 8 9 Definitely likely

It is part of the definition of a ball that it is the kind of thing that people kick, but one would not conclude that balls cause events of kicking in quite the same way kickers do.

In order to see whether the Brown and Fish task diverges from intuitions about who caused the event, in Exp. 7, I compare it to a task that directly asks participants who caused the event.

Method

Participants. Forty-eight participants were tested through Amazon Mechanical Turk with the same exclusion criteria as above (29 female (two no response); 18–62 years old (2 no response), M = 37, SD = 14; 3 non-native English speakers). An additional nine participants were excluded for failing to complete all items

Materials and Procedure. Materials and procedure were identical to the causal attribution task in Exp. 1, except that participants were directed to determine who was responsible for the event:

(14) The butler blamed the duke.Who is most likely responsible for this: The butler The duke

Results and discussion

The verb-specific biases of Exp. 7 correlated more strongly with the analogous re-mention biases in Exp. 1 ($r = 0.87$, $p < 0.001$; 92% with same bias) than did the causal attribution biases of Exp. 1 ($r = 0.73$, $p < 0.001$; 71% with same bias), a statistically significant difference ($t = 2.3$, $p = 0.03$).[15] There was no effect of the social hierarchy manipulation in Exp. 7 ($t(46) < 1$), nor did the by-verb high-status biases correlate between Exp. 7 and the causal attribution task of Exp. 1 ($r = 0.28$, $p = 0.18$).

Thus, implicit causality re-mention biases are better predicted by intuitions about who is responsible for the event (Exp. 7) than by the Brown and Fish task. This suggests that the Brown and Fish task does not measure the same causal intuitions employed in re-mention. What, then, does it measure?

One possibility is that it simply does not measure a meaningful construct. The degree to which people believe that an event more likely happened because one person is the sort to engage in that event as opposed to because the other person is the sort to engage in it is measurable, but not necessarily meaningful. Similarly, one can measure the difference in the number of vowels in *Moby Dick* and the number of zebras in Africa; the measurability of a construct does not guarantee its meaningfulness. To demonstrate that the outcome measure of the Brown and Fish task is psychologically meaningful, one would want to show that it predicts some other behaviour above and beyond other available predictors. Note that, in contrast, the re-mention phenomenon is a common and important human behaviour.

Alternatively, the Brown and Fish task might diverge from re-mention because of the kind of explanation of the event it requires. Researchers as far back as Aristotle have noted that there are multiple types of explanations, each invoking different kinds of information (for review, see Lombrozo, 2010). Recently, Bott and Solstad (under review) have argued that the type of explanation can modulate re-mention biases. Consider:

(15) Sally frightened Mary because she was a dax.

While the typical interpretation is that Sally is the dax (Mary finds daxes scary), an equally plausible interpretation is that Sally likes instilling fright in daxes, and Mary happens to be one. Thus, argue Bott and Solstad, a verb-specific re-mention bias is in part a bias for certain types of explanations. By focusing on explaining events in terms of the "type of person" each event participant is, the Brown and Fish task might focus comprehenders on a specific type of explanation. Note that this type of explanation would have to be one not normally considered by comprehenders, since if it was, it would driving interpretation in sentence continuation tasks, in which case one would expect Brown and Fish causal attribution biases to better match sentence continuation biases (cf. Exps. 2 and 4) than they do.

Experiment 8

Modifying the Brown and Fish task to focus on the kind of causal information implicated in re-mention led to results that better matched re-mention than the original Brown and Fish task (Exp. 7). In Exp. 8, I modify a re-mention task to make it focus more on the kind of causal information invoked in the Brown and Fish task. The stimuli from Exps. 2 and 4 were modified as shown below:

(16) The butler blamed the duke because he is the kind of person that…
Who does "he" refer to? The butler The duke

The difference between (16) and the pronoun judgement tasks used above (e.g., *The butler blamed the duke because he is a froom*) is subtle, since both reference the kind of person the responsible party is. However, this reference in (16) is much more explicit. This could make the explanation type more salient, affecting interpretation that way (*cf* Bott & Solstad, under review). This could also make the event-participant manipulation more salient.

Method

Participants. One hundred and eighteen native English speakers older than 13 who reported not being repeat subjects participated (83 female, 15–68 years old, M = 21, SD = 11) were recruited and tested through the experiment portal gameswithwords.org: 69 were tested using the stimuli from Exp. 2 (Appendix 2) and 49 using the stimuli from Exp. 4 (Appendix 3). An additional 71 participants were excluded for missing any unambiguous item. A more lenient screen (missing no more than one item) resulted in qualitatively similar results.

Materials and procedure. The stimuli from the re-mention task in Exps. 2 and 4 were modified as shown in (16). Concluding the sentences with an ellipsis helped avoid providing additional, potentially disambiguating information. This was explained to the participants as a device to make the task more challenging and interesting. The same stimulus lists that counterbalanced whether the subject was high- or low status were employed, with the addition of four filler trials in which the pronoun was unambiguous (*The reporter believed the actor because he is a gullible person*). Participants were randomly assigned to a list, and the order of sentences within the list was randomised separately for each participant.

Results and discussion

Of the 43 verbs tested, five (*condemned*, *disliked*, *mourned*, *supported* and *weakened*) showed significant high-status biases and one (*revived*) showed a significant low-status bias (ps < 0.05). Only two of these survive Sidak correction for multiple comparisons. This was sufficient to drive a small but significant effect of the social hierarchy manipulation (high-status bias: +6; *Wald's* $z = –2.4$, $p = 0.02$) and a significant correlation with the high-status biases of the causal attribution tasks in Exps. 2 and 4 ($r = 0.32$, $p = 0.04$). Both these latter effects lost significance if the four most high-status biased verbs were removed (ps > 0.1).

Although the changes in design led to some limited sensitivity to the event-participant manipulation, this came at the cost of otherwise matching the re-mention task data less well. The overall subject-/object biases of the verbs also became less well-matched to re-mention biases, whether measured by continuation tasks (Exps. 1 and 3: $r = 0.75$, $p < 0.001$) or pronoun judgement tasks (Exps. 2 and 4: $r = 0.81$, $p < 0.001$) than either were to each other ($r = 0.93$, $p < 0.001$; difference: ps < 0.001). Comparing the present results to re-mention biases reported in the literature (Ferstl et al., 2011; Hartshorne & Snedeker, in press) yields nearly identical results.[16] These correlations were not affected by excluding the verbs with significant high-status biases, suggesting that these weaker correlations are not an artifact of increased sensitivity to the event-participant manipulation.

One intriguing possibility is that comprehenders were more focused on explanations involving stable traits; such explanations are not the most typical (hence the weaker correlation with re-mention biases in the neutral sentence-continuation context), but are more sensitive to information about event participants. However, it is also possible that the explicit reference to the "type of person" information helped participants guess the nature of the research question, causing them to engage in extra, top-down conscious processing. The online nature of these experiments means there was no debriefing interview, but one research participant wrote in to spontaneously remark on how the "stature of the person (senator vs. page)" affected pronoun interpretation. Ideally, one would want to rule out this latter possibility, perhaps with an implicit measure that also captures time course information. The re-mention bias appears to guide pronoun interpretation quite rapidly (Cozijn, Commandeur, Vonk, & Noordman, 2011; Pyykkonen & Jarvikivi, 2010), and thus could be dissociated from effects of later, conscious reasoning.

Regardless of the mechanism, altering a pronoun comprehension task to more explicitly require the same kind of causal reasoning as the Brown and Fish task made the results less like those of the most neutral re-mention task (sentence continuation).

General discussion

Combining two linguistic theories – theories of argument structure and of discourse structure – has proven remarkably successful at producing a large number of precise, verified predictions about re-mention biases, including implicit causality re-mention biases (Hartshorne & Snedeker, in press; Hartshorne, in press; Kehler et al., 2008; see also Arnold, 2001; Crinean & Garnham, 2006; Stevenson et al., 1994). Nonetheless, two lines of research reported in the literature have been difficult for the argument structure + discourse structure account to explain: (a) the apparent relationship between re-mention biases and the Brown and Fish causal attribution task, which intuitively invokes high-level, non-linguistic cognition, and (b) effects of (non-linguistic) knowledge about the participants in the event.

Exps. 1–6 demonstrate that the relationship between re-mention biases and the Brown and Fish causal attribution task, while reliable, is much weaker than the relationship between different re-mention bias tasks (sentence continuation and pronoun resolution). In fact, the relationship between the Brown and Fish task and causal judgements itself has been overstated (Exp. 7). The above experiments also show that the event-participant manipulations most widely discussed have limited, if any, effect on re-mention biases.

Below, I first review other findings which have been taken to support placing re-mention biases in high-level, non-linguistic cognition, as well as possible places to look for additional effects. Next, I discuss how theories that place re-mention biases in high-level cognition or squarely within linguistic processing might account for the findings above and in the literature. Finally, I consider how the weak relationship between implicit causality re-mention biases and the Brown and Fish task affects interpretation of the literature.

Additional event-participant manipulations discussed in the literature

On re-mention biases

Beyond gender and social hierarchy manipulations, few other event-participant manipulations have been investigated in terms of re-mention biases. The only additional study comes from Garvey et al. (1974), who manipulated whether the event participants were congruent with the event (e.g., *The prisoner confessed to the guard because he*... vs. *The guard confessed to the prisoner because he*...). Two verbs (*praise* and *criticise*) became significantly more subject-biased when the event participants were incongruent, one became significantly more *object*-biased (*argue with*), and two were not significantly affected (*confess to*, *join*). Thus, if there is a systematic effect, it is complex.

This raises an important point: The above analyses have assumed that the effects of event-participant manipulations are consistent across verbs. In fact, while re-mention biases are generally impervious to the gender manipulation, a very small number of verbs do show a reliable effect (Exp. 4). Could the above analyses have missed complex patterns in the social hierarchy data? I compared the effect of the social hierarchy manipulation on re-mention in Exps. 1 and 5, which used the same stimuli. The correlation in high-status biases by verb was not significant ($r = 0.11$, $p = 0.62$), suggesting that for these stimuli, at least, there is no reliable effect of the manipulation, however complex.

On causal attribution

Though the above review exhausts the event-participant manipulations attempted in the literature, it does not exhaust all *possible* event-participant manipulations. Although we cannot assume that event-participant manipulations which affect causal attribution affect re-mention, it would be reasonable to check the causal attribution literature for ideas. Unfortunately, here, too, little has been done beyond gender and social status manipulations.

Pickering and Majid (2007); see extended quote above) cite two causal attribution studies as evidence that animacy and typicality of event participants affects implicit causality, but whether they even affect causal attribution is unclear. The animacy data, due to Corrigan (1988, 1992) involves comparing sentences like *The custodian pushed the mop* with *The sportscar pushed the woman*. Although it was demonstrated that subjects were more willing to accept "The custodian is the kind of person who pushes things" than that "The sportscar is the kind of thing that pushes people", it is unclear whether this reflects beliefs about animates and inanimates *per se* or just about, e.g., custodians and sportscars and their roles in specific events. The effect of "typicality" is in fact a study by Corrigan (1992) showing *no* relationship between the direction of causal attribution bias and the degree to which the *sentence* is a typical member of the category "sentence."[17]

Two other manipulations tested are more compelling. Kasof and Lee (1993, Exp. 2) report a causal attribution study in which they manipulate whether the sentential subject or object is "you", finding a self-serving bias such that more causality is attributed to oneself for positive than negative events, at least for some verbs. Van Kleeck and colleagues (1988) presented subjects with trials like the following:

(17) Ted admires Bill.
Many other people admire Bill.
Ted admires Bill. Why?
Definitely something about Ted 1 2 3 4 5 6 7 8 definitely something about Bill

As expected, individuals were more likely to attribute this episode of admiration to Bill than if they were told instead that Ted admires many people.

Either of these manipulations could be readily adapted for re-mention, though one potential concern about the van Kleeck manipulation is that it may be obvious to the participants what they "should" say. This concern could be ameliorated by using an implicit dependent measure, such as eye-tracking or reading time. Regardless, whether either of these manipulations would have an effect on re-mention biases is an empirical question that remains to be tested.

Linguistic structure vs. high-level cognition

In this section, I consider how the above results can be accounted for by theories which place the re-mention effect within linguistic processing or within non-linguistic, high-level cognition.

Linguistic structure

The argument structure + discourse structure account has several significant strengths as a scientific theory. First, it makes numerous concrete predictions. Second, it ties together the implicit causality along with a number of other re-mention phenomena into a larger framework (Hartshorne & Snedeker, in press; Kehler et al., 2008; see also Crinean & Garnham, 2006). Finally, it does so without positing any representations or processes that aren't independently motivated; theories of argument structure and discourse structure are required to explain core syntactic phenomena, so if they explain implicit causality as well, they simplify our account. The motivating question in the present study is whether we can, in fact, get away with only invoking linguistic representations and avoid making recourse to general knowledge and higher-level cognition.

As a first approximation, we can. In most cases investigated, event-participant manipulations have no effect, the only clear exceptions being a small percentage of verbs that affected by gender, and, if the broadest interpretation of Exp. 8 is taken, perhaps a small percentage of verbs that are affected by social hierarchy. Thus, most of the time one can correctly predict the results ignoring event participants entirely.

Nonetheless, even exceptions that prove the rule are data that need to be accounted for. One option is to abandon the argument structure + discourse structure account entirely. I explore that possibility in the next section. A second option is that since gender does play other roles in language (such as determining agreement for *he* and *she*), one could develop an account on which it plays a role in re-mention biases, entirely within the linguistic system. A third option, which I explore here, is to attribute effects of gender to revision processes.

As discussed in the introduction, the re-mention effect is a pragmatic inference subject to subsequent revision, not a hard grammatical constraint. Thus, while (18a) reads like a reasonable, if awkward, clarification (18b–c) are much harder to accommodate:

(18) a. Sally frightened Mary because she is a dax – that is, Mary is a dax.
b. Sally frightened Bill because she is a dax – that is, Bill is a dax.
c. Sally frightened Bill because Sally is a dax – that is, Bill is a dax.

In this, re-mention biases are similar to many pragmatic inferences. The fact that the re-mention inference is revisable has been explored in a number of psycholinguistic studies, with the finding that overriding the re-mention bias comes with a processing cost (Caramazza et al., 1977; Koornneef & van Berkum, 2006; Stewart et al., 2000). For instance, bias-congruent sentences (19a) can be read faster than bias-incongruent sentences (19b):

(19) a. Sally frightened Bill because she . . .
b. Sally frightened Bill because he . . .

This processing cost suggests that an initial expectation is being revised. Indeed, eye-tracking studies have shown that re-mention biases can act to guide pronoun interpretation within several hundred milliseconds (Cozijn et al., 2011; Pyykkonen & Jarvikivi, 2010). These considerations form a key part of the argument that there is some re-mention bias independent of final pronoun interpretation that needs to be explained.

This means that one possible mechanism for general knowledge to affect re-mention is via revision processes. The linguistic structures (argument structure and discourse structure) provide an initial re-mention expectation (e.g., pronoun interpretation). In cases where this initial expectation results in a sufficiently implausible interpretation of the sentence given what is known about the event participants that initial expectation can be revised.

This account has the advantage of requiring relatively little adjustment to the theory in order to account what are apparently rare – barring the future discoveries – effects of general knowledge, and requires only mechanisms which must already be posited to handle other phenomena (3, 18). However, at the moment it is *ad hoc* and un-tested. One way of testing this hypothesis would be to study online processing, such as in a Visual World Paradigm eye-tracking study (e.g., Cozijn et al., 2011; Pyykkonen & Jarvikivi, 2010). If the occasional effect of event participants is due to top-down revision processes, one might expect to see it appear only after an initial, verb bias-consistent interpretation of the pronoun appears, assuming the top-down processes are not so rapid as to render undetectable the initial interpretation. In fact, the present study began as pilot work for just such an experiment. It should be obvious why that experiment has not yet been run: with only a few stimuli that show reliable effects of event-participant manipulations, and given that those effects are relatively small, back-of-the-envelop calculations suggest that an eye-tracking study would need hundreds of subjects to have a reasonable statistical power. Thus, such a study must wait discovery of additional stimuli

High-level cognition

An alternative account of implicit causality re-mention biases is that the relevant causal information is inferred from a high-level, non-linguistic representation of the event. Brown and Fish (1983) and many subsequent researchers assumed that the sentences they studied did not directly encode causality. For that reason, it was necessary to explain how causation was inferred, and this motivated recourse to theories of high-level cognition. However, the last several decades of research in linguistics suggests that many verbs do directly encode causality, leaving less to be explained.

Nonetheless, there is something to be explained, namely the occasional effects of general knowledge such as those embodied in event-participant manipulations. That is the greatest advantage of this account. Unfortunately, such effects are – barring future discoveries – so rare as to render that advantage superfluous most of the time. If additional effects were identified – particularly if they were broadly applicable – that would provide a strong argument for placing more the explanatory burden in high-level cognition.

Another way in which this account could gain ground is to become further specified and, importantly, constrained. This theory can account for, e.g., effects of event-participant manipulations only in the sense that it predicts that whatever matters for re-mention biases matters for re-mention biases. A theory that can account for anything predicts nothing, and as such is not much more valuable than a highly constrained but incorrect theory. It may actually be less valuable, since there is something to be learned in the ways an incorrect theory is wrong.

Re-evaluation of the literature

The literature has generally assumed that findings true of the Brown and Fish task generalise to re-mention and *vice versa*. How does our understanding of the literature change given that such generalisation turns out not to be justified? The fact that event-participant manipulations known to affect the Brown and Fish task cannot be assumed to affect re-mention has already been mentioned. Below, I describe four other important implications.

First, much of the research in implicit causality has been devoted to finding a verb taxonomy that would accurately predict verb biases (Hartshorne & Snedeker, in press; Rudolph & Forsterling, 1997). Until recently, these taxonomies were developed primarily based on data from the Brown and Fish task. Hartshorne and Snedeker (in press) reported that these taxonomies generally performed at chance levels, whereas a taxonomy based on argument structure patterns performed much better. Hartshorne and Snedeker concluded that the older taxonomies were likely over-fit to small samples of verbs. However, it may be that one (or more) of these taxonomies actually fits the Brown and Fish data quite well. As no large-scale studies analogous to Hartshorne and Snedeker (in press) or Ferstl et al. (2011) have been conducted for causal attribution, answering this question remains for future work.

Second, much of the recent research on implicit causality in psycholinguistics has focused on when implicit causality biases begin to drive pronoun interpretation (e.g., Cozijn et al., 2011; Guerry, Gimenes, Caplan, & Rigalleau, 2006; Featherstone & Sturt, 2010; Koornneef & van Berkum, 2006; Pyykkonen & Jarvikivi, 2010; Stewart et al., 2000). Most studies have reported that implicit causality affects pronoun interpretation quite rapidly, and some suggest that implicit causality information is available soon after encountering the verb. These results cannot be generalised to the Brown and Fish causal attribution task – both for the trivial reason that we can no longer assume that results generalise from the one phenomenon to the other and because what these studies show is that the re-mention bias observable in off-line judgements is already detectable very early in online measures. Since those off-line re-mention biases do not match the Brown and Fish biases, by definition these studies do not show that the biases seen offline in Brown and Fish tasks are detectable in early online processing. How quickly the information governing the Brown and Fish causal attribution is available remains an open question.

Third, implicit causality has been used to probe multiple aspects of non-linguistic cognition, such as the development of causal schemas in children (Au, 1986; Corrigan & Stevenson, 1994), the stability of these schemas across cultures (Brown & Fish, 1983), and the conceptualisation of social relationships and dominance hierarchies (Corrigan, 2001; LaFrance et al., 1997). To the extent these studies are conducted using the Brown and Fish task, this may be appropriate.

Finally, these results raise important questions as to what exactly the Brown and Fish task measures, if anything. Although Brown and Fish offered it as an assay of the causal inferences underlying the re-mention bias, it does not appear to be. Nor does it map well onto judgements as to who is responsible for an interpersonal event (Exp. 7). Brown and Fish (1983) reported that verbs that were subject-biased on their task were more likely to have subject-derived adjectives (*helpful*, *cheating*) and verbs object-biased on their task were more likely to have object-derived adjectives (*likable*, *noticeable*). They suggested that this was evidence of thought (biases as to typical causes of events) shaping language (specifically, vocabulary). However, these results were based on a small set of 36 verbs, for which re-mention and causal attribution biases correlate fairly strongly (see Introduction), which raises the question of which bias best predicts the derived adjective patterns. Answering that question will require a more comprehensive study.

Conclusion

If there is no unified phenomenon "implicit causality" and if the re-mention effect is largely a function of linguistic structure rather than more general and less constrained cognitive processes, does that make it less interesting? Not at all! For one, the fact that the linguistic system is able to approximate complex inferences with "simple" representations actually represents a very smart solution to a difficult problem. Given the speed at which language comprehension needs to proceed, any shortcuts to approximately correct inferences is potentially quite helpful, especially if those inferences can be revised later as necessary. As we understand language and human cognition better, we may find more and more such examples (*cf* Chierchia, Fox, & Spector, in press; Levinson, 2000). Second, the re-mention biases are emerging as a complex but relatively well-understood phenomenon where relatively accurate predictions can be made for a wide range of situations. Every additional such phenomenon greatly enhances our understanding of how language – and the human mind – works.

Acknowledgements

The author wishes to thank Jesse Snedeker, Steven Pinker, Susan Carey, Josh Tenenbaum, Manizeh Khan, Hugh Rabagliati, Timothy O'Donnell, Rebecca Nappa, Patrick Sturt and several anonymous reviewers for comments and discussion. Portions of this work were presented at the 2011 CUNY Conference on Human Sentence Proncessing and benefitted from discussion there. This material is based on work supported by a National Defense Science and Engineering Graduate Fellowship and the Gordon W. Allport Memorial Fund.

Notes

1. Psycholinguists have been particularly interested in implicit causality as it relates to pronoun interpretation. However, given that implicit causality affects production – even when a pronoun is not used – I will use the more inclusive terms "re-mention effect" and "re-mention bias".
2. Not all approaches to all pragmatic phenomena follow this pattern (*cf* Chierchia, et al., in press).
3. "Our main finding may seem to be a Whorfian one, a demonstration that language affects thought. We think it is not that but is, rather, a demonstration that a mode

of thought that is universally human affects language use" (Brown & Fish, 1983, p. 271).

4. Though Brown and Fish (1983) and Ferstl et al. (2011) note similarity in findings across the tasks, neither they nor Corrigan (1988) report these correlations; I calculated them from data presented in tables and appendices. Note: Corrigan tested each verb in five conditions; as there were large effects of condition, each result is treated separately in this analysis.
5. In addition to Brown and Fish causal attribution, researchers have asserted that "implicit causality" explains a number of other phenomena (e.g., Blankenship & Craig, 2012). There is even less evidence linking these phenomena to re-mention biases and the phenomena themselves are much less well-established than causal attribution. Though worthy of investigation, they will not be discussed further here.
6. This analysis is not reported, but can be reconstructed from the tables.
7. The relevant analyses are not discussed in the text, but can be reconstructed from the tables.
8. There was complex effect of gender on participants' confidence in those interpretations. While intriguing, it is not clear how to interpret this result.
9. As revealed in Table 1. This finding is not discussed in the paper, nor are any statistical analyses reported.
10. These analyses are calculated from the condition means in Tables 1 and 2.
11. Though many of these verbs have appeared frequently in the implicit causality literature and all of them appeared in Hartshorne and Snedeker (in press), the classification system is different from the ones familiar to most implicit causality researchers. The classes are derived instead from VerbNet (Kipper, Korhonen, Ryant & Palmer, 2006), which developed independently of the implicit causality literature, and are classes 31.1 (experiencer–object emotion verbs), 31.2 (experiencer–subject emotion verbs), 33 (judgement verbs) and 45.1 (causal verbs).
12. Except where otherwise specified, statistical analyses were mixed effects linear models (continuous or binomial, as appropriate) with maximal random effects. Each causal attribution trial was converted to a difference score (attribution to subject – attribution to object) prior to analyses. For continuous models, I estimate *p*-values treating *t*-values as normally distributed (Barr, Levy, Scheepers & Tily, 2013), whereas for binomial models, I use Wald's z on the normal distribution (Baayen, 2008).
13. Biases for causal and judgement verbs were reported in Exp. 1 of Hartshorne and Snedeker (in press). For the experiencer–subject and experiencer–object verbs, I used the results in Exp. 2.
14. In order to balance both the needs of choosing strongly valenced verbs and verbs with male or female biases, *mourned* was included because it is strongly negatively valenced, though it showed a male bias of exactly 0. Note also that one of the verbs – admired – was also used in Exps. 1–2.
15. The verb-specific biases of Exp. 7 also correlated well with the causal attribution biases of Exp. 1 ($r = 0.83$, $p < 0.001$; 80% with same bias), though not significantly better that the correlation between the tasks in Exp. 1 ($t = 1.7$, $p = 0.11$).
16. Interestingly, the correlation with causal attribution data remained high but below ceiling ($r = 0.82$, $p < 0.001$; 82% with same bias; based on Exps. 1 and 3), suggesting that the task in Exp. 8 retained some differences to the causal attribution task.
17. There is a positive correlation between typicality of sentence and the combined causal weight of the subject and object, such that for more typical sentences, *both* the subject and object are judged to be more causal.

References

Ambridge, B., Pine, J. M., Rowland, C. F., Chang, F., & Bidgood, A. (2013). The retreat from overgeneralization in child language acquisition: Word learning, morphology, and verb argument structure. *Wiley Interdisciplinary Reviews: Cognitive Science*, *4*(1), 47–62. doi:10.1002/wcs.1207

Arnold, J. E. (2001). The effect of thematic roles on pronoun use and frequency of reference continuation. *Discourse Processes*, *31*(2), 137–162. doi:10.1207/S15326950DP3102_02

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. doi:10.1016/0749-596X(86)90024-0

Baayen, R. H. (2008). *Analyzing linguistic data*. Cambridge: Cambridge University Press.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. doi:10.1016/j.jml.2012.11.001

Blankenship, K. L., & Craig, T. Y. (2012). Something about Mary: Information processing and the persistence of implicit causality. *Social Cognition*, *30*(1), 71–93. doi:10.1521/soco.2012.30.1.71

Bott, O., & Solstad, T. (under review). *From verbs to discourse – A novel account of implicit causality.*

Brown, R., & Fish, D. (1983). The psychological causality implicit in language. *Cognition*, *14*(3), 237–273. doi:10.1016/0010-0277(83)90006-9

Brown, R., & van Kleeck, M. H. (1989). Enough said: Three principles of explanation. *Journal of Personality and Social Psychology*, *57*(4), 590–604. doi:10.1037/0022-3514.57.4.590

Caramazza, A., Grober, E., Garvey, C., & Yates, J. (1977). Comprehension of anaphoric pronouns. *Journal of verbal learning and verbal behavior*, *16*(5), 601–609. doi:10.1016/S0022-5371(77)80022-4

Clark, H. H. (1996). *Using Language*. Cambridge: Cambridge University Press.

Chierchia, G., Fox, D., & Spector, D. (in press). The grammatical view of scalar implicatures and the relationship between semantics and pragmatics. In P. Portner, C. Maienborn, & K. von Heusinger (Eds.), *Handbook of Semanticsk*, Vol. 3. Mouton de Gruyter.

Corrigan, R. (1988). Who dun it? The influence of actor-patient animacy and type of verb in the making of causal attributions. *Journal of Memory and Language*, *27*(4), 447–465. doi:10.1016/0749-596X(88)90067-8

Corrigan, R. (1992). The relationship between causal attributions and judgements of the typicality of events described by sentences. *British Journal of Social Psychology*, *31*(4), 351–368. doi:10.1111/j.2044-8309.1992.tb00978.x

Corrigan, R. (2001). Implicit causality in language: Event participants and their interactions. *Journal of Language and Social Psychology*, *20*(3), 285–320. doi:10.1177/0261927X01020003002

Corrigan, R. (2002). The influence of evaluation and potency on perceivers' causal attributions. *European Journal of Social Psychology*, *32*(3), 363–382. doi:10.1002/ejsp.96

Corrigan, R. (2003). Preschoolers' and adults' attributions of who causes interpersonal events. *Infant and Child Development*, *12*(4), 305–328. doi:10.1002/icd.291

Corrigan, R., & Stevenson, C. (1994). Children's causal attributions to states and events described by different classes of verbs. *Cognitive Development*, *9*(2), 235–256. doi:10.1016/0885-2014(94)90005-1

Cozijn, R., Commandeur, E., Vonk, W., & Noordman, L. G. M. (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. doi:10.1016/j.jml.2011.01.001

Crinean, M., & Garnham, A. (2006). Implicit causality, implicit consequentiality and semantic roles. *Language and Cognitive Processes*, *21*(5), 636–648. doi:10.1080/01690960500199763

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. doi:10.1080/17470210903134344

Ferstl, E. C., Garnham, A., & Manouilidou, C. (2011). Implicit causality bias in English: A corpus of 300 verbs. *Behavior Research Methods*, *43*(1), 124–135. doi:10.3758/s13428-010-0023-2

Garvey, C., & Caramazza, A. (1974). Implicit causality in verbs. *Linguistic Inquiry*, *5*(3), 459–464.

Garvey, C., Caramazza, A., & Yates, J. (1974). Factors influencing assignment of pronoun antecedents. *Cognition*, *3*(3), 227–243. doi:10.1016/0010-0277(74)90010-9

Gleitman, L. (1990). The structural sources of verb meanings. *Language Acquisition*, *1*(1), 3–55. doi:10.1207/s15327817la0101_2

Goikoetxea, E., Pascual, G., & Acha, J. (2008). Normative study of implicit causality in 100 interpersonal verbs in Spanish. *Behavior Research Methods, Instruments, & Computers*, *40*, 760–772. doi:10.3758/BRM.40.3.760

Goodman, N. D., & Stuhlmuller, A. (2013). Knowledge and implicature: Modeling language understanding as social cognition. *Topics in Cognitive Science*, *5*(1), 173–184. doi:10.1111/tops.12007

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. doi:10.1111/j.1467-9280.1995.tb00509.x

Grice, P. (1989). *Studies in the way of words*. Cambridge, MA: Harvard University Press.

Guerry, M., Gimenes, M., Caplan, D., & Rigalleau, F. (2006). How long does it take to find a cause? An online investigation of implicit causality in sentence production. *The Quarterly Journal of Experimental Psychology*, *59*(9), 1535–1555. doi:10.1080/17470210500269105

Hartshorne, J. K. (in press). Are implicit causality pronoun resolution biases consistent across languages and cultures? *Experimental Psychology.*

Hartshorne, J. K., & Snedeker, J. (in press). Verb argument structure predicts implicit causality: The advantages of finer-grained semantics. *Language and Cognitive Processes.*

Holtgraves, T., & Raymond, S. (1995). Implicit causality and memory: Evidence for a priming model. *Personality and Social Psychology Bulletin*, *21*(1), 5–12. doi:10.1177/0146167295211002

Jackendoff, R. (1990). *Semantic structures*. Cambridge, MA: The MIT Press.

Kako, E. (2006). Thematic role properties of subjects and objects. *Cognition*, *101*(1), 1–42. doi:10.1016/j.cognition.2005.08.002

Kamp, H., van Genabith, J., & Reyle, U. (2011). Discourse representation theory. In D. Gabbay & F. Guenthner (Eds.), *Handbook of philosophical logic* (pp. 125–394). New York, NY: Springer.

Kasof, J., & Lee, J. Y. (1993). Implicit causality as implicit salience. *Journal of Personality and Social Psychology*, *65*(5), 877–891. doi:10.1037/0022-3514.65.5.877

Kehler, A. (2002). *Coherence, reference, and the theory of grammar*. Stanford, CA: CSLI Publications.

Kehler, A., Kertz, L., Rohde, H., & Elman, J. L. (2008). Coherence and coreference revisited. *Journal of Semantics*, *25*(1), 1–44. doi:10.1093/jos/ffm018

Kelley, H. H. (1967). Attribution theory in social psychology. In D. Levine (Ed.), *Nebraska symposium on motivation* (Vol. 15, pp. 192–238). Lincoln: University of Nebraska Press.

Kelley, H. H., & Michela, J. L. (1980). Attribution theory and research. *Annual Review of Psychology*, *31*(1), 457–501. doi:10.1146/annurev.ps.31.020180.002325

Kipper, K., Korhonen, A., Ryant, N., & Palmer, M. (2006). Extending VerbNet with novel verb classes. *Proceedings of the Fifth International Conference on Language Resources and Evaluation*.

Koornneef, A., & 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*(4), 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*(2), 138–152. doi:10.2307/2787101

Lee, J. J., & Pinker, S. (2010). Rationales for indirect speech: The theory of the strategic speaker. *Psychological Review*, *117*(3), 785–807. doi:10.1037/a0019688

Levin, B. (1993). *English verb classes and alternations: A preliminary investigation*. Chicago, IL: University of Chicago Press.

Levin, B., & Rappaport Hovav, M. (2005). *Argument realization*. Cambridge: Cambridge University Press.

Levinson, S. (2000). *Presumptive meanings: The theory of generalized coversational implicature*. Cambridge, MA: MIT Press.

Lombrozo, T. (2010). Causal-explanatory pluralism: How intentions, functions, and mechanisms influence causal ascriptions. *Cognitive Psychology*, *61*(4), 303–332. doi:10.1016/j.cogpsych.2010.05.002

Mannetti, L., & De Grada, E. (1991). Interpersonal verbs: Implicit causality of action verbs and contextual factors. *European Journal of Social Psychology*, *21*(5), 429–443. doi:10.1002/ejsp.2420210506

McArthur, L. A. (1972). The how and what of why: Some determinants and consequences of causal attributions. *Journal of Personality and Social Psychology*, *22*(2), 171–193. doi:10.1037/h0032602

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

Noveck, I. A., & Reboul, A. (2008). Experimental pragmatics: A Gricean turn in the study of language. *Trends in Cognitive Sciences*, *12*(11), 425–431. doi:10.1016/j.tics.2008.07.009

Pesetsky, D. (1995). *Zero syntax: Experiencers and cascades*. Cambridge, MA: The MIT Press.

Pickering, M. J., & Majid, A. (2007). What are implicit causality and consequentiality? *Language & Cognitive Processes*, *22*(5), 780–788. doi:10.1080/01690960601119876

Pinker, S. (1989). *Learnability and cognition*. Cambridge, MA: The MIT Press.

Pynte, J., Kennedy, A., Murray, W. S., & Courrieu, P. (1988). The effects of spatialisation on the processing of ambiguous pronominal reference. In G. Lueer, U. Lass & J. Shallo-Hoffman (Eds.), *Eye movement research: Physiological and psychological aspects* (pp. 214–225). Toronto: Hogrefe and Huber.

Pyykkönen, P., & Järvikivi, J. (2010). Activation and persistence of implicit causality information in spoken language comprehension. *Experimental Psychology*, *57*(1), 5–16. doi:10.1027/1618-3169/a000002

Rudolph, U. (2008). Covariation, causality, and language: Developing a causal structure of the social world. *Social Psychology*, *39*(3), 174–181. doi:10.1027/1864-9335.39.3.174

Rudolph, U., & Forsterling, F. (1997). The psychological causality implicit in verbs: A review. *Psychological Bulletin*, *121*(2), 192–218. doi:10.1037/0033-2909.121.2.192

Semin, G. R., & Fiedler, K. (1991). The linguistic category model, its bases, applications and range. *European Review of Social Psychology*, *2*, 1–30. doi:10.1080/14792779143000006

Sperber, D., & Wilson, D. (1986). *Relevance: Communication and cognition*. Oxford: Basil Blackwell.

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. doi:10.1080/01690969408402130

Stewart, A. J., & Pickering, M. J. (1998). Implicit consequentiality. In M. A. Gernsbacher & S. J. Derry (Eds.), *Proceedings of the 20th Annual Conference of the Cognitive Science Society* (pp. 1031–1036). Mahwah, NJ: 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*(3), 423–443. doi:10.1006/jmla.1999.2691

van Kleeck, M. H., Hillger, L. A., & Brown, R. (1988). Pitting verbal schemas against information variables in attribution. *Social Cognition*, *6*(2), 89–106. doi:10.1521/soco.1988.6.2.89

Vorster, J. (1985). Implicit causality in language: Evidence from Afrikaans. *South African Journal of Psychology*, *15*(2), 62–67. doi:10.1177/008124638501500203

Wolf, F., & Gibson, E. (2006). *Coherence in natural language: Data structures and applications.* Cambridge, MA: The MIT Press.

Appendix 1: Argument structure & implicit causality

Early discussion of argument structure in the implicit causality focused on thematic role theories (Brown & Fish, 1983; Crinean & Garnham, 2006; Rudolph & Forsterling, 1997; see also Arnold, 2001). On such accounts, every argument of a verb fits into one of a short list of "thematic roles", which are abstract generalisations of the role an entity can play in an event (e.g., AGENT, PATIENT, EXPERIENCER, STIMULUS). Below are examples of the thematic roles assigned by Brown and Fish (1983) to several verbs:

(I) a. Sally frightened Mary.
b. STIMULUS V EXPERIENCER
(II) a. Sally loved Mary.
b. EXPERIENCER V STIMULUS
(III) a. Sally criticised Mary.
b. AGENT V PATIENT

Some of these roles are inherently causal (like AGENT, an animate causal actor, and STIMULUS, the source/cause of a mental state), and it was argued that it is exactly such is entities playing such roles that attract causal biases in implicit causality tasks.

Although subsequent to Brown and Fish (1983) thematic role theories developed largely independently in the argument structure and implicit causality literatures, they faced many of the same problems. One was the difficulty in finding a definitive set of thematic roles (Levin & Rappaport Hovav, 2005; Rudolph & Forsterling, 1997). A related but deeper problem is that thematic role theories provide a single label for each argument, whereas different phenomena sometimes appear to call for different labels. In the implicit causality literature, this arose first in the context of implicit consequentiality (Crinean & Garnham, 2006; Stewart & Pickering, 1998; Pickering & Majid, 2007):

(IV) a. Sally frightened Mary because/so she…
b. Sally feared Mary because/so she…
c. Sally criticised Mary because/so she…

Whereas for implicit causality, the relevant argument is the causal one, for implicit consequentiality, it appears to be the affected entity. In some cases (*fear*, *frighten*) these are different entities, whereas for others (*criticise*), they seem to be the same. One option is to create a greater number of more specific thematic roles – e.g., label the object of *criticise* a CAUSAL AFFECTED ENTITY. However, many researchers working on argument structure have simply abandoned thematic roles as atomic primitives (Levin & Rappaport Hovav, 2005).

One particularly successful line of work falls under the rubric "predicate decomposition" (see especially Jackendoff, 1990; Levin, 1993; Pinker, 1989). On these accounts, the core aspects of verbs semantics – specifically, those parts relevant to argument structure – are built out of more primitive predicates. Here, for instance, is a possible deconstruction of *criticise* (see Hartshorne & Snedeker, in press):

(V) a. Sally criticised Mary.
b. DECLARE (DURING(E), Sally, Mary, CRITICISM-WORTHY) IN_RESPONSE_TO (BEFORE(E), Mary, ACTION)

where *E* is the event described by criticise and *ACTION* is some prior action Mary undertook, that is now being criticised. What distinguishes *criticise* from *praise* or *euologise* is the final argument of the DECLARE predicate (CRITICISM-WORTHY).

Note that in thematic role terms, this decomposition leaves Mary as the patient of the DECLARE predicate (and thus, for the purposes of implicit consequentiality, the likely recipient of any consequences) but the stimulus of the IN_RESPONSE_TO predicate (and thus, for the purposes of implicit causality, the cause).

Hartshorne and Snedeker (in press) have argued that predicate decomposition is considerably more successful at explaining implicit causality re-mention biases as well as explaining the existence of different biases in different discourse contexts.

Bott and Solstad (under review) have recently proposed that the semantic representations of Discourse Representation Structure (Kamp, van Genabith, & Reyle, 2011) provide an even better fit to re-mention phenomena. Although the details have not been fully reported, it promises to account for the differential effects of different explanation types described in the General Discussion.

Appendix 2

The high-status subject sentences from the causal attribution task in Exp. 2 are given below. Low-status subject sentences were created by reversing the subject and object.

Causal verbs:
The boss balanced the employee
The CEO improved the clerk
The king revived the knight
The master softened the apprentice
The senator strengthened the page
The duke weakened the butler

Judgement verbs:
The duke blamed the butler
The senator condemned the page
The king criticised the knight
The master cursed the apprentice
The boss denounced the employee
The CEO excused the clerk

Experiencer-Subject verbs:
The CEO admired the clerk
The master despised the apprentice
The senator disliked the page
The boss hated the employee
The duke resented the butler
The king respected the knight

Experiencer-Object verbs:
- The duke affected the butler
- The king aroused the knight
- The CEO bored the clerk
- The senator frustrated the page
- The boss puzzled the employee
- The master satisfied the apprentice

Appendix 3

The high-status subject sentences from the causal attribution task in Exp. 6 are given below. Low-status subject sentences and re-mention-bias sentences were created as described in the method section for Exp. 4.

Negatively valenced:
- The CEO deceived the clerk
- The general distressed the soldier
- The boss dreaded the employee
- The king killed the knight
- The senator loathed the page
- The master mourned the apprentice
- The general persecuted the soldier
- The parent plagued the child
- The king repulsed the knight
- The teacher tormented the student

Positively valenced:
- The master admired the apprentice
- The boss applauded the employee
- The duke celebrated the butler
- The CEO comforted the clerk
- The parent complemented the child
- The duke cuddled the butler
- The manager embraced the intern
- The manager kissed the intern
- The teacher relaxed the student
- The senator supported the page