UC Merced
Proceedings of the Annual Meeting of the Cognitive Science Society

Title
When does suggestive language shape memory for car accidents? Assessing the role of elaboration and pragmatics in a classic framing effect

Permalink
https://escholarship.org/uc/item/0wk430xs

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 46(0)

Authors
Holmes, Kevin J.
Kassin, Lena
Flusberg, Stephen

Publication Date
2024

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

When does suggestive language shape memory for car accidents? Assessing the role of elaboration and pragmatics in a classic framing effect

Kevin J. Holmes (kjholmes@reed.edu)
Lena Kassin (lenakassin@reed.edu)
Department of Psychology, Reed College
3203 SE Woodstock Blvd, Portland, OR 97202, USA

Stephen J. Flusberg (sflusberg@vassar.edu)
Department of Cognitive Science, Vassar College
124 Raymond Ave, Poughkeepsie, NY 12604, USA

Abstract

Does linguistic framing shape memory for consequential events? An influential study by Loftus and Palmer (1974) found that people estimated higher speeds when asked how fast the vehicles involved in an accident were going when they *smashed* (vs. *hit*) each other. This finding has proven difficult to replicate, however. Based on a key difference between the original study and previous replications, as well as recent work on linguistic framing, we hypothesized that verbal elaboration and pragmatic inference might moderate this classic effect. In two experiments (N = 1204), participants viewed a brief car accident video. They either wrote a verbal description of the event or did not before answering the verb-framed speed question. Participants who wrote longer descriptions and inferred a greater difference in intensity between the two verb frames were less likely to show the expected framing effect. These findings advance our understanding of how suggestive language influences recollections.

Keywords: framing; language; eyewitness memory; verbal elaboration; pragmatic inference; replication

Introduction

Suppose you witness a head-on collision between two cars and are later asked to recount what you saw to the police or legal authorities. According to an influential body of research on language and eyewitness memory, the way the question is framed could have a strong impact on what you recall (Brainerd & Reyna, 2005; Frenda et al., 2011). For example, you may misremember the accident if the question contains a presupposition (e.g., "*How fast was the white car going when it ran the stop sign?*"; Loftus, 1975), even when asked directly about a key detail (e.g., "*Did you see the broken headlight?*"; Loftus & Zanni, 1975). Framing effects like these are striking because they challenge the conventional wisdom that our memories are precise replicas of our experiences (Simons & Chabris, 2011). They also have important implications for understanding the nature of memory, the relationship between language and thought, and the trustworthiness of real-world eyewitness testimony.

Many of the seminal studies in this area were conducted by Elizabeth Loftus and her colleagues in the 1970s. One of Loftus' most famous findings concerns the influence of verb intensity on memory for car accidents. Loftus and Palmer (1974; henceforth, L&P) showed participants several brief video clips of such accidents. After each one, participants wrote an account of the accident and then estimated the speed of the cars involved, among other questions. The main verb in the speed question varied across conditions: "*About how fast were the cars going when they [smashed into/collided with/bumped into/hit/contacted] each other?*" When the verb implied a more severe accident (e.g., *smashed*), participants estimated higher speeds. In a follow-up experiment, L&P replicated this framing effect using a single accident clip and the verbs *smashed* and *hit*. They also showed that, compared to *hit*, *smashed* led many participants to mistakenly report seeing broken glass in the video when asked a week later.

To date, L&P's article has been cited more than 3,400 times according to Google Scholar, with nearly 1,500 citations in the past 10 years. Yet despite its enduring impact, the findings have not been consistently replicated. Some studies have failed to observe the effect of verb framing entirely (McAllister et al., 1988; Raghunath et al., 2021; Read et al., 1978). For example, Read et al. (1978) found no difference in participants' car speed estimates in response to a question containing *smashed*, *hit*, or *bumped*. Other studies have found that the framing effect is limited to certain contexts (e.g., when the setting of the accident is unfamiliar; Read & Bruce, 1984) or certain populations (e.g., 7th and 8th graders, but not younger children or college students; Lipscomb et al., 1985), with little theoretical rationale for the differences. Conceptual replications of L&P (e.g., a hockey collision described as a "smash" or a "bump"; Goldschmied et al., 2017) have been similarly inconclusive.

What explains this mixed record of replication? One possible factor is that, to our knowledge, all published replication attempts have omitted a key element of L&P's procedure: the account of the accident written by participants immediately after the video, before making their speed estimate. This methodological difference—unrecognized in the literature—may be important because research on *verbal overshadowing* shows that freely describing a previously seen event can impair recollection by biasing memory toward verbalized details (Alogna et al., 2014; Schooler & Engstler-Schooler, 1990). In L&P's experiments, participants' verbal descriptions may have overshadowed their visual memory of the accident, leading them to weight the verb frame more heavily in their speed estimates.

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.* ©2024 The Author(s). This work is licensed under a Creative Commons Attribution 4.0 International License (CC BY).

An alternative account of the impact of verbal description is suggested by contemporary research on linguistic framing. Framing effects are often reduced when people have a stronger or more elaborate representation of the framed situation (Flusberg et al., in press). In one study, for example, participants judged an NBA basketball player as more valuable when he was described as "making 60%" of his free throws than when he was described as "missing 40%." However, this effect was eliminated for those with extensive NBA knowledge (Leong et al., 2017). In a similar fashion, writing a detailed description of a just-seen car accident may result in a more elaborate representation of the event (cf. Klatzky et al., 1982), making participants less susceptible to subsequent reframing by a single verb. On this account, the inconsistent reproducibility of L&P's findings may be due in part to variation in the extent to which participants verbally elaborate on the accident before making their speed estimates.

Recent framing research points to another factor that might be at play: *pragmatic inference*. Many framing effects stem from the ability to "read between the lines" and infer what the speaker or writer intended to communicate through their choice of words (Flusberg et al., 2022; Sher & McKenzie, 2006). For example, the aforementioned effect of framing on basketball player evaluations has been linked to the inference that one of the two seemingly equivalent frames ("makes 60%" or "misses 40%") was chosen over the other to communicate that the player's performance is better or worse than average (Leong et al., 2017). Similarly, statements like "girls are just as good as boys at math" communicate that the group in the complement position ("boys") sets the standard for the other group (Chestnut & Markman, 2018; Holmes et al., 2022), and people who are more sensitive to this implication show stronger framing effects from such statements (Holmes et al., under review). Pragmatic inference may likewise contribute to L&P's verb framing effect. Specifically, the strength of the effect may be predicted by what individual participants believe the verbs in the speed question imply about the severity of an accident.

In sum, there are methodological differences between L&P's study and subsequent replication attempts, as well as moderating factors identified in other work that might help explain the conflicting findings. Across two experiments, we sought to systematically assess the impact of these factors. Participants viewed a single video clip of a car accident and were later asked to estimate how fast the cars were going. We manipulated not only which verb appeared in the speed question, but also whether participants wrote a description of the accident before answering this question. Additionally, we examined whether the amount of detail in participants' descriptions and their pragmatic inferences about the verbs of interest predicted the magnitude of the framing effect.

We preregistered our methods and analysis plans on AsPredicted. However, some of our key findings are exploratory, as noted below. Our preregistrations, materials, and data are available on the Open Science Framework (https://osf.io/x7ntz/).

Experiment 1

In Experiment 1a, participants watched an accident video, described the accident in writing or completed an unrelated task, and then answered the critical speed question and other specific questions about the accident. The speed question included either the verb *smashed* or *hit*. L&P found that *smashed* yielded higher speed estimates.

In Experiment 1b, we attempted to strengthen the verbal description manipulation by increasing the length of the description task and inserting a delay before it. Prior work suggests that a verbal description is more likely to "overshadow" visual memory when it is provided some time after an event is witnessed (Alogna et al., 2014; Schooler & Engstler-Schooler, 1990).

Participants in Experiment 1b also completed a task assessing their pragmatic inferences about the verbs used in L&P's studies. Responses on this task revealed participants' assumptions about the severity of an accident described with *smashed* compared to *hit*. If participants rely on the implied severity of the verb in the speed question to estimate the speed of the cars in the video, those who interpret *smashed* and *hit* as implying a relatively *large* difference in severity should exhibit a larger framing effect. Alternatively, for these participants, *smashed* may seem *too* severe for the target accident (or *hit* not severe enough). If so, they may discount the verb frame when making their speed estimate, since it does not seem consistent with their visual memory. As a result, participants who interpret *smashed* and *hit* as implying a relatively *small* difference in severity may be the ones to show a larger framing effect. For these participants, either verb would be an apt descriptor of the accident, so they may interpret the particular verb chosen as communicating relevant information about the accident, and thus rely on it when making their speed estimate.

Method

Participants Using Prolific, we recruited 401 participants in Experiment 1a and 400 participants in Experiment 1b (430 male, 354 female, 17 nonbinary/other; M_{age} = 44 years, range = 19–83). All were native English speakers located in the U.S., had >95% approval on at least 100 prior studies, and passed an attention check at the beginning of the study. In each experiment, we aimed for 100 participants in each cell of the design, similar to other contemporary linguistic framing studies (Flusberg et al., in press).

Design and Procedure Participants in both Experiment 1a and 1b were randomly assigned to either the verbal description condition or the control condition, and to receive the *smashed* or *hit* version of the speed question. Both experiments were administered online via Qualtrics.

Experiment 1a First, participants viewed a 9-second video clip of a two-car accident (see our OSF page). On the next screen, participants in the verbal description condition were asked to write an account of the accident in their own words. Participants in the control condition were asked to list as many countries and their capitals as they could from memory, as in studies of verbal overshadowing (e.g., Alogna et al., 2014). Both groups were given 2 minutes to complete their task and encouraged to use the full time.

On subsequent screens, all participants answered a series of specific questions about the accident. The first was the speed question: "*About how fast were the cars going when they [smashed into/hit] each other?*" Participants used a 0–100 slider to log their estimate in miles per hour. On the remaining questions, participants used an 11-point scale to rate—separately for each car—how seriously the driver was injured, how severe the damage was, how much smoke was produced, and how shaken the driver felt. Participants were also asked to estimate the cost of repairs for each car and how high off the ground one of the cars went during the accident.

Finally, participants answered a series of demographic questions, as well as questions about their driving experience and familiarity with the L&P study. These questions were included for exploratory purposes; due to space constraints, the results are not reported here.

Experiment 1b The procedure was identical to Experiment 1a with four exceptions. First, participants were given 3 minutes (instead of 2) to complete the verbal description or control task. Second, this task was immediately preceded by a series of filler tasks: the demographic and driving experience questions from Experiment 1a, as well as the 36-item Internal Representations Questionnaire (e.g., "My mental images are very vivid and photographic"; 5-point agreement scale; Roebuck & Lupyan, 2020).

Third, after answering the series of specific questions about the accident in the video, participants completed a task assessing their pragmatic inferences about the verbs of interest. They read, "*Suppose a friend described a different car accident by saying, 'The cars _____ each other.'*" For each of the 5 verbs in L&P's studies that could be chosen to fill in the blank ("smashed into," "collided with," "bumped into," "hit," "contacted"), participants were asked to indicate what the verb implied about how fast the cars were going (using a 0–100 mph slider), how severe the accident was, how reckless the drivers were, and how serious the drivers' injuries were (using 11-point scales). For each of these questions, the 5 verbs appeared in a randomized order.

Finally, after this pragmatics task, participants responded to a recognition memory prompt asking them to select which verb had been presented in the question about the speed of the cars in the video. Responses to this exploratory question did not moderate the main results, and are not reported here.

Results

Speed Estimates For both Experiment 1a and 1b, we analyzed participants' speed estimates for the cars in the video using a 2 (condition) × 2 (verb frame: *smashed* vs. *hit*) ANOVA. In both experiments, neither main effect was significant (condition: Fs < 2.9, ps > .09; $\eta_p^2 < .008$; verb frame: Fs < 2.6, ps > .11; $\eta_p^2 < .007$), nor was the interaction (Fs < 0.7, ps > .4; $\eta_p^2 < .003$). As shown in Figure 1, L&P's framing effect (*smashed* > *hit*) was not replicated overall in either experiment.

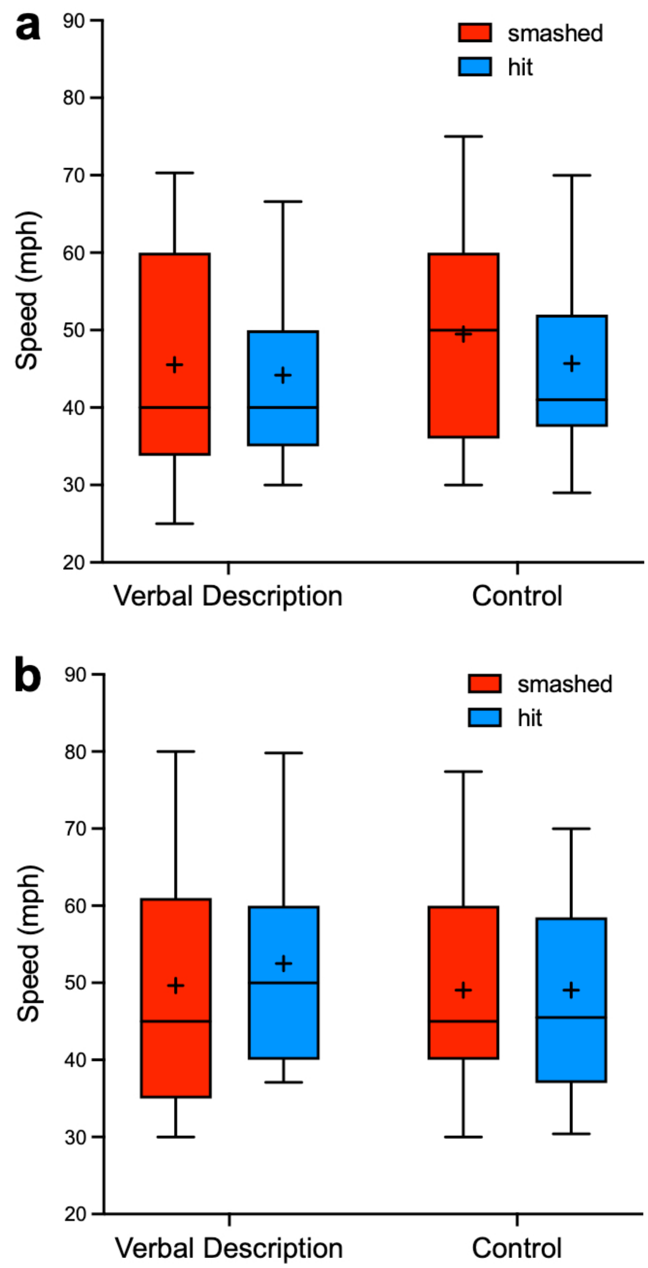

Figure 1: Speed estimates in (a) Experiment 1a and (b) Experiment 1b. Boxes denote the interquartile range (whiskers = 10^{th}–90^{th} percentile, middle line = median, + sign = mean).

Other Questions We conducted corresponding analyses on responses to the other specific questions about the accident in the video. The 8 rating questions had good reliability (Cronbach's α > .77 in both Experiment 1a and 1b), so each participant's ratings were averaged to form a composite rating of *accident severity*. ANOVAs on these composite ratings and on estimates of repair cost and car height yielded no significant effects in either experiment (ps > .05).

Verbal Elaboration In an exploratory analysis, we assessed whether the effect of the verb frame was moderated by the extent to which participants in the verbal description condition elaborated on the accident in writing. As a rough index of elaboration, we computed the length of participants' descriptions across Experiment 1a and 1b (M = 56 words, SD = 32). A linear regression model with description length, verb frame (1, *smashed*; 0, *hit*), and their interaction as predictors of participants' speed estimates yielded a significant interaction, b = -0.11 [95% CI: -0.22 to -0.004], p = .04. The longer participants' descriptions were, the less likely they were to exhibit the expected framing effect (see Figure 2).

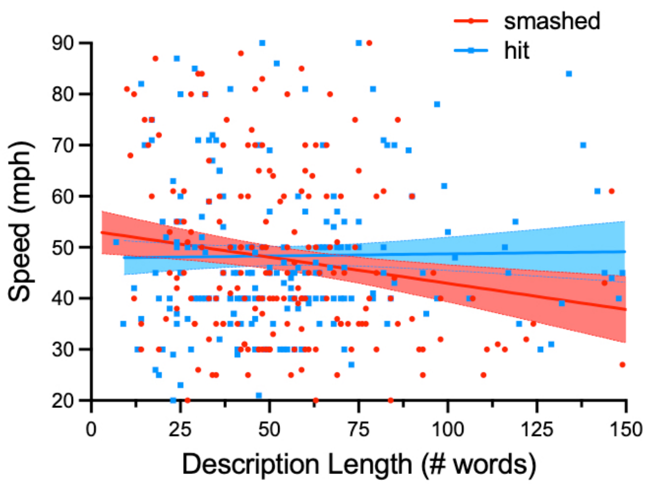

Figure 2: Relationship between the length of participants' verbal descriptions and their speed estimates for each verb frame in Experiment 1. Bands denote 95% CIs.

Pragmatic Inference In Experiment 1b, the 15 rating questions on the pragmatics task (3 per verb) had good reliability (Cronbach's α = .91). Each participant's ratings were averaged to form a composite rating of *implied* severity for each verb. As shown in Table 1, participants differentiated systematically among the verbs in both implied speed and implied severity. An exploratory repeated-measures ANOVA showed that both sets of responses differed across verbs (ps < .001). Holm-Bonferroni post hoc tests indicated that each of the 5 verbs differed significantly from all the others (ps < .001), except *contacted* vs. *bumped* (ps > .07).

Table 1: Rank-ordered mean implied speed (mph) and implied severity (1-to-11 scale) for each verb on the pragmatics task in Experiments 1b and 2.

	Experiment 1b		Experiment 2	
Verb	**Speed**	**Severity**	**Speed**	**Severity**
smashed	60.9	9.5	59.1	9.2
crashed	–	–	56.0	8.7
collided	49.8	7.9	47.5	7.1
ran	–	–	40.2	6.3
hit	44.7	6.8	39.4	6.2
contacted	20.7	3.8	17.6	3.3
bumped	18.7	3.6	15.4	3.2

To assess whether the effect of the verb frame was moderated by participants' pragmatic inferences about the verbs in the speed question (*smashed* and *hit*), we used an individual-level index of these inferences. For each participant in Experiment 1b, we computed dSeverity: the difference in their composite ratings of implied severity for *smashed* and *hit* (higher values = larger difference in implied severity between the two verbs). We preregistered this measure because we expected that it would yield fewer extreme values than difference scores based on implied speeds. In an exploratory linear regression analysis predicting participants' speed estimates, we entered dSeverity, verb frame, and their interaction as predictors. The interaction did not reach significance, b = -1.40 [95% CI: -3.01–0.21], p = .09, though participants who judged *smashed* as only somewhat more severe than *hit* (smaller dSeverity) were descriptively more likely to exhibit the expected framing effect than those who judged *smashed* as far more severe than *hit* (larger dSeverity).

Discussion

Consistent with other attempts to replicate L&P (McAllister et al., 1988; Raghunath et al., 2021; Read et al., 1978), we observed no overall effect of verb framing on speed estimates (or other specific questions) for cars involved in an accident. While other replications failed to include the verbal description task from the original L&P studies, we found that this task did not moderate the effect of the verb frame.

However, our results provide clues that other moderating factors identified in the linguistic framing literature may help explain the apparently fragile nature of this effect. Specifically, exploratory analyses revealed that the more participants elaborated on their written description of the accident, the less likely they were to exhibit the expected framing effect. This suggests that people who develop a richer memory representation of an event are more impervious to subsequent suggestive framing. Additionally, we found no evidence that pragmatic inferences about the intensity of *smashed* and *hit* moderated the framing effect. However, our pragmatics task revealed that these two verbs might not provide the strongest test of the framing effect or potential moderators. While *smashed* was rated high in implied speed and severity, *hit* was rated intermediate relative to the other verbs in L&P's studies (see Table 1). A starker verb contrast might maximize the potential for verb framing (and pragmatic inferences about the verbs) to have an impact. We designed Experiment 2 to address this possibility and replicate the exploratory findings from Experiment 1.

Experiment 2

Given that the verbal description task did not moderate the framing effect in Experiment 1a or 1b, we dropped this manipulation in Experiment 2. As in L&P's studies, all participants wrote a verbal description before making their speed estimates. The speed question contrasted *smashed* and *contacted*, which differed markedly in intensity on the pragmatics task of Experiment 1b.

We also varied the position of the speed question within the series of specific questions about the accident. L&P did not report the order of their questions, nor what the other questions were. If some of these questions came before the speed question, this may have directed participants' attention to verbalizable aspects of the accident and made them more susceptible to subsequent verb framing. To examine this possibility, we presented the speed question either first in the series of specific questions (as in Experiment 1) or last.

Method

Participants We recruited 403 U.S. participants on Prolific using the same criteria as in Experiment 1 (165 male, 226 female, 12 nonbinary/other; M_{age} = 43 years, range = 18–79).

Design and Procedure As in Experiment 1, participants were randomly assigned to receive one of two versions of the speed question, this time with either *smashed* or *contacted*.

All participants watched the video of the accident and then wrote a description of it for 3 minutes. Then they answered the series of specific questions about the accident (with the speed question appearing either first or last), completed the pragmatics task and memory prompt from Experiment 1b, and reported their demographics, driving experience, and familiarity with the L&P study. The pragmatics task assessed participants' inferences about "crashed into" and "ran into" (which were among the verbs most frequently generated in participants' verbal descriptions in Experiment 1) in addition to the 5 verbs presented in Experiment 1b. All other aspects of the procedure were identical to Experiment 1b.

Results

Speed Estimates A 2 (verb frame: *smashed* vs. *contacted*) × 2 (question order: speed question first vs. last) ANOVA on speed estimates yielded no main effects (verb frame: $F(1, 399) = 1.03$, $p = .31$, $\eta_p^2 = .003$; question order: $F(1, 399) = 0.31$, $p = .58$, $\eta_p^2 < .001$) and no interaction, $F(1, 399) = 2.79$, $p = .10$, $\eta_p^2 = .007$. As shown in Figure 3, the expected framing effect (*smashed* > *contacted*) was not observed, regardless of whether the speed question appeared first or last in the series of specific questions.

Other Questions Corresponding ANOVAs on responses to the other specific questions yielded a main effect of question order on composite accident severity ratings, $F(1, 399) = 13.39$, $p < .001$, $\eta_p^2 = .03$, and estimates of repair cost for the second car seen in the video, $F(1, 375) = 4.02$, $p = .05$, $\eta_p^2 = .01$. Ratings and estimates were higher when these questions appeared before the speed question, which may have anchored participants on relatively low values. No other significant effects were observed (*ps* > .05).

Verbal Elaboration As in Experiment 1, we explored whether description length (M = 66 words, SD = 32) moderated the effect of the verb frame. A linear regression model with description length, verb frame (1, *smashed*; 0, *contacted*), and their interaction as predictors yielded a significant interaction, $b = -0.12$ [95% CI: -0.21 to -0.02], $p = .02$. Participants who wrote shorter descriptions were more likely to exhibit the expected framing effect, replicating Experiment 1 (see Figure 4).

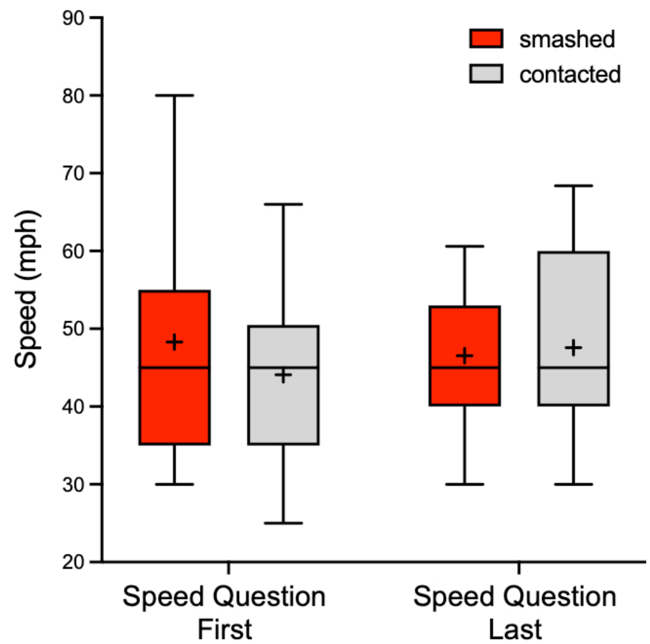

Figure 3: Speed estimates in Experiment 2. Boxes denote the interquartile range (whiskers = 10th–90th percentile, middle line = median, + sign = mean).

Pragmatic Inference Once again, participants differentiated systematically among the verbs on the pragmatics task (see Table 1). An exploratory repeated-measures ANOVA showed that both implied speed and implied severity differed across verbs (*ps* < .001). Holm-Bonferroni post hoc tests

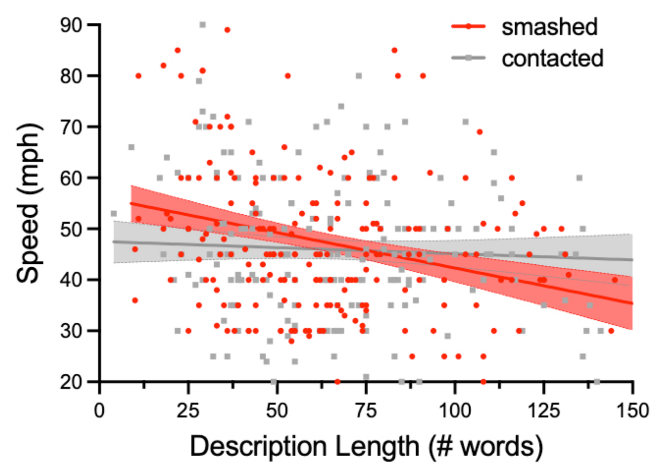

Figure 4: Relationship between the length of participants' verbal descriptions and their speed estimates for each verb frame in Experiment 2. Bands denote 95% CIs.

indicated that each of the 7 verbs differed significantly from all the others (*ps* < .001), except *ran* vs. *hit* and *contacted* vs. *bumped* (*ps* > .05).

As in Experiment 1b, we explored whether participants' pragmatic inferences moderated the effect of the verb frame. For each participant, we computed dSeverity for their responses on the pragmatics task (*smashed* minus *contacted*). A linear regression model with dSeverity, verb frame, and their interaction as predictors of speed estimates yielded a significant interaction, b = -1.38 [95% CI: -2.56 to -0.20], p = .02. Consistent with the trend observed in Experiment 1b, participants who judged *smashed* as only somewhat more severe than *contacted* (smaller dSeverity) were more likely to exhibit the expected framing effect than those who judged *smashed* as far more severe (larger dSeverity; see Figure 5).

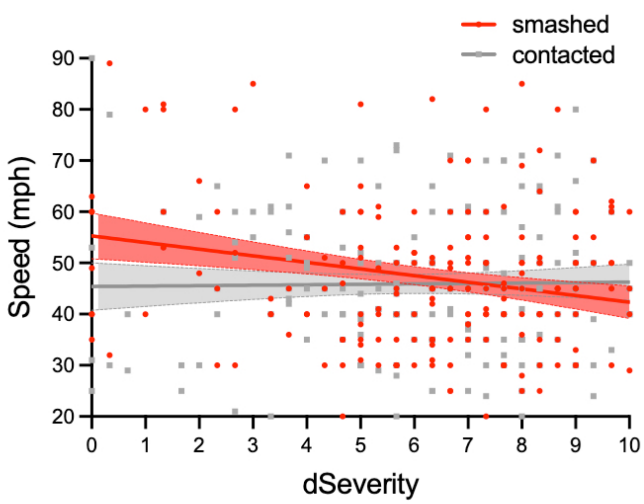

Figure 5: Relationship between dSeverity ($\text{severity}_{smashed}$ – $\text{severity}_{contacted}$ on the pragmatics task; 11 point-scale) and speed estimates for each verb frame in Experiment 2. Bands denote 95% CIs.

Discussion

We again found no overall framing effect on speed estimates even though *smashed* and *contacted* were chosen to be maximally different in intensity. We also found no evidence that presenting the speed question before or after the other specific questions moderated the framing effect.

However, Experiment 2 replicated our other observations from Experiment 1. Participants who (a) elaborated *less* on their description of the accident and (b) inferred a *smaller* difference in severity between the two verb frames were *more* likely to show the expected framing effect. The latter effect was statistically significant in Experiment 2, reinforcing the importance of using participant norming data to select linguistic frames rather than relying on researcher intuitions (cf. Thibodeau & Boroditsky, 2015).

General Discussion

Scientists, lawyers, law enforcement, and the public have a longstanding interest in the effects of linguistic framing on eyewitness memory. A classic, oft-cited finding by L&P has proven difficult to replicate, however. Based on a key difference between the L&P study and previous replication attempts, as well as recent work on linguistic framing, we hypothesized that verbal elaboration and pragmatic inference might help explain this discrepancy.

In two experiments, participants viewed a brief car accident video. They either wrote a verbal description of the event or completed an unrelated task before answering the verb-framed question eliciting their speed estimate. We observed no main effect of the verb frame, and the act of writing a description of the accident—which was present in L&P but omitted in all other replication attempts—was not a significant moderator. However, participants who wrote less elaborate descriptions and inferred a smaller difference in intensity between the two verb frames were more likely to show the expected framing effect. These results suggest that verb framing may have a greater impact on recollections when people have a relatively weak representation of an event and when the framing language is subtle compared to alternative wording, or at least believed to be so. This is consistent with other work showing that representation strength and pragmatic inferences about the linguistic frame moderate a range of framing effects (Flusberg et al., in press). That said, some of our key findings were derived from exploratory analyses and thus would benefit from corroboration in future work.

There are also several differences between our experiments and L&P's that are worth bearing in mind. First, we used only a single accident video, as opposed to the seven different clips used in their first study. However, L&P also used a single clip in their second study and assumed their findings would generalize to a wide range of accidents of different speeds and configurations. Second, our study was conducted online as opposed to in the lab, potentially increasing variability in speed estimates and making it more difficult to detect the effect of the verb frame. Third, a lot has changed since the 1970s, especially in access to media. It is likely that our participants had viewed many more real and simulated car accidents in movies, video games, and online—all in high definition—compared to L&P's participants. The single clip we showed them may have seemed mundane by comparison, perhaps tempering their reactions to the accident and to our framing language. Additional studies are needed to address these issues.

In sum, however, our findings may help resolve an ongoing controversy in the framing literature that might appear to undermine prevailing accounts of the fallibility of human memory. They also illuminate the importance of assessing factors that moderate the impact of framing. Most behavioral effects are heterogeneous across contexts and experimental manipulations (Bryan et al., 2021), including the effects of other kinds of suggestive language on memory (O'Donnell & Chan, 2023; Read & Bruce, 1984). In light of this heterogeneity, failed replications in the framing literature serve as opportunities to understand when and how—not just whether—language shapes the way we think and remember.

Acknowledgments

This work was supported by a James McKeen Cattell Sabbatical Award to KJH.

References

Alogna, V. K., Attaya, M. K., Aucoin, P., Bahnik, Š., Birch, S., Birt, A. R., . . . Zwaan, R. A. (2014). Registered replication report: Schooler & Engstler-Schooler (1990). *Perspectives on Psychological Science*, *9*(5), 556–578. https://doi.org/10.1177/1745691614545653

Brainerd, C. J., & Reyna, V. F. (2005). *The science of false memory*. Oxford University Press.

Bryan, C. J., Tipton, E., & Yeager, D. S. (2021). Behavioural science is unlikely to change the world without a heterogeneity revolution. *Nature Human Behaviour*, *5*, 980–989. https://doi.org/10.1038/s41562-021-01143-3

Chestnut, E. K., & Markman, E. M. (2018). "Girls are as good as boys at math" implies that boys are probably better: A study of expressions of gender equality. *Cognitive Science*, *42*(7), 2229–2249. https://doi.org/10.1111/cogs.12637

Flusberg, S. J., Holmes, K. J., Thibodeau, P. H., Nabi, R. L., & Matlock, T. (in press). The psychology of framing: How everyday language shapes the way we think, feel, and act. *Psychological Science in the Public Interest*.

Flusberg, S. J., Thibodeau, P. H., & Holmes, K. J. (2022). Even simple framing effects are rational: Commentary on Bermúdez. *Behavioral & Brain Sciences*, *45*, e228. https://doi.org/10.1017/S0140525X22000942

Frenda, S. J., Nichols, R. M., & Loftus, E. F. (2011). Current issues and advances in misinformation research. *Current Directions in Psychological Science*, *20*(1), 20–23. https://doi.org/10.1177/0963721410396620

Goldschmied, N., Sheptock, M., Kim, K., & Galily, Y. (2017). Appraising Loftus and Palmer (2017) post-event information versus concurrent commentary in the context of sport. *Quarterly Journal of Experimental Psychology*, *70*(11), 2347–2356. https://doi.org/10.1080/17470218.2016.1237980

Holmes, K. J., Doherty, E. M., & Flusberg, S. J. (2022). How and when does syntax perpetuate stereotypes? Probing the framing effects of subject-complement statements of equality. *Thinking & Reasoning*, *28*(2), 226–260. https://doi.org/10.1080/13546783.2021.1963841

Holmes, K. J., Wu, S. H., Elpers, N., Doherty, E., & Flusberg, S. J. (under review). How syntax promotes stereotypes: Assessing the role of pragmatic inference.

Klatzky, R. L., Martin, G. L., & Kane, R. A. (1982). Semantic interpretation effects on memory for faces. *Memory & Cognition*, *10*(3), 195–206. https://doi.org/10.3758/BF03197630

Leong, L. M., McKenzie, C. R. M., Sher, S., & Müller-Trede, J. (2017). The role of inference in attribute framing effects. *Journal of Behavioral Decision Making*, *30*(5), 1147–1156. https://doi.org/10.1002/bdm.2030

Lipscomb, T. J., Bregman, N. J., & McAllister, H. A. (1985). A developmental inquiry into the effects of postevent information on eyewitness accounts. *Journal of Genetic Psychology*, *146*(4), 551–556. https://doi.org/10.1080/00221325.1985.10532475

Loftus, E. F. (1975). Leading questions and the eyewitness report. *Cognitive Psychology*, *7*(4), 560–572. https://doi.org/10.1016/0010-0285(75)90023-7

Loftus, E. F., & Palmer, J. C. (1974). Reconstruction of automobile destruction: An example of the interaction between language and memory. *Journal of Verbal Learning and Verbal Behavior*, *13*(5), 585–589. https://doi.org/10.1016/S0022-5371(74)80011-3

Loftus, E. F., & Zanni, G. (1975). Eyewitness testimony: The influence of the wording of a question. *Bulletin of the Psychonomic Society*, *5*(1), 86–88. https://doi.org/10.3758/BF03336715

McAllister, H. A., Bregman, N. J., & Lipscomb, T. J. (1988). Speed estimates by eyewitnesses and earwitnesses: How vulnerable to postevent information? *Journal of Genetic Psychology*, *115*(1), 25–35. https://doi.org/10.1080/00221309.1988.9711085

O'Donnell, R., & Chan, J. C. K. (2023). Does blatantly contradictory information reduce the misinformation effect? A Registered Report replication of Loftus (1979). *Legal and Criminological Psychology*. https://doi.org/10.1111/lcrp.12242

Raghunath, N., Fultz, A. A., & Sanchez, C. A. (2021). Car crash speed estimates not biased by leading questions. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *65*(1), 848–853. https://doi.org/10.1177/1071181321651302

Read, J. D., Barnsley, R. H., Ankers, K., & Whishaw, I. Q. (1978). Variations in severity of verbs and eyewitnesses' testimony: An alternative interpretation. *Perceptual and Motor Skills*, *46*(3), 795–800. https://doi.org/10.2466/pms.1978.46.3.795

Read, J. D., & Bruce, D. (1984). On the external validity of questioning effects in eyewitness testimony. *International Review of Applied Psychology*, *33*(1), 33–49. https://doi.org/10.1111/j.1464-0597.1984.tb01415.x

Roebuck, H., & Lupyan, G. (2020). The Internal Representations Questionnaire: Measuring modes of thinking. *Behavior Research Methods*, *52*, 2053–2070. https://doi.org/10.3758/s13428-020-01354-y

Schooler, J. W., & Engstler-Schooler, T. Y. (1990). Verbal overshadowing of visual memories: Some things are better left unsaid. *Cognitive Psychology*, *22*(1), 36–71. https://doi.org/10.1016/0010-0285(90)90003-M

Sher, S., & McKenzie, C. R. M. (2006). Information leakage from logically equivalent frames. *Cognition*, *101*(3), 467–494. https://doi.org/10.1016/j.cognition.2005.11.001

Simons, D. J., & Chabris, C. F. (2011). What people believe about how memory works: A representative survey of the U.S. population. *PLoS One*, *6*(8), e22757. https://doi.org/10.1371/journal.pone.0022757

Thibodeau, P. H., & Boroditsky, L. (2015). Measuring effects of metaphor in a dynamic opinion landscape. *PLoS One*, *10*(7), e0133939. https://doi.org/10.1371/journal.pone.0133939