

The verbatim access effect: implicature in experimental context

MUFFY SIEGEL

JÉRÉMY ZEHR

HEZEKIAH AKIVA BACOVGIN

University of Pennsylvania

LYNNE STEUERLE SCHOFIELD

Swarthmore College

AND

FLORIAN SCHWARZ

University of Pennsylvania

(Received 07 June 2018 – Revised 14 September 2018 – Accepted 08 October 2018)

ABSTRACT

Implicature interpretation is sensitive to many contextual factors. This experimental study investigates two:

- (A) instructions to think carefully about exactly what is said
- (B) access to the verbatim form of what has been said

Participants encountered (1) below, which can give rise to the contradictory relevance implicature in (2), as feedback during a decoy task:

- (1) I'm not suggesting that you're responding too slowly, but it's important to give the first response that comes to mind.
- (2) (I am suggesting that) you're responding too slowly.

[*] We gratefully acknowledge NSF grant BCS-1349009 to Florian Schwarz for support of this research. We are also grateful for helpful discussions with Joe Bowering, Alice Hausman, Jeff Kaplan, Mandy Simons, Rob Wilder, and the participants in Florian Schwarz's lab seminar. We thank *Meaning in Flux* (2016) audience members, especially Lyn Frazier and Adele Goldberg, for their incisive questions and invaluable recommendations. Any remaining errors are our own. Address for correspondence: Muffy Siegel, Department of Linguistics, University of Pennsylvania, 3401-C Walnut St., suite 300, Philadelphia, PA 19104-6228. e-mail: muffy.siegel@temple.edu

When participants were questioned post-task, (B) significantly reduced rates of agreement that the speaker of (1) had said (2), whether the verbatim form provided was written (Experiment 1) or audio (Experiment 2). (A) had no such effect. In Experiment 3, we added a final task for participants: to recall (1) verbatim. One-third had forgotten it, typically substituting the implicature (2). We argue that this memory loss can explain the lower implicature rates associated with verbatim access: verbatim access reminds forgetful participants of (1)'s compositional interpretation, and that interpretation is inconsistent with the implicature in (2). Consequently, verbatim access reduces the chances of endorsing (2), thus introducing an inherent literal meaning bias in interpreting previous conversation.

KEYWORDS: implicature, pragmatics, semantics, context, verbatim memory.

1. Introduction

Under what circumstances do we understand (1) as insulting the speaker's boss?

(1) I'm not saying my boss is stupid, but ...

Intuitively, how we interpret (1) depends largely on what we know about the speaker and the context of the utterance. If the speaker is an earnest employee engaged in a serious discussion about a difficult work situation, we would probably take (1) literally: the speaker is not saying that his boss is stupid. If, on the other hand, the speaker is an obviously disgruntled employee, or he is a stand-up comedian performing a routine, we would probably believe that he does mean to insult his boss by saying that she is stupid. That is, in natural conversational settings, many extralinguistic factors influence how listeners interpret sentences like (1). Since we aspire to account in a principled way for how listeners interpret what they hear, it is important, for theoretical reasons, to learn about these influential contextual factors. Moreover, what we learn about such factors will have practical applications as well: in experimental work, and even in legal proceedings, it is critical to know if certain contextual factors are likely to disproportionately promote particular interpretations. Therefore, in the present study, we expose experimental participants to a conversational utterance like (1) and record, a few minutes later, their preferred interpretations of it in the presence and absence of two contextual factors that may affect those preferred interpretations: instructions to think carefully about the utterance and access to its verbatim form.

1.1. CONVERSATIONAL IMPLICATURE

Utterances like (1) are unusual in having two, equally accessible interpretations which contradict each other. The first, literal meaning ('I'm not saying that my boss is stupid.') is the product of ordinary compositional semantic interpretation, the putting together of word meanings according to the structure of the sentence. In contrast, the second meaning ('I am saying that my boss is stupid') is a *CONVERSATIONAL IMPLICATURE*, an additional or alternate meaning derived from what the speaker has said, on the basis of conversational principles first defined by Grice (1975). These principles comprise Grice's overarching Universal Cooperative Principle: speakers and hearers attempt to cooperate with each other in order to help all parties achieve their conversational goals. More specifically, Grice (1975, pp. 45–46) identified four conversational sub-principles, or maxims: listeners assume that speakers try to make their contributions true (Maxim of Quality), that speakers include as much information as is required for the exchange, but no more (Maxim of Quantity), that speakers' contributions are related to the message they want to convey (Maxim of Relevance/Relation), and that those contributions are worded clearly and concisely (Maxim of Manner).

Subsequent to Grice's groundbreaking work, several scholars have suggested recombining, reorganizing, or recasting these four Gricean conversational maxims (Horn, 1984; Levinson, 2000; Sperber & Wilson, 1995). Even in these revised formats, the maxims maintain similar explanatory power for our purposes. That is, when speakers appear to be in danger of violating any of these maxims, cooperative listeners can be relied upon to make sense of the possible breach in conversational rules by inferring (usually unconsciously) that the speaker is also being cooperative: he merely intends to convey a conversational implicature in addition to or instead of the literal meaning. It is up to the listener to use her knowledge of the context of the conversation, the language being spoken, and Grice's maxims to compute that implicature. Grice (1975) described conversational implicatures that depend only upon specific conversational contexts as *PARTICULARIZED*, and those that spring from agreed-upon word meanings as *GENERALIZED*.

Consider, for instance, (2) and (3) below:

- (2) My love is a rose
- (3) I ate some of the cookies.

If we take (2) literally, it appears that the speaker has violated the Maxim of Quality; there are very few contexts in which it would be true that a person is in love with a flower. However, a cooperative listener is unlikely to give up on the conversation immediately, having concluded that the speaker is lying, uncooperative, or simply deluded. Instead, she is much more likely to assume

that the speaker, too, is being cooperative; that is, that he is trying to convey something sensible by purposely disobeying, or flouting, the Maxim of Quality. In this case, he seems to want the listener to infer that his lover is a person who is like a rose in some ways (left to the listener to imagine). Such a metaphorical meaning can be analyzed as a (particularized) conversational implicature, because the listener derives it from (2) on the basis of her knowledge that (2) is false in the conversational context and that a cooperative speaker would not so obviously disobey the Maxim of Quality unless he meant to signal that a conversational implicature was to be derived.

A different kind of conversational implicature, a *SCALAR IMPLICATURE*, can be derived from (3) on the basis of the meaning of *some* and the Maxim of Quantity. The literal meaning of (3) ('I ate some of the cookies.') is consistent with the speaker having eaten either all of the cookies in question or just a subset of those cookies. (If you have eaten all the cookies, it is still literally true that you ate some of the cookies.) Yet (3) is usually understood to have the second, (generalized) conversational implicature meaning, that the speaker has eaten some *BUT NOT ALL* of the cookies. This comes about because listeners assume that the speaker of (3) is being cooperative and trying to adhere to the maxims, especially Quantity, in giving as much reliable, relevant information as she can about the fate of the cookies. If she had eaten all the cookies, she would have known that, and, consequently, as a cooperative speaker, she would have felt obliged by the Maxim of Quantity to provide that information, since it would probably be useful to the listener. Since she said, instead, that she ate some of the cookies, listeners infer that the speaker did not eat them all, because she would have shared that fact if she had known it to be true. Hence, *some* is typically taken to conversationally implicate 'some but not all', although its literal meaning is 'some and possibly all'.

Thus, the meanings we take from quite ordinary utterances like (2) and (3) are often not the literal ones associated with their semantics, but conversational implicatures, which listeners compute from semantic interpretations on the basis of pragmatic factors associated with the participants, the context, and our shared rules of conversation. Returning to our example in (1), we can now trace the generation of its conversational implicature meaning as we did those of (2) and (3). First, since the Cooperative Principle and the maxims are universal, those who hear (1) will assume that the person who has said it is trying his best to adhere to all the maxims. Thus, listeners will assume, by virtue of the Maxim of Relevance, that whatever the speaker has said in (1) is relevant to what he wants to convey in this particular conversational context. In that case, though, why would the speaker even mention the inflammatory proposition that his boss is stupid as something he is *NOT* saying? We do not normally go about announcing things that we are not saying, especially not provocative, but irrelevant things. A likely explanation is that the speaker of

(1) actually wants to be taken as saying that his boss is stupid, even though he appears to be saying literally the opposite. Thus, (1) allows both a literal interpretation, that the speaker is not saying his boss is stupid, and a (particularized) conversational implicature interpretation, that he is saying that the boss is stupid.

1.2. CONTEXTUAL FACTORS

Much previous research has shown that the inclination to embrace a conversational implicature, rather than the literal interpretation, as the meaning of an utterance depends upon a variety of specific contextual factors. These factors include beliefs about the communicative situation (Clark, 1979; Grodner & Sedivy, 2011; Siegel, 2005), beliefs about the interlocutors (Sikos, Kim, Anchirai, Lam, & Grodner, 2016), lexical variation in the utterance itself (Clark, 1979; van Tiel, van Miltenburg, Zevakhin, & Geurts, 2014), lexical variation in the immediate linguistic context (Degen, 2015; Doran, Ward, Larson, McNabb, & Baker, 2012; Feeney, Scafton, Duckworth, & Handley, 2004), age-related communicative preferences (Katsos & Bishop, 2011; Bill, Romoli, Schwarz, & Crain, 2016), and even politeness considerations (Bonnefon, Feeney, & Villejoubert, 2009; Mazzarella, 2015). Experimental investigation of such contextual effects on rates of implicature interpretation is challenging, in part because implicature rates are typically depressed in contexts that do not faithfully mimic natural conversation (Guasti, Chierchia, Crane, Foppolo, Gualmini, & Meroni, 2005; Papafragou & Musolino, 2003; Papafragou & Tantalou, 2004). For instance, Smith (1980) and Noveck (2001) found, in studies employing artificial laboratory tasks, that young children did not readily generate implicatures. However, when other researchers had companionable puppets address their target utterances to children in natural, conversational settings, five- and seven-year-olds regularly gave them implicature interpretations (Guasti et al., 2005; Papafragou and Musolino, 2003). In view of such results, a naturalistic conversational setting is regarded as a valued design feature of experiments meant to investigate natural language behavior (Gurevich, Johnson, & Goldberg, 2010).

Still, even when the experimental task involves participants in a reasonably natural interaction, other features peculiar to the experimental context can affect the likelihood of implicature interpretation. This paper investigates two such common, but potentially confounding, features. First, instructions to participants are almost universal in experiments, but not present in typical conversations, and these instructions can have significant effects. Doran et al. (2012) found that pre-trained participants instructed to interpret material with potential implicature readings “literally” exhibited significantly lower

implicature rates than a baseline group with neutral instructions. Moreover, others, who were instructed to interpret the material as if they were *Literal Lucy*, a fictional character who takes everything literally, exhibited even lower rates. Our Experiment 1 briefly investigates the effects on implicature rates of a more natural version of Doran et al.'s instructions. It addresses the question of whether participants with no special training who are instructed merely to "pay careful attention ... think carefully ... [and] answer questions about exactly what was said" will be less likely to endorse implicature readings than a baseline group who receive no such instructions.

However, the central question addressed in this paper concerns a different, previously unstudied element of many experimental designs: participant access to the verbatim form of a conversational target utterance at a time of interpretation just a few minutes after the target is uttered. That is, we investigate, in three experiments, whether having unlimited access to a written transcript or audio-recording of a conversational contribution affects how likely speakers are to accept, post-utterance, a conversational implicature as its meaning. Of course, ordinary conversation rarely comes with a transcript or recording of what our interlocutors have said, but it is very common for experiments in semantics/pragmatics to allow participants access to the verbatim form of targets as they record their judgments. Moreover, courts also present conversational evidence in transcripts and/or recordings for the judge and jury to pore over, and it has been suggested that this could result in an unrealistically low rate of implicature interpretation which could compromise the judicial process (Prince, 1990; Siegel, 2005). Thus it is pertinent both to experimental pragmatics and to the legal system to investigate whether post-stimulus verbatim access affects implicature rates and, if so, why.

1.3. DESIGN CONSIDERATIONS

To collect realistic judgments bearing on the effect of instructions and verbatim access on implicature rates, we chose a target utterance on the model of (1), with equally plausible and easily distinguishable literal and implicature interpretations, and presented it in a setting that allowed participants to feel, as much as possible, that they were involved in a natural conversation with a cooperative interlocutor. The first clause of (4) below is our target utterance, and (5) is its possible implicature. (The second clause of (4) was included in order to enhance the naturalness of our target. It was chosen to be as semantically neutral as possible and consistent with either reading of the first clause, but neither entailed by nor entailing either first-clause reading. We leave for future research a full investigation of the role of such exception clauses.)

- (4) I'm not suggesting that you're responding too slowly, but it's important to give the first answer that comes to mind.
 (5) (The speaker is suggesting that) you're responding too slowly.

As in (1), the implicature in (5) arises from the first clause of (4) via Grice's Maxim of Relevance: "Be Relevant" (Grice, 1975, p. 46). An addressee may endorse (5) as what was said in (4) because he grasps intuitively that a cooperative speaker would not mention the proposition that the listener is responding too slowly if it were not relevant to what she was trying to communicate.

Our choice to work with a relevance implicature like (5) is unusual; most prior experimental research on implicatures has employed scalar ones, as in (3). However, using a relevance implicature not only broadens the range of implicature types under study, but also avoids several difficulties that scalar implicatures introduce. Consider (6) and its possible scalar implicature in (7), from Bott and Noveck (2004):

- (6) Some elephants are mammals.
 (7) Some, but not all, elephants are mammals.

First, it is difficult to tell whether someone is interpreting (6) as (7) because their truth conditions largely overlap. *Some, but not all, elephants are mammals* entails that some elephants are mammals. In contrast, the literal and implicature meanings of the first clause of (4) are easily distinguishable because they are inconsistent. Second, scalar implicatures are sensitive to the contextual availability of other items on the pertinent scale (Barner, Brooks, & Bale, 2011; Degen & Tanenhaus, 2015; Grodner, Kim, & Russell, 2016), the distinctness of these scalemates (van Tiel et al., 2014), and the addressee's perception of the speaker's knowledge of such factors (Sauerland, 2004). Relevance implicatures have no such sensitivities. Third, under-informative scalar examples like (6) are difficult to place in natural conversations for experimental participants. Testing them typically requires the creation of somewhat artificial tasks which elicit participants' judgments of truth or 'correctness'. Such truth judgments indicate the participants' choice of literal or implicature interpretation because under-informative scalar generalizations like (6) are true on only their literal readings. (This asymmetry might also bias participants toward those literal readings.) In contrast, in our experiments, we easily constructed a natural conversational context in which the literal reading of the first clause of (4) ('The speaker is not suggesting *p*') and the contradictory implicature meaning in (5) ('The speaker is suggesting *p*') are clearly distinguishable from each other, yet about equally natural and equally likely to be true.

In order to create this naturalistic conversational context in which to present our target utterance (4), we led participants to believe that our study

involved only a lexical decision task. At the start of the experiment, a friendly, personable audio guide who introduces herself as Sarah gives the participants conversational-sounding instructions for doing the lexical decision task: "... Please let us know which are words by immediately pressing the F key for those that are words and the J key for those that are not words. Okay? ..." The participants then start to hear the lexical decision stimuli and record their judgments. One-third of the way through the lexical decision stimuli, Sarah addresses (4), repeated here, to each participant in a conversational tone:

- (4) I'm not suggesting that you're responding too slowly, but it's important to give the first answer that comes to mind.

This is followed, two-thirds of the way through the lexical decision stimuli, by the distractor advice in (8):

- (8) It would be good for you to take a deep breath, just to clear your mind.

After the last third of the lexical decision stimuli, approximately three minutes after starting the experiment, participants move on to a question page where they indicate on a five-point Likert scale to what degree they agree that Sarah said the implicature meaning (5) (and four other statements) during the experiment.

Thus, each participant hears only one target (4) and one piece of distractor advice (8), and these are the same for all participants in all three experiments. We could expose each participant to only one target because participants' interaction with Sarah is short, so it would have sounded odd if she had uttered more than one sentence of the form of (4). Being able to record only one judgment from each participant meant that we needed a relatively large number of participants to ensure that we had sufficient power for our statistical analyses. However, the fact that those participants encountered only a single target utterance had the advantage of preventing them from adopting a test-taking strategy, possibly distinct from their usual interpretive strategies in conversation, a problem detected in some implicature studies with multiple parallel examples (Feeney et al., 2004; Guasti et al., 2005). Also, using the same target utterance for all participants avoided the effects of lexical variation on implicature rates, which can be considerable (Clark, 1979; van Tiel et al., 2014). Finally, including (8) as a uniform distractor allowed us to use it as a control to eliminate participants who were not paying enough attention to Sarah to be able to agree that they recalled her having uttered (8).

1.4. ORGANIZATION

In this paper, we report on three web-based experiments in which the critical example (4) is spoken to participants by the personable guide called Sarah

during a decoy lexical decision task, and participants are asked later to what degree they agree that Sarah said what amounts to the implicature in (5). First, in Experiment 1, we investigate the effect on implicature agreement rates of our two independent contextual factors:

- (9) (a) instructions to think carefully about exactly what is said in (4)
- (b) access for the participants to a written verbatim version of (4)

Next, in Experiment 2, we expand the mode of presentation of verbatim access to include audio as well as writing. We replicate our test of (9b) and then add a new condition, in which participants are given the opportunity to replay verbatim audio of (4), in place of reading a transcript.

Finally, in Experiment 3, we investigate a possible mechanism whereby verbatim access might affect implicature rates. Following classic studies on verbatim memory, such as Sachs (1967, 1974), we hypothesize that some participants might have forgotten Sarah's actual words in (4), having stored only something like the contradictory implicature in (5) as the gist of (4). For such participants, renewed access on the question page to the verbatim form of (4) might decrease their chances of agreeing that Sarah had said (5) in uttering (4), since such renewed verbatim access would remind forgetful participants of the exact wording of (4), and the literal compositional meaning of that wording is inconsistent with the implicature in (5). Experiment 3 explores whether such memory restoration through verbatim access could have affected implicature rates in Experiments 1 and 2. We identify a group of participants who are, in fact, forgetful regarding the literal, truth-conditional contribution of the verbatim form of (4), and we test whether these forgetful participants endorse the implicature in (5) as the intended meaning of (4) at a different rate from those who can, on their own, successfully recall something consistent with Sarah's actual words in (4).

The paper ends with discussion of the implications of our verbatim access studies for experimental pragmatics, court proceedings, and the role of literal compositional semantic interpretation.

Archived versions of Experiments 1, 2, and 3 (including all conditions) can be viewed as participants experienced them at <http://spellout.net/ibexexps/VerbatimAccessEffect/Archive/>.

2. Experiment 1

2.1. INTRODUCTION

In order to test the effects on implicature rates of our two contextual factors, (9a), the presence of special instructions to think carefully about exactly what has been said, and (9b), access to a written verbatim transcript, we crossed them in a 2×2 design. One group of participants saw both the special

instructions and a transcript (+Instr/+Trans), a second saw only the instructions (+Instr/-Trans), a third saw only a transcript (-Instr/+Trans), and a fourth saw neither (-Instr/-Trans).

2.2. PARTICIPANTS

We recruited 200 unique native English-speaking participants through Amazon Mechanical Turk. Each had completed at least 1000 Human Intelligence Tasks (HITs) with a minimum 95% approval rating. They were paid \$0.65 for their participation. 254 others participated through the subject pool of the University of Pennsylvania's psychology department, partially fulfilling a course requirement. There were no significant differences in the performance of the Mechanical Turk and student groups in our study, so we have collapsed their results in what follows.

Seven original participants were excluded because they reported that they were not native English speakers. Forty others were excluded because they did not meet the accuracy criteria in (10) (22 participants were excluded by (10a); 18 by (10b)):

- (10) (a) giving a 4 or 5 rating on our Likert scale for the control distractor advice (5), indicating that they agreed that Sarah had said it
 (b) scoring 65% or more correct answers on the lexical decision experiment

This left 407 participants.

To access the experiment, participants recruited through Mechanical Turk were redirected to Ibex, where the experiment was implemented and hosted. Participants from the university pool were linked to the Ibex experiment from the pool's recruitment portal hosted by SONA.

2.3. MATERIALS

The critical items for our implicature study, examples (4) and (8), were spoken by Sarah, a fictional audio guide, during a decoy lexical decision experiment. Sarah's voice was recorded by the first author, a female native speaker of American English from the Philadelphia and New York regions. Intuitive efforts were made to have Sarah sound warm, personable, and genuine as a real guide to the decoy lexical decision experiment. While we did not ask participants explicitly whether they reacted to Sarah as a real person who was actively monitoring their activity and addressing them, the optional comments (offered by only about 12% of participants) were consistent with this view.

The decoy experiment consisted of 32 of the experimental stimuli from Wilder (2016), which were recorded by a male native speaker of American English from the Philadelphia region. Half were words and half non-words;

the non-words were designed to be as similar to words as possible. The series of 32 stimuli was presented three times in random orders.

2.4. PROCEDURE

For participants in the +Instr conditions, the experiment began with the written instructions to pay careful attention to Sarah's utterances in (11). These instructions were designed to mirror naturally occurring directions to pay careful attention, such as people might encounter in courts, classrooms, workplaces, and some experimental contexts:

- (11) This experiment includes an audio guide who will give you instructions for the experiment and then return to give you extra advice. We are especially interested in the accuracy of your reports about what the audio guide says when she gives you advice. Please pay very careful attention to the audio guide's advice, since we will ask you to answer questions about exactly what she said.

After seeing (11), those in +Instr conditions heard Sarah introduce herself, greet them warmly with a promise that "later, I'll be checking in with you with some advice, if it seems like I can help", and give them instructions for the decoy lexical decision task (see 'Appendix A'). When Sarah was speaking, participants saw just a microphone clip art image on their screens.

In contrast, those in the -Instr conditions saw no initial written instructions. For them, the experiment started with Sarah's conversationally delivered introduction, greeting, and decoy-experiment instruction message in 'Appendix A', which was the same for all participants. After Sarah's message ended, all participants clicked to begin the first 32 lexical decision stimuli. As they responded to the stimuli by pressing the F or J keys, they saw on their screens only a progress bar and the reminders "F: Word" and "J: Not a Word". Progress was self-paced, as a new sound was not presented until the participants had responded to the previous one. The mean time taken to complete one 32-item series was approximately 44 seconds.

After the first 32 word recognition stimuli, Sarah interrupted with (12), a greeting followed by the critical sentence from (4):

- (12) Hi, it's Sarah again. I'm not suggesting that you're responding too slowly, but it's important to give the first answer that comes to mind.

Participants clicked to continue to the next third of the decoy experiment, after which Sarah interrupted with (13), another greeting followed by the distractor advice from (8):

- (13) Hi, it's Sarah. It would be good for you to take a deep breath, just to clear your mind.

After clicking to hear the third and last presentation of the 32 lexical decision stimuli, all participants clicked to move on to the question page (see ‘Appendix B’). There, they were thanked for their participation and asked to answer “some questions about the advice that Sarah the audio guide gave you during the experiment”. Next, those in the +Instr conditions were reminded of their task with (14):

- (14) Please think very carefully about exactly what the guide actually says in these pieces of advice, in order to answer the questions below:

Then, all those in the +Trans conditions were presented with transcripts of (4) and (8), as in (15):

- (15) **We’ve written below the two pieces of advice that the audio guide gave you during the experiment:**

I’m not suggesting that you’re responding too slowly, but it’s important to give the first answer that comes to mind.

It would be good for you to take a deep breath, just to clear your mind.

Still on the question page, beneath any special instructions and/or written verbatim transcripts appropriate to their assigned condition, all participants were asked to record their degree of agreement on a five-point Likert scale with five statements. These included three unrelated fillers and (16) and (17) below. (16) is the critical question, measuring to what degree the participants agree that the implicature in (5) is what Sarah said in (4). It was the first question asked of all participants. (17) was the last question and served as a control, measuring to what degree participants agreed that Sarah had said the distractor piece of advice, to which all of them had, in fact, been exposed.

- (16) **Sarah the audio guide said that I was responding too slowly.**

	1	2	3	4	5	
Strongly disagree	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Strongly agree

- (17) **Sarah the audio guide said that it would be good for me to take a deep breath, just to clear my mind.**

	1	2	3	4	5	
Strongly disagree	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Strongly agree

2.5. RESULTS

We asked our participants to provide implicature-agreement rankings on a five-point scale in order to offer flexibility to those who might feel

uncomfortable entirely committing themselves to one of the two logically inconsistent, but equally plausible interpretations of (4). However, the dependent variable for our research question was binary: Did the participant finally prefer to agree, to some degree, with the implicature interpretation of (4) found in (5), or not? Consequently, we treated our question as a yes-or-no one: for the critical question (16), ratings of 4 or 5 were scored as agreeing with (4)'s implicature meaning in (5), while ratings of 1, 2, or 3 were scored as not agreeing with the implicature. For the control question (17), ratings of 4 or 5 were scored as correct; 1, 2, and 3, as incorrect (see 'Appendix C', Tables 2–4, for raw scores and means).

Pearson's chi-squared tests with Yates' continuity correction were performed to compare the +Instr and –Instr conditions across Transcript conditions and the +Trans and –Trans conditions across Instruction conditions. No significant association was found between the presence of special instructions to pay careful attention and low (1–3) scores, indicating that participants did not agree with the implicature interpretation of (4) ($\chi^2(1, N = 407) = 0.21, p > .64$). The proportion of participants agreeing with the implicature interpretation was 57.5% in the –Instr condition, and 56% in the +Instr condition. However, we found a significant association between access to a written transcript of (4) and the low (1–3) scores that indicate a lack of agreement with (4)'s implicature interpretation as (5) ($\chi^2(1, N = 407) = 9.16, p < .003$). The proportion of participants agreeing with the implicature interpretation was 64.9% in the –Trans condition, but only 49.5% in the +Trans condition. When we excluded the neutral 3 ratings (about 15% of total responses) from the 'not-agreeing' group, we found similar results, so the effect was not driven by participants who could not make up their minds about accepting the implicature.

We also recorded reaction times during the decoy experiment, so we could measure whether participants did, in fact, speed up their answers in response to the implicature interpretation of Sarah's advice. That is, we wanted to find out whether participants who indicated that they agreed with the implicature interpretation of (4) started to react measurably faster after hearing it, since they believed they had been told that they were responding too slowly. However, the primary reaction time effect we found was that all groups of participants answered more quickly in each subsequent block of the lexical decision task, as the lexical decision stimuli became more familiar to them. In the context of this strong overall speed-up, we did not find consistent evidence across our experiments that 4–5 responders speeded up significantly more than 1–3 responders. There are many possible explanations for this. It may be that our design did not allow us to detect such a relatively small difference, or perhaps even those who disagreed with the implicature interpretation when giving an explicit rating were still aware of its suggestion

that they should speed up. We leave investigation of these and other possible explanations to future research.

2.6. DISCUSSION

The main results of Experiment 1 reveal that access to a transcript of a previous conversational utterance is associated with significantly lower rates of agreement with the implicature interpretation, while extra instructions to pay attention and think carefully about what the speaker has said are not. (Responses from a post-experiment exit question about participants' awareness of instructions indicated that +Instr participants generally remembered encountering the think carefully instructions during the experiment; they just had no significant effect on implicature rate.) This is a bit surprising since Doran et al. (2012) reported an association between somewhat similar instructions and reduced implicature rates. However, there were crucial differences between that study and ours. First, although Doran et al. tested many different kinds of conversational implicatures, they did not include relevance implicatures like (5). (They also found that different types of implicature yielded different rates of implicature response, so we cannot safely generalize to a new type of implicature.) Second, all their stimuli were presented to participants in writing on the same page as the truth-value judgment task that revealed whether the subject gave the stimulus an implicature interpretation. There was no time-consuming decoy task between utterance and interpretation, as in our Experiment 1. Finally, Doran et al. started with a baseline condition in which trained participants were asked to judge the truth of a written stimulus with a potential implicature. They then added instructions asking subjects to consider the literal meaning of the written stimulus they had read. A third version asked subjects to base their judgments on what they thought a fictional character (Literal Lucy, who takes everything literally) would say. The literal condition showed a significantly lower implicature response rate relative to the baseline condition, and the Literal Lucy condition showed a rate significantly lower than the literal condition.

However, our instructions, unlike Doran et al.'s (2012), did not include the word 'literal' or evoke fictional third parties, but asked for the participants' own carefully thought-out opinions of what Sarah had said when she addressed them. It is not that surprising, then, that our instructions in (11) and (14) did not significantly increase literal interpretations (which we did not ask for), and thereby decrease implicature agreement. As Doran et al. point out, their study shows that ordinary speakers can be taught to distinguish literal from implicature interpretations as linguists would. In contrast, our study shows that speakers with no special training do not

necessarily take being told to think carefully in order to ascertain exactly what has been said to them as a call to find more literal readings and fewer implicature ones. For the majority of our +Instr participants, a careful account of what had been said in a conversation naturally included any implicatures they had derived.

Why, then, is access to a transcript of a previous conversational utterance associated with lower rates of agreement with its implicature meaning? Why might renewed exposure, in writing, to the verbatim form of an earlier utterance correlate with forsaking (or not deriving) an implicature that otherwise would have been embraced? We investigate two features of our transcripts that offer possible explanations: mode of presentation and timing. First, the fact that the transcripts are presented in writing could itself disrupt the process that leads to deriving and committing oneself to an implicature interpretation. This is because conversational implicature depends to a large extent upon the speaker's perception that the norms of conversation are in force (Guasti et al., 2005; Grodner & Sedivy, 2011; Papafragou & Musolino, 2003; Siegel, 2005), but written transcripts are not a normal part of face-to-face conversation. Indeed, writing need not even come from an actual conversational partner who can be assumed to be following Gricean norms. Accordingly, we might expect speakers to be less likely to attribute implicature meaning to sentences presented in writing than to utterances that they hear only conversationally.

Still, the timing of the presentation of our transcript could also have contributed to the lower implicature rates found in our +Trans condition. Sachs (1967) showed that, within 27 seconds of continuing speech after an utterance, addressees forget the specific linguistic form of the utterance and retain only the gist of its meaning. Consequently, participants in our experiments could well have forgotten Sarah's actual words by the time they accessed our question page. (The mean (and median) time lag was 2.6 minutes, and that time was spent, for the most part, doing the lexical decision task, which imposed its own cognitive load.) The remembered gist of Sarah's utterance of (4), for many of them, might have been a version of the implicature in (5), that is, that Sarah had suggested that they were responding too slowly. In such circumstances, seeing the transcript of (4) on the question page would have reminded them of what Sarah had actually said and that its literal compositional meaning was inconsistent with the implicature in (5): Sarah had actually said that she was NOT suggesting that they were responding too slowly. This would have led fewer of them to agree, by giving a 4 or 5 response to the first question on the question page, that Sarah "said that I was responding too slowly".

Thus, we could hypothesize that participants in our +Trans condition exhibited lower implicature rates than those in our -Trans group because the

transcript reminded many of them of the verbatim linguistic form of (4) – and its compositional meaning – which they otherwise would have forgotten by the time they provided their ratings. However, there is a possible objection to such an explanation: subsequent research has shown that the findings of Sachs (1967) (and many others) that memory for verbatim linguistic form disappears almost immediately do not tell the whole story (see Gurevich et al., 2010, and references therein). Verbatim memory fades very quickly for some material in some settings, such as Sachs' passages about impersonal topics like astronomy read by subjects in a formal lab experiment. However, for other material in more natural settings, verbatim memory can persist for nearly a week. Of particular relevance to our example (4), Gibbs (1981), Keenan, MacWhinney, and Mayhew (1977), and Murphy and Shapiro (1994) show that many speakers retain for several minutes or hours the verbatim form of utterances that share (4)'s salient properties: interactiveness, direct emotional connection with the listener, and the ability to give rise to conversational implicature. For longer time spans, Gurevich et al. (2010) provide evidence that the verbatim forms of sentences presented in natural, connected children's stories are remembered at better than chance level even after six days. Taking these findings into account, we cannot assume that many of our participants would have forgotten, by the time they got to the question page, the linguistic form of the potentially insulting implicated personal comment that Sarah addressed to them in (4). If participants had not forgotten the wording of (4), of course, the transcript would not have reminded them of anything, so there could be no memory restoration explanation for Experiment 1's transcript effect.

Further research was necessary to ascertain which, if either, of our explanations could account for the decrease in implicature agreement in the +Trans condition. Is this decrease connected with the transcript's written mode or with its ability to remind participants of the literal meaning associated with the verbatim form?

Experiment 2 was designed to test the mode of presentation explanation. It omits the special think carefully instructions of Experiment 1, since they had been shown to have no significant effect, but adds new conditions that vary the mode of presentation (audio or written) of both Sarah's original advice during the decoy experiment and its verbatim repetition on the question page. Our aims were to replicate the transcript effect of Experiment 1 and, further, to see whether a switch from writing to audio (or vice versa) had any impact on that effect. In particular, if the lowered rate of agreement with the implicature we saw with the written transcript in Experiment 1 disappears with our new audio version of the transcript, that would constitute evidence that the transcript's written form was responsible for the implicature-lowering effect we found in Experiment 1.

TABLE 1. *Experiment 2 conditions by advice mode and repetition mode*

CONDITION NAME:	Written–none	Audio–none	Audio–written	Audio–audio	
				Replay	No replay
Advice mode	written (chatbox)	audio	audio	audio	audio
Repetition mode	none	none	written (transcript)	audio (replay)	none

3. Experiment 2

3.1. INTRODUCTION

The four basic conditions of Experiment 2 were distinguished by the mode of presentation (Written or Audio) of both Sarah’s initial introduction during the decoy experiment of the target advice (4) and the repetition of that advice, if any, on the question page. These conditions are summarized in Table 1.

Note that in Table 1 we have divided the rightmost, Audio–audio condition into two sub-conditions, according to whether Sarah’s audio advice was replayed or not. This was necessary because only 45% of the 175 participants in the Audio–audio condition exercised the option to replay the audio of Sarah uttering (4). Thus, in order to be able to measure any effect associated with hearing Sarah’s advice replayed, we treated participants who actually replayed the audio as belonging to the Audio–audio–replay sub-condition, and those who did not as belonging to the Audio–audio–no replay sub-condition.

3.2. PARTICIPANTS

We recruited 796 unique native English-speaking participants through Amazon Turk. None had participated in Experiment 1, and each had completed at least 1000 HITs with a minimum 95% approval rating. They were paid \$0.65.

We excluded 11 participants because they reported being non-native English speakers, 20 more by (10a), and 71 by (10b), leaving 694 participants for the study.

3.3 MATERIALS

The materials – Sarah’s recordings and the decoy lexical decision experiment – were the same as for Experiment 1, except in the Written–none condition. In that condition, in order to mimic the conversational tone of Sarah’s interactions across modes of presentation as much as possible, we presented

participants with a chatbox. Rather than hearing Sarah say (12) and (13), participants watched her typing each sentence in the chatbox, accompanied by chatbox sounds and the written message “Sarah is typing a message” (see ‘Appendix D’).

3.4. PROCEDURE

The Audio–none and Audio–written conditions of Experiment 2 reproduced exactly the –Instr/–Trans and –Instr/+Trans conditions of Experiment 1, respectively; the procedures for them were identical to those for the corresponding Experiment 1 conditions.

The Written–none condition of Experiment 2 differed from –Instr/–Trans only in that, during the decoy experiment, participants saw Sarah deliver (12) and (13) as written chat messages, rather than hearing recorded speech.

Similarly, the procedure for the Audio–audio condition differed minimally from that for Experiment 1’s –Instr/+Trans condition. There, participants had read on their question page that we had written below the two pieces of advice that Sarah had given them, followed by (4) and (8) written out, as in (15). Participants in Experiment 2’s Audio–audio condition saw (18) instead:

(18) We’ve provided below recordings of the two pieces of advice that Sarah the audio guide gave you during the experiment.

Sarah’s first piece of advice:



Sarah’s second piece of advice:



Participants could click on the buttons to replay the audio of Sarah’s original utterance of (4) and (8) as many times as they liked, parallel to the written transcripts, which participants could also reread as they pleased (see ‘Appendix B’).

3.5. RESULTS

The major finding of Experiment 1 was replicated in Experiment 2. A χ^2 test comparing the Audio–none and Audio–written conditions revealed a significant association between access to a written transcript of (4) on the question page and low (1–3) scores, indicating a lack of agreement with (4)’s implicature interpretation as (5) ($\chi^2(1, N = 349) = 5.14, p < .02$). The proportion of participants agreeing with the implicature interpretation was 65.9% in the Audio–none condition, but only 53.4% in Audio–written.

As for the new conditions of Experiment 2, which reversed modes of presentation, χ^2 tests revealed no significant association between mode of presentation and implicature rate either for Sarah's first utterance during the decoy experiment (Audio–none vs Written–none) or for its repetition on the question page (full Audio–audio/Audio–audio–replay vs. Audio–written). A χ^2 test comparing the full Audio–audio condition with Audio–none also revealed no significant association between access to the audio version of the transcript and more low (1–3) implicature agreement scores.

However, differentiating between the Audio–audio sub-conditions, a χ^2 test comparing Audio–audio–replay and Audio–audio–no replay revealed an association between actually hearing Sarah's advice replayed and more low implicature agreement scores ($\chi^2(1, N = 175) = 4.99, p < .03$). The proportion of participants agreeing with the implicature interpretation was 71.9% in the Audio–audio–no replay sub-condition, but only 54.4% in Audio–audio–replay (see 'Appendix C', Tables 5–6).

3.6. DISCUSSION

The replication, in Experiment 2, of the lowering of implicature rates with access to a written transcript confirms the existence of a transcript effect. Moreover, substituting audio repetition for the written transcript does not significantly alter this effect, so what we have is truly an effect of access during later interpretation to the verbatim form of an utterance, no matter its mode of presentation. That is, there is a general Verbatim Access Effect (VAE).

A cross-modal VAE would seem to predict an association of low (1–3) ratings with the full Audio–audio condition compared with Audio–none, but such an association did not emerge. However, when participants in the Audio–audio condition were distinguished by whether they actually played the audio that defined that condition, we found that fewer than half of them did so. That is, even though (18), which introduced the audio replay buttons, differs minimally from (15), which introduced the written transcripts, just having buttons which had to be clicked to activate the audio effectively made playing the audio of Sarah's advice optional, in a way that seeing the written transcripts provided directly on the question page was not. While we cannot be sure how many participants provided with written transcripts actually read them with any care, it would have been difficult to avoid looking at them at all. Thus, it is not surprising that we found significantly lower implicature agreement ratings in the Audio–written condition (compared with Audio–none), but not with the full Audio–audio group, most of whom had not, in fact, heard the audio repetition. (We did not think it wise to correct the problem of participants' not activating the

audio by having the audio play automatically on the question page, as that would introduce more departures from the written transcript context, in which participants can read and reread exactly when and how often they choose.) Consequently, a proper comparison between written and audio repetition required splitting participants in the Audio–audio condition into replay and no-replay sub-conditions.

When we did this, we found clear evidence that being exposed to the verbatim form of a previous conversational utterance in audio form is associated with significantly lower rates of agreement with the implicature interpretation, compared with no repetition. Not only is the association between audio repetition and low implicature rates significant, but the percentage of Audio–audio–replay participants agreeing with the implicature interpretation by giving a 4/5 rating, 54.4, is a very close match to the 53.4% for the Audio–written condition. Thus, we can conclude that access to audio verbatim form has an implicature-rate lowering effect like that of access to a written transcript, even though it is difficult to present audio in exactly the same way as written transcripts. Experiment 2 shows that differences in the mode of presentation of the verbatim linguistic form cannot account for the effect we found in Experiment 1.

Consequently, we designed Experiment 3 to test our second explanation for the VAE: having access to the verbatim linguistic form of the original conversational utterance reminded participants of the speaker’s exact words, which many participants had forgotten. The renewal of this lost verbatim memory made them less likely to agree with the implicature interpretation, because the implicature was inconsistent with the literal compositional meaning of the speaker’s actual words.

4. Experiment 3

4.1. INTRODUCTION

In order to detect, and then measure, any effect of the restoration of forgotten verbatim memory on implicature agreement rates, Experiment 3 included Sarah’s initial audio presentation of (4) and (8), but no repetition on the question page to refresh participants’ memories. At the end of the experiment, we tested participants’ unaided recall of what Sarah had said in (4) and, on the basis of their responses, divided participants into two groups, verbatim contribution recallers (VRs) and verbatim contribution forgetters (VFs). Since we were interested in the interaction of memory with implicature construals, we did not score for whether participants were able to reproduce Sarah’s utterance word-for-word. Rather, we counted as VR those who recalled something on the topic that was consistent with the literal compositional meaning of (4), that is, any explicit disavowal of (5)’s implicature that the

participant was responding too slowly. For example, those who wrote “I’m not saying that you’re moving too slow ...” or “Not to say that you’re going too slowly ...” were scored as VR, while those who wrote “You’re responding too slowly ...”, “Go faster ...”, or “I’m not trying to correct you, but you may be responding too slowly ...” were scored as VF. Distinguishing these two groups allowed us to test whether VRs and VFs agreed with the implicature interpretation at significantly different rates and to measure whether that difference alone could account for the VAEs of Experiments 1 and 2.

4.2. PARTICIPANTS

We recruited 405 unique native English-speaking participants through Amazon Turk. None had participated in Experiment 1 or 2, and each had completed at least 1000 HITs with a minimum 95% approval rating. They were paid \$0.65.

Ten participants were excluded because they reported being non-native speakers of English, nine more were excluded by (10a), and 45 by (10b), leaving 341 participants for the study.

4.3. MATERIALS

The materials – Sarah’s recordings and the decoy lexical decision experiment – were the same as for Experiments 1 and 2.

4.4. PROCEDURE

The procedure was the same as for the –Instr/–Trans condition of Experiment 1 and the Audio–none condition of Experiment 2, except that, after the final question page, an additional page appeared which asked participants to try to type in a text box Sarah’s entire comment about responding too slowly, exactly as she had said it (see ‘Appendix E’). The median time between hearing Sarah say (4) and reaching the recall questionnaire was 4.3 minutes (mean 4.7 min.; minimum 2.8 min.; maximum 19.3 min.)

Participants’ renderings of (4) were then scored as VR if they expressed something consistent with a disavowal of ‘You’re responding too slowly’ and scored as VF otherwise.

4.5. RESULTS

A χ^2 test comparing the implicature agreement ratings of VFs and VRs reveals that the VR individuals give significantly more 1–3 ratings, indicating that VRs fail to agree with the implicature meaning more often

than VFs ($\chi^2(1, N = 341) = 23.98, p = 9.74e-07$). Of the 117 VF participants, 86.3% agreed with the implicature interpretation, but only 59.8% of the 224 VRs did so (see 'Appendix C', Table 7).

Having ascertained that there is an association between being able to recall (4) verbatim and lower (1–3) Likert ratings indicating lack of agreement with the implicature interpretation, it was important to find out whether this recall effect is large enough to account for the entire VAE of Experiments 1 and 2. If it were not, we would have to look for additional contributors to the VAE. However, it was not possible for us to compare the size of the recall effect we found in Experiment 3 with that of the VAE of Experiments 1 and 2 directly, with a single population, because we could not meaningfully combine the +Transcript condition of Experiment 1 or the Audio–written condition of Experiment 2 with Experiment 3's verbatim recall question. If the written transcript of (4) were presented first, that would be likely to affect the participants' subsequent recall of (4). Similarly, if the recall question were presented first, participants' attempts to recall (4)'s verbatim form would be likely to vitiate the effect of their later exposure to a transcript of (4). Consequently, we made the assumption that the percentage of VFs in Experiment 2 and Experiment 3 are equal, which seems plausible, given our methods of recruiting individuals into the study. Under this assumption, we removed from the results of the Audio–none condition of Experiment 2 those that would have come from the VFs among the Audio–none participants, and analyzed the result. That is, on the basis of Experiment 3, in which 34.31% of participants were VFs, we assumed that the Audio–none condition of Experiment 2 also included 34.31% VFs, or 59.36 of the 173 participants. We removed these 59.36 presumptive VF individuals from the Audio–none results proportionally, according to the distribution of the 1–5 ratings given by those in the Experiment 3 VF group. For instance, 82 Experiment 3 VFs gave 5 ratings, 70.09% of the total 117. To remove the Audio–none 5 ratings coming from VFs, we subtracted 70.09% of the total 173 (41.60), leaving 37.40 5 ratings which would have come from VRs (see 'Appendix F', Tables 8–9).

A χ^2 test comparing the resulting new Audio–none condition (now stripped of its presumptive VF responses in each rating category) with the Audio–written condition from Experiment 2 showed that there was no longer any significant difference in implicature rates between the two ($\chi^2(1, N = 290) = 0.033, p > .8$). In fact, removing the presumed VFs from Audio–none virtually erases the VAE: the proportion of implicature agreement (4/5 ratings) for the original Experiment 2 Audio–none condition had been 65.9%, but for the new, VF-less Audio–none, it is down to 55.2%, very close to the 53.4% implicature agreement rate of the Audio–written condition (see 'Appendix F', Table 10).

4.6. DISCUSSION

The results of Experiment 3 support our second, memory restoration, explanation of the VAE. First, we find that about a third of participants did, indeed, forget exactly what Sarah had said during our experiments, along with its associated compositional meaning. For almost all these VFs, the remembered gist of (4) was only the contradictory implicature (5). (The 16 exceptional VFs who gave 1–3 ratings, indicating lack of agreement with (5), were evenly divided between those whose responses to the recall question indicated that they actually agreed that Sarah had said the implicature in (5), their low agreement rating notwithstanding, and those who had responded to an utterance of Sarah's other than the first clause of (4).)

Second, we found that recall – or lack thereof – for the verbatim form's semantic contribution affects one's ultimate choice of interpretation. Our newly identified VF group was significantly more likely than the VR group to endorse the implicature in (5) as what Sarah said in (4).

Finally, this higher rate of implicature agreement found among VFs reveals a mechanism that can account for the VAE. In Experiments 1 and 2, the evidence for the VAE is that participants who are given renewed access to the verbatim form of (4) agree with its implicature meaning significantly less often than those who enjoy no renewed access as they later commit themselves to an interpretation. On the basis of Experiment 3, we assumed that Experiment 2 also included about one-third VFs and two-thirds VRs. When we took the responses in Experiment 2's Audio–none condition, which offered no renewed access to the verbatim form of (4), and eliminated responses from its presumptive VFs, we were left with a presumptive VR implicature-agreement rate that matched that of Audio–written, whose participants had been provided a written transcript of (4) on their question page. That is, for the purposes of implicature agreement, giving a group of participants access to the verbatim form of a previous conversational utterance turns them all into good verbatim recallers. The lowering effect on implicature rate is the same, whether a person independently remembers what the speaker said or whether she is reminded of it by an external source.

5. Conclusion

We have shown that later access to the verbatim form of a previous conversational utterance is associated with a significant decrease in agreement with its implicature interpretation. Thus, when courts, legislative bodies, or even news outlets, present decision-makers with ongoing access to the verbatim form of past critical conversations, they reduce the chances of realistic construals of those conversations by introducing an inherent literal meaning bias. Our studies have similar implications for experimental

pragmatics, where it is also common to provide ongoing access to the verbatim form of utterances meant to be taken conversationally. Of course, experiments do not often include our paradigm's three-minute time lag between a naturally occurring conversation and its interpretation, so further research would be required to ascertain how long a delay, if any, is necessary for the VAE to manifest itself. Such research focusing on the timecourse of the VAE might also answer other questions about it: Do we observe a VAE because, over the three-minute lag time, as participants' memory of the verbatim form fades, implicature readings strengthen? Or does the initial derivation and encoding of an implicature reading somehow interfere with retrieving the verbatim form beginning immediately at the time of utterance, with no change over time? Whatever role time plays in producing the VAE, our findings suggest strongly that those who seek to create contexts in which speakers will interpret utterances as they would in naturally occurring conversation serve their purpose better by creating natural conversational contexts and withholding ongoing verbatim access.

Our findings also bear on some theoretical issues. About two-thirds of our participants in all three experiments agreed that Sarah had said that they were responding too slowly (the implicature). Even though this implicature contradicts the verbatim compositional meaning of (4), it was taken to be what the speaker of (4) had said, even by those who had been instructed to think carefully about what they had heard and report "exactly what was said" in the utterance. Thus, we know that naive speakers often take a relevance conversational implicature, not the literal compositional meaning, as *WHAT WAS SAID*. (Horton, Schmader, & Ward, 2016, show similar behavior with other types of conversational implicature.)

In addition, Experiment 3 makes it clear that accepting the implicature in (5) as what Sarah has said is still consistent with participants' retaining access to the literal compositional meaning associated with the verbatim form (4): 'I'm not suggesting that you're responding too slowly.' About half of the VRs, good recallers of the semantic contribution of (4), nevertheless endorsed the implicature interpretation in (5) with a 4 or 5 rating. Thus, we know that many speakers will retain forms of both the literal semantic contribution of the verbatim form in (4) and the contradictory implicature in (5). (We can conclude further that VFs who gave 4/5 ratings retained only the implicature, since they failed to write the verbatim meaning when asked to do so. However, we cannot be sure that VRs who gave low Likert ratings retained only the literal meaning, as we did not ask them to write down the implicature.) Presumably, speakers who retain forms of both interpretations make a decision about which is the intended meaning on the basis of complex online contextual cues about the situation, the speaker, and their goals (Clark, 1979; Roberts, 2017).

What is more mysterious is what these stored interpretations look like and how they are derived. Predicting the derivational history and structure of a freshly generated conversational implicature is beyond the scope of this paper, but our Experiment 3 sheds some light on the form of the stored literal semantic interpretations of Sarah's utterance of (4). Consistent with Sachs (1967, 1974), we found that precise verbatim memory fades quickly. By the time even the VRs, our good recallers of the verbatim contribution in Experiment 3, got to our recall question, only 15.7% of them recalled the exact, word-for-word form of (4). However, Sachs and subsequent verbatim memory researchers have noted that memory for meaning is extremely durable. Accordingly, we saw that two-thirds of our Experiment 3 participants recalled the form of (4) with enough accuracy to be classified as VRs in the first place. That is, they wrote down something that made it clear that they recalled that a literal compositional account of Sarah's utterance constituted a disavowal of the implicature interpretation that they were responding too slowly. They were able to recall this much of the literal meaning even though half of these VRs had indicated with a 4 or 5 rating that this very disavowed implicature was, in fact, what Sarah had said to them.

Examination of Experiment 3 participants' renderings of Sarah's verbatim utterance raises an important question: How are such retained representations related to the predictable output of compositional semantics operating on linguistic structure? While more study would be required to reach definite conclusions, we can see that participants' most common deviations from correct verbatim form are mediated by systematic pragmatic factors. First, of course, there are the implicature interpretations: nearly all the VFs simply substituted for the verbatim form of (4) something that shared the truth conditions of the implicature (5). Another popular substitution involved replacing a less conventional form with a more conventional one, a tendency noted by Clark (1979). We had purposely written (4) as "I'm not suggesting *p*" in order to avoid using the more conventionalized 'I'm not saying *p*'. Nevertheless, most (86%) of our VR subjects substituted 'saying' for 'suggesting' in their attempts to recall Sarah's exact words. Finally, a much smaller group exhibited the effects of another pragmatic pressure noted by Clark: the tendency to treat lexically distinct politeness formulae as equivalent. Thus, ten people replaced 'I'm not suggesting *p*' with other politeness expressions, including 'I'm not trying to be mean', 'Not that I want to worry you', and 'Sorry to bother you'.

We have shown that, after just a few minutes, what addressees retain already shows signs of adjustment to systematic pragmatic forces of Gricean principles, conventionality, and the interchangeability of politeness formulas. Thus, our experiments suggest that there are limitations to how much linguistic structure determines the interpretations that speakers actually take away. While we cannot fully address this large question here,

our findings are consistent with the kind of parallel processing models suggested by Roberts (2017) and Huang and Snedeker (2018), in which there is ongoing interaction between top-down pragmatic forces and bottom-up compositional interpretation. The Verbatim Access Effect that we have documented here is just one manifestation of the complex interplay between contextual factors and compositional semantics in determining speakers' interpretation of what has been said.

REFERENCES

- Barner, D., Brooks, N. & Bale, A. (2011). Accessing the unsaid: the role of scalar alternatives in children's pragmatic inferences. *Cognition* **188**, 87–96.
- Bill, C., Romoli, J., Schwarz, F. & Crain, S. (2016). Scalar implicatures versus presuppositions: the view from acquisition. *Topoi* **35**, 57–71.
- Bonnefon, J.-F., Feeney, A. & Villejoubert, G. (2009). When some is actually all: scalar inferences in face-threatening contexts. *Cognition* **112**, 249–258.
- Bott, L. & Noveck, I. A. (2004). Some utterances are underinformative: the onset and time course of scalar inferences. *Journal of Memory and Language* **51**, 437–457.
- Clark, H. H. (1979). Responding to indirect speech acts. *Cognitive Psychology* **11**, 430–477.
- Degen, J. (2015). Investigating the distribution of *some* (but not *all*) implicatures using corpora and web-based methods. *Semantics and Pragmatics* **8**, 1–55.
- Degen, J. & Tanenhaus, M. K. (2015). Availability of alternatives and the processing of scalar implicatures: a visual world eye-tracking study. *Cognitive Science* **40**, 172–201.
- Doran, R., Ward, G., Larson, M., McNabb, Y. & Baker, R. E. (2012). A novel experimental paradigm for distinguishing between what is said and what is implicated. *Language* **88**(1), 124–154.
- Feeney, A., Scafton, S., Duckworth, A. & Handley, S. J. (2004). The story of *some*: everyday pragmatic inference by children and adults. *Canadian Journal of Experimental Psychology* **58**(2), 121–132.
- Gibbs, R. W. (1981). Memory for requests in conversation. *Journal of Verbal Learning and Verbal Behavior* **20**, 630–640.
- Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (eds.), *Syntax and Semantics vol. 3: Speech Acts* (pp. 41–58). New York: Academic Press.
- Grodner, D., Kim, M. & Russell, B. (2016, May). A Bayesian account of conversational implicature. Paper presented at a meeting of Common Ground Colloquium, University of Pennsylvania.
- Grodner, D. & Sedivy, J. C. (2011). The effect of speaker-specific information on pragmatic inferences. In E. Gibson and N. J. Pearlmuter (eds.), *The processing and acquisition of reference* (pp. 239–272). Cambridge, MA: MIT Press.
- Guasti, M. T., Chierchia, G., Crane, S., Foppolo, F., Gualmini, A. & Meroni, L. (2005). Why children and adults sometimes (but not always) compute implicatures. *Language and Cognitive Processes* **20**(5), 667–696.
- Gurevich, O., Johnson, M. & Goldberg, A. (2010). Incidental verbatim memory for language. *Language and Cognition* **2**(1), 45–78.
- Horn, L. R. (1984). Toward a new taxonomy of pragmatic inference: Q- and R-based implicature. In D. Shiffrin (ed.), *Meaning, form, and use in context* (pp. 11–42). Washington, DC: Georgetown University Press.
- Horton, W. S., Schmader, C. & Ward, G. (2016). On the incorporation of generalized conversational implicatures into what is said: an experimental investigation. Poster presented at the meeting of the Linguistic Society of America, Washington, DC.
- Huang, Y. T. & Snedeker, J. (2018). *Some* inferences still take time: prosody, predictability, and the speed of scalar implicatures. *Cognitive Psychology* **102**, 105–126.

- Katsos, N. & Bishop, D. V. M. (2011). Pragmatic tolerance: implications for the acquisition of informativeness and implicature. *Cognition* **120**(1), 67–81.
- Keenan, J. M., MacWhinney, B. & Mayhew, D. (1977). Pragmatics in memory. *Journal of Verbal Learning and Verbal Behavior* **16**, 549–560.
- Levinson, S. C. (2000). *Presumptive meanings: the theory of generalized conversational implicature*. Cambridge, MA: MIT Press.
- Mazzarella, D. (2015). Politeness, relevance and scalar inferences. *Journal of Pragmatics* **79**, 93–106.
- Murphy, G. L. & Shapiro, A. M. (1994). Forgetting of verbatim information in discourse. *Memory and Cognition* **22**, 85–94.
- Noveck, I. A. (2001). When children are more logical than adults: experimental investigations of scalar implicatures. *Cognition* **80**, 253–282.
- Papafragou, A. & Musolino, J. (2003). Scalar implicatures: experiments at the semantics-pragmatics interface. *Cognition* **86**, 253–282.
- Papafragou, A. & Tantalou, N. (2004). Children’s computation of implicatures. *Language Acquisition*. **12**(1), 71–82.
- Prince, E. F. (1990). On the use of social conversation as evidence in a court of law. In J. N. Levi & A. G. Walker (eds.), *Language in the judicial process: Vol. 5* (pp. 279–289). Berlin: Springer.
- Roberts, C. (2017). Linguistic convention and the architecture of interpretation. *Analytic Philosophy* **58**(4), 418–439.
- Sachs, J. S. (1967). Recognition memory for syntactic and semantic aspects of connected discourse. *Perception and Psychophysics* **2**(9), 437–442.
- Sachs, J. S. (1974). Memory in reading and listening to discourse. *Memory and Cognition* **2**(4), 95–100.
- Sauerland, U. (2004). Scalar implicatures in complex sentences. *Linguistics and Philosophy* **27**(3), 267–291.
- Siegel, M. (2005). Finding conversational facts: a role for linguistics in court. *International Journal of Speech, Language and the Law* **12**(2), 255–278.
- Sikos, L., Kim, M., Anchirai, R., Lam, H. & Grodner, D. J. (2016). Speaker likeability leads to utterance acceptability: social context modulates tolerance for pragmatic violations in adults. Poster presented at the meeting of the CUNY Sentence Processing Conference, Gainesville, FL.
- Smith, C. L. (1980). Quantifier and question answering in young children. *Journal of Experimental Child Psychology* **30**, 191–205.
- Sperber, D. & Wilson, D. (1995 [1986]). *Relevance: communication and cognition* (2nd ed.). Oxford: Blackwell.
- van Tiel, B., van Miltenburg, E., Zevakhin, N. & Geurts, B. (2014). Scalar diversity. *Journal of Semantics* **33**(1), 137–175.
- Wilder, R. (2016). *Examining episodic information in speech perception* (Unpublished doctoral dissertation proposal). University of Pennsylvania, Philadelphia, PA.

Appendix A

Audio guide’s initial greeting for all experiments

Hi, I’m Sarah, and I’m going to be your guide for this experiment. I’ll let you know what you need to do, and, then, later, I’ll be checking in with you with some advice, if it seems like I can help. OK, so here are your instructions for the experiment: You’re going to be hearing some sounds. Some of them will be words, and some of them will not be words. Please let us know which are words by immediately pressing the F key for those that are words and

the J key for those that are not words. Okay? Now, please get your fingers ready on the F and J keys. Remember, press F, as in Frank, for words and J, as in John, for not words. Now, we're going to get started.

Appendix B

Question page for all experiments (indicating adaptations for different conditions)

Thank you for completing our word experiment. Now, we'd like you to answer some questions about the advice that Sarah the audio guide gave you during the experiment.

(For +Transcript conditions of Experiment 1, insert (15) here.)

(For +Instructions conditions of Experiment 1, insert (14) here.)

(For Audio-audio condition of Experiment 2, insert (18) here.)

Please indicate how strongly you agree or disagree with each statement below by clicking the button beneath the number 1, 2, 3, 4, or 5. 1 means that you strongly disagree with the statement, 2 means you disagree somewhat, 3 means you neither agree nor disagree, 4 means you agree somewhat, and 5 means you strongly agree with the statement.

Sarah the audio guide said that I was responding too slowly.

1 2 3 4 5

Strongly disagree Strongly agree

Sarah the audio guide said that I should probably have tried a little harder to pay attention.

1 2 3 4 5

Strongly disagree Strongly agree

Sarah the audio guide said that the experiment was boring.

1 2 3 4 5

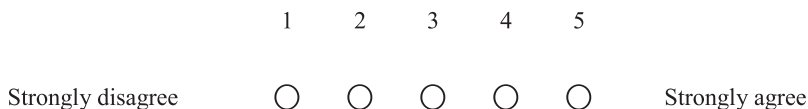
Strongly disagree Strongly agree

Sarah the audio guide said that it was important to give the first answer that came to mind.

1 2 3 4 5

Strongly disagree Strongly agree

Sarah the audio guide said that it would be good for me to take a deep breath, just to clear my mind.



Appendix C

Experiments 1, 2, and 3, 1–5 responses

TABLE 2. *Experiment 1: 1–5 responses by Instructions and Transcript*

Instr	Trans	1	2	3	4	5	%4/5	N	Mean(SD)
–	–	9	13	22	27	46	62.4	117	3.75(1.29)
+	–	13	5	5	18	33	68.9	74	3.72(1.51)
–	+	14	13	14	18	24	50.6	83	3.30(1.45)
+	+	30	18	20	22	43	48.9	133	3.23(1.56)

TABLE 3. *Experiment 1: 1–5 responses by Instructions, across ±Transcript*

Instructions	1	2	3	4	5	%4/5	N	Mean(SD)
–	43	23	25	40	76	56.0	207	3.40(1.56)
+	23	26	36	45	70	57.5	200	3.56(1.38)

TABLE 4. *Experiment 1: 1–5 responses by Transcript, across ±Instructions*

Transcript	1	2	3	4	5	%4/5	N	Mean(SD)
–	22	18	27	45	79	64.9	191	3.74(1.38)
+	44	31	34	40	67	49.5	216	3.25(1.52)

TABLE 5. *Experiment 2: 1–5 responses by Advice and Repetition Modes*

Advice Mode	Repetition Mode	1	2	3	4	5	%4/5	N	Mean(SD)
Written	None	12	16	15	43	84	74.7	170	4.01(1.26)
Audio	None	21	14	24	35	79	65.9	173	3.79(1.39)
Audio	Audio	18	22	23	44	68	64.0	175	3.70(1.36)
Audio	Written	27	34	21	35	59	53.4	176	3.37(1.48)

TABLE 6. *Experiment 2: 1–5 responses from Audio–audio participants by Replay option*

Audio advice replayed?	1	2	3	4	5	% 4/5	<i>N</i>	Mean(<i>SD</i>)
No	4	11	12	23	46	71.9	96	4.00(1.19)
Yes	14	11	11	21	22	54.4	79	3.33(1.46)

TABLE 7. *Experiment 3: 1–5 responses by verbatim contribution recall*

Verbatim contribution recalled?	1	2	3	4	5	% 4/5	<i>N</i>	Mean(<i>SD</i>)	
No (VF)		6	5	5	19	82	86.3	117	4.41(1.10)
Yes (VR)		23	34	33	63	71	59.8	224	3.56(1.34)

Appendix D

Experiment 2: Written–none chatboxes



Appendix E

Experiment 3: post-experiment recall question

During the experiment, Sarah made a comment that mentioned something about responding too slowly.

Please try to remember all the wording of this particular comment and do your best to type the entire comment in the box below, exactly as Sarah said it.

Appendix F

Experiment 3: removing presumed verbatim contribution forgetters' 1–5 ratings from Experiment 2 Audio–none condition

TABLE 8. *Experiment 3: percentage of 1–5 ratings by verbatim contribution recall*

Verbatim contribution Recalled?	%1	%2	%3	%4	%5	N	% total	N = 341
No (VF)	5.13	4.27	4.27	16.24	70.09	117	34.31	
Yes (VR)	10.27	15.18	14.73	28.12	31.70	224	65.59	

TABLE 9. *Experiment 3: number of presumed Verbatim Contribution Forgetter (VF) responses to remove, by 1–5 rating, from Experiment 2 Audio–none condition (Table 5), based on Table 8 percentages*

Rating	1	2	3	4	5	total
# of responses	3.04	2.54	2.54	9.64	41.60	59.36

TABLE 10. *Experiment 3: Audio–none and Audio–written 1–5 responses (from Experiment 2, Table 5) compared with revised 1–5 Audio–none responses after removing presumed Verbatim Contribution Forgetters (VFs)*

	Advice Mode	Repetition Mode	1	2	3	4	5	%4/5	N
Exp. 2:	Audio	None	21	14	24	35	79	65.9	173
Exp. 2:	Audio	Written	27	34	21	35	59	53.4	176
Exp. 3: (noVF)	Audio	None	17.96	11.46	21.46	25.36	37.40	55.2	113.64