Tomorrow isn't always a day away: non-utterance time readings of *tomorrow* *

Carolyn Jane Anderson
University of Massachusetts, Amherst

Abstract This paper explores non-utterance time readings of *tomorrow*, which, though unexpected under the pure indexical view of *tomorrow*, are found in American English corpus data:
1. Last week, UPS said that the package would be delivered tomorrow.
(a) Utterance time reading: I hope it arrives on time!
(b) Non-utterance time reading: It still hasn't shown up!
I present experimental evidence that non-utterance time readings of *tomorrow* are accepted by many American English speakers and are not restricted to environments for which context shift has been proposed, such as Free Indirect Discourse and attitude reports. I propose an analysis of *tomorrow* as anaphoric to perspectives, and present quantificational binding evidence in support of this view.

Keywords: temporal adverbials, context shift, perspective

Word count: 11653

1 Introduction

Indexicality and anaphoricity are two types of context sensitivity. Indexicality is reference relative to the context of utterance, while anaphoricity is reference to objects introduced by the preceding discourse context. While some expressions are purely indexical or purely anaphoric, a growing number of expressions have been found to have both anaphoric and indexical uses (as recently discussed in Maier (2017)). In some cases, anaphoric uses have been miscategorized as indexical uses because the referent is something unexpectedly complex; for instance, the temporal adverbial *now*, once though to refer indexically to the time index of the utterance context, has been reanalyzed as anaphoric to a prominent event and its result state (Stojnić & Altshuler 2019).

* Many thanks to Daniel Altshuler, Brian Dillon, and John Kingston for their guidance and encouragement; and to Ana Arregui, Rajesh Bhatt, Amy Rose Deal, Jack Duff, Lyn Frazier, Emar Maier, Kristen Syrett, Ming Xiang, and reviewers and participants at the 2018 CUNY Sentence Processing Conference and Sinn und Bedeutung 2018 for their feedback and suggestions.

Anaphoric reference and indexical reference can be particularly difficult to distinguish in the case of perspectival expressions. For instance, although the perspectival motion verb *come* has been analyzed as indexical (Oshima 2006), Barlew (2017) argues that it is, in fact, anaphoric to perspectives. Because the speaker's utterance-time perspective is often the preferred perspective (Harris & Potts 2009), and because it is usually consistent with the context of utterance, it is often challenging to tease apart expressions that are perspective-anaphoric from those that are indexical.

I argue that *tomorrow* is, for a substantial group of American English speakers, only illusorily a pure indexical (Kamp 1971, Kaplan 1989); like *come*, it is actually anaphoric to perspectives and appears to be a pure indexical only because of the significant overlap in licensing conditions between pure indexicals and perspective-anaphoric expressions. I look at *tomorrow* in environments in which the prominent perspective is not an utterance-time perspective, and show that its behavior is not consistent with pure indexicality. The data that I consider are non-utterance time readings of *tomorrow*, as in (1).

(1) Last week, Jane said that she would order the cake tomorrow, but she didn't.

This sentence does not make sense if *tomorrow* is interpreted as the day after utterance time; if it is judged felicitous, *tomorrow* must refer to the day after the saying event. Although such readings have been reported to be unavailable, I will show that they are accepted by many (but not all) American English speakers.

The existence of non-utterance time interpretations of *tomorrow* does not disprove that it is a pure indexical, since non-utterance time readings of pure temporal indexicals can arise when the context parameter is manipulated. This has been proposed to occur in Free Indirect Discourse (Schlenker 2004, Sharvit 2008, Eckardt 2014) and in speech and attitude reports in a number of languages (Schlenker 2003, Quer 2005, Anand & Nevins 2014, Deal 2014).

I present a series of experimental studies that narrow the hypothesis space for non-utterance time readings of *tomorrow*, considering two main possibilities: (1) that they arise through context shift and *tomorrow* is indeed a pure indexical; and (2) that they are anaphoric uses of *tomorrow*. In the first half of the paper, I take up the context shift hypothesis. In Experiments 1a and 1b, I establish a baseline of acceptability for non-utterance time readings. In Experiments 2a, 2b, and 3, I show that such readings arise outside of two environments for which context shift has been proposed: Free Indirect Discourse and attitude reports.

In the second half, I lay out a perspective-anaphoric account of *tomorrow*. I present two quantificational binding tasks that show that *tomorrow* is not anaphoric to just any discourse-given time, and that the acceptability of *tomorrow* in binding environments improves when the context allows perspectives to co-vary with the

quantifier. On the basis of this data, I argue that *tomorrow* is anaphoric to the time index of a perspective, and that this behavior has been overlooked because this index is almost always identical to the time index of the context parameter.

My data and analysis suggest that perspective may play a role in the interpretation of more expressions than previously thought, highlighting the need to explore whether context-sensitive items like shifty indexicals are also sensitive to perspective. In addition, this work reveals a temporal parallel to perspective-anaphoric expressions in the spatial domain like the motion verb *come*.

2 Analyzing non-utterance time *tomorrow*

A central question about non-utterance time interpretations of *tomorrow* is whether they are indexical or anaphoric. An expression is used *indexically* when it refers relative to the context parameter of the sentence; an expression is used *anaphorically* when it refers relative to a reference point given by the discourse context.[1] While some expressions have both kinds of uses, some do not: Kaplan (1989) defines a class of *pure indexicals* whose interpretation is always indexical.

In Kaplan (1989), *tomorrow* is described as a pure indexical, which makes non-utterance time interpretations surprising unless they occur only when the context parameter has been manipulated. Alternatively, these uses of *tomorrow* may be anaphoric, in which case *tomorrow* is not a pure indexical. Thus, there are two types of explanations for non-utterance time interpretations of *tomorrow*: indexical accounts, which maintain the pure indexical status of *tomorrow*, and anaphoric ones, which posit that non-utterance time readings arise through anaphoric reference.

2.1 Context shift accounts of non-utterance time *tomorrow*

Under a pure indexical view of *tomorrow*, *tomorrow* refers strictly to the time index of the context parameter.

(2) Pure indexical semantics for *tomorrow*:
$[[\text{tomorrow}]]^{c,g} = \lambda Q.\lambda e.\tau(e) \subset \iota t.\text{DAY-AFTER}(t, c_t) \wedge Q(e)$
where c_t is the temporal index of the context parameter c

If this is the case, then non-utterance time readings can only arise if the context parameter has been manipulated so that its time index differs from utterance time. There are two environments which are proposed to involve manipulations of the context parameter: Free Indirect Discourse, and speech / attitude reports.

1 This is one common view of context sensitivity; there are also views of context in which anaphoric expressions are treated as pure indexicals (Stojnić 2016), and views in which pure indexicals are treated as anaphoric expressions (Roberts 2015).

2.1.1 Analysis 1: non-utterance time readings are due to Free Indirect Discourse effects

One environment in which context shift has been proposed to occur is Free Indirect Discourse (FID), a literary style in which tense and person pronouns are used relative to the narrator's perspective, while temporal and locative indexicals, expressives, and perspectival items are used relative to that of the protagonist (Banfield 1982).[2]

Many analyses of Free Indirect Discourse posit that some manipulation of the context parameter is responsible for the protagonist-oriented interpretations. One approach is to use two types of context parameters: the external context parameter C, representing the narrator's utterance situation, and an internal context parameter c, representing the protagonist's thought situation.[3] While in direct speech, only the external context is available, Free Indirect Discourse environments introduce an internal context. When there are two context parameters, the shiftable indexicals (temporal and locative) refer relative to the internal context, while the rigid indexicals (person) remain fixed to the external context.

(3) FID derivation of non-utterance time *tomorrow*:
$[[\text{tomorrow}]]^{C,c,g} = \lambda Q.\lambda e.\tau(e) \subset \imath t.\text{DAY-AFTER}(t, c_t) \wedge Q(e)$
where C is the external context parameter, c is the internal context parameter, and c_t is the temporal index of c

One possible explanation of non-utterance time readings of *tomorrow* therefore is that they arise only in Free Indirect Discourse environments. In a FID system like the one sketched above, non-utterance time readings would arise when *tomorrow* is interpreted relative to an internal context. This account is consistent with pure indexicality, since *tomorrow* would refer to a context parameter even when receiving protagonist-oriented interpretation.

A Free Indirect Discourse context-shift account of non-utterance time *tomorrow* makes several predictions. First, we would expect that other expressions, such as epithets and expressives, that are protagonist-oriented in FID would receive non-utterance time interpretations whenever *tomorrow* does. Second, non-utterance time readings should not be possible in narrator-oriented clauses, clauses where a first-person subject is reporting their own thoughts (Banfield 1982).

2 Protagonist Projection is another environment in which context-sensitive expressions are interpreted relative to the protagonist; however, in Protagonist Projection, temporal indexicals are fixed to the narrator's perspective, which makes it not directly relevant to our discussion of non-utterance time *tomorrow* (Stokke 2013, Abrusán 2018).

3 This analysis was proposed by Doron (1991) and adopted by Eckardt (2014), but see Schlenker (2004), Sharvit (2008), and Maier (2015) for other analyses.

2.1.2 Analysis 2: *tomorrow* is a shifty indexical

Another environment in which context shift has been proposed is under speech and attitude verb embedding. In a number of languages, there are so-called *shifty indexicals*, indexicals whose interpretation changes when they are embedded under speech or attitude verbs.[4] When embedded, they are interpreted relative to the embedded context rather than the matrix context.

Although person indexicals are the most common kind of shifty indexical, shifty temporal indexicals have also been documented. In the Korean example in (4), for example, *nayil* 'tomorrow' can refer either to the day after utterance time (matrix interpretation), or the day after Mary's speech act (shifted interpretation).

(4) Context: It is January 8th.

cinan	cwu-ey	Mary-ka	nwuka	nayil	ttenanta-ko
last	week-in	Mary-NOM	who-NOM	tomorrow	leave-C
malhayss-ni?					
said-Q					

'Who did Mary say a week ago would leave on January 2nd/9th?' (Park 2018)

The dominant analysis of indexical shift posits a covert syntactic operator that can shift the context parameter in speech-embedding environments (Schlenker 2003, Anand & Nevins 2014, Deal 2014). Under such analyses, the indexicals involved are still considered pure indexicals, because they are evaluated relative to a context parameter.

(5) Indexical shift derivation of non-utterance time *tomorrow*:
$[[\text{Mary said OP}_t\text{John would come tomorrow}]]^{c,g} = \exists e.\text{SAY}(e) \wedge \text{Agent}(Mary, e) \wedge \text{Theme}([[\text{John would come tomorrow}]]^{c^{[TIME(e) \rightarrow Time]},g}, e)$

There are two diagnostics for indexical shift that arise from the context shift operator view of the phenomenon: the existence of Shift Together effects for shifty indexicals; and the unacceptability of shifty indexicals in quantificational binding contexts.

Because the shift operator overwrites the matrix context parameter, if one indexical shifts, all others in the same syntactic domain ought to shift as well,[5] since the matrix context parameter is no longer accessible.[6] This springs from an obvious

4 For more in-depth discussion of the phenomena, see Deal (2017).

5 At least, if they are the same type; Deal (2014) proposes separate shift operators for time and person parameters.

6 The data on Shift Together effects in some languages is contested; see Deal (2017) for more discussion.

property of pure indexicals, which is that they receive uniform interpretations since they all refer relative to the context parameter.

In addition, because shifty indexicals are still pure indexicals, they are not licit in quantificational binding contexts. In (6), for example, the indexical *cikum* 'now', despite being shiftable, is infelicitous because of the quantification (Park 2018).

(6)

*Obama-ka	malhal	ttyaymyun	manhun	salamtul-i	cikum
Obama-NOM	speaks	when	many	people-NOM	now
pakswuchinta.					
clap					

Intended: 'When Obama speaks, many people clap now.' (Park 2018)

Thus, shifty indexicals behave like pure indexicals except when they are embedded under an attitude predicate.

One possible explanation for non-utterance time interpretations of *tomorrow* in English, then, is that *tomorrow* is a shifty indexical, and that such readings arise from context-shift. This account, like the FID account, is consistent with pure indexicality, given that the dominant analyses of indexical shift posit that the original context parameter is overwritten by a context parameter representing the embedded context.

An indexical shift account makes three predictions. First, non-utterance time interpretations should only arise when *tomorrow* is embedded under a speech or attitude verb. Second, like other indexicals, *tomorrow* should be infelicitous in quantificational binding environments. Lastly, we expect uniform behavior across temporal indexicals in the same clause (with the caveat that some indexicals seem to be rigidly referent to the matrix context); we would expect other temporal adverbials, like *yesterday*, to behave similarly to *tomorrow*.

2.2 Analysis 3: *tomorrow* is anaphoric

Above I have outlined two context-shift accounts of non-utterance time *tomorrow*. There may be other environments where context-shift can occur, in which case, *tomorrow* may fail to meet the predictions of either account and yet be a pure indexical. Another possibility, however, is that *tomorrow* is not a pure indexical and that non-utterance time readings are in fact anaphoric uses of *tomorrow*.

(7) Anaphoric semantics for *tomorrow*:
$[[\text{tomorrow}]]^{c,g} = \lambda Q.\lambda e.\tau(e) \subset \iota t.\text{DAY-AFTER}(t,t') \wedge Q(e)$
where t' is a prominent time in the Common Ground

Under an anaphoric account of *tomorrow*, we would expect *tomorrow* to be licit in quantificational binding contexts. Pure indexicals resist quantificational binding;

because pure indexicals refer strictly relative to the utterance context, which does not vary, examples like (8) can only mean that for each place that the speaker visits, it rains at the utterance location, which is nonsensical.

(8) # Everywhere I visit, it rains here.

By contrast, the referent for the anaphoric expression *the next day* can co-vary with the quantification over times,[7] leading to a felicitous interpretation.

(9) Whenever I wash my car, it rains the next day.

In order to be able to apply the quantificational binding diagnostic, however, we have to understand the referent of the expression of interest. As work on *now* highlights, temporal adverbials with anaphoric uses are not always anaphoric to any prominent time in the discourse context as in the anaphoric semantics in (7); they may be anaphoric to something more complex, such as a result state (Altshuler & Stojnić 2015, Altshuler 2017, Stojnić & Altshuler 2019).

When this is the case, anaphoric expressions may appear indexical because their referent does not co-vary with the quantifier in the particular quantificational binding cases tested. For instance, *now* is infelicitous in (10) but felicitous in (11), where for each artist that instantiates the quantifier, there is a different prominent result state.

(10) #Whenever I am in North Hadley, I'm happy now. (Altshuler 2017)

(11) Every artist reaches that point, where he is now ready for his masterpiece. (Altshuler & Stojnić 2015)

Saying that non-utterance time interpretations of *tomorrow* are anaphoric therefore is only a partial account, since in order to understand such readings, we also need to know what kind of object *tomorrow* can reference anaphorically. For now, I set aside this question and sketch out the predictions that all anaphoric accounts make.

First, all anaphoric accounts predict that non-utterance time readings could arise even when there is no manipulation of the context parameter. Because of this, an embedding speech verb is not predicted to be obligatory. Second, under an anaphoric account, *tomorrow* will not necessarily pattern together with other context-sensitive expressions in the same clause, since anaphoric reference is not as constrained as indexical reference. We might expect *tomorrow* to be behave similarly to *the next day*, since *the next day* is also anaphoric; however, if *tomorrow* is anaphoric to a different kind of referent than *the next day*, it may behave differently.

7 The quantification could be viewed as quantification over situations rather than times; nothing in my analysis rests on this distinction.

2.3 Evaluating analyses of non-utterance time *tomorrow*

I have outlined three analyses for non-utterance time *tomorrow*: two that appeal to context-shift, and one that is anaphoric. These accounts make different predictions about the availability of non-utterance time readings, summarized in Fig. 1.

Hypothesis	Quant. binding	UT-oriented epithets	No embedding	Attitude reports
Anaphoric	★	★	★	★
Shifty indexical		★		★
FID			★	★

Figure 1 Predicted availability of non-UT readings of *tomorrow* by analysis

Using experimental methods, we can evaluate these accounts by testing the availability of non-utterance time readings in the environments for which their predictions differ. In Sections 3 - 5, I present a series of experiments evaluating the context shift accounts. Experiments 1a and 1b establish a baseline, and Experiments 2a, 2b, and 3 test the predictions of the context-shift accounts by manipulating the environment in which *tomorrow* occurs. In Section 6, I turn to the anaphoric account and discuss evidence from post-experiment quantificational binding tasks.[8]

3 Experiments 1a and 1b: establishing a baseline

In Experiments 1a and 1b[9], I establish a baseline of acceptability for non-utterance time interpretations of *tomorrow*. I measure their acceptability with a task that avoids explicit grammaticality judgments, since non-utterance time readings may violate the usage rules for *tomorrow* that speakers have been taught. Instead of asking participants to rate the grammaticality of sentences, I ask them to rate how well a sentence fits the context depicted by a comic strip. All sentences presented to participants are grammatical, but not all of them are felicitous in the given context.

Experiment 1a uses an environment that is predicted to allow non-utterance time readings under all of the accounts discussed above: embedding under a speech verb.

Experiment 1b is a replication of Experiment 1a that controls for the possibility of a quoted interpretation of *tomorrow*.

8 All materials are publicly available through the Open Science Foundation: https://osf.io/8x9w2/.

9 The experiments are presented in the order that makes for the clearest argument; their chronological order was as follows: Experiment 1a, Experiment 2b, Experiment 3, Experiment 1b, Experiment 2a.

3.1 Methods

Data on the acceptability of non-utterance time interpretations of *tomorrow* in American English was collected through a comic-captioning task where participants rated captions for comic strips on a 7-point Likert scale (where 7 indicates high naturalness). Ratings for *tomorrow* were compared against the anaphoric expression *the next day* and a factually correct and factually incorrect baseline; captions that are accurate according to the comic strip depiction are expected to receive higher ratings than ones which are false. If speakers of American English accept non-utterance time readings of *tomorrow*, then participants should rate the *tomorrow* items higher than the false baseline and close to *the next day*.

Condition	Truth	Predicted ratings
False control	False	Low
True control	True	High
the next day	True	High
tomorrow	True if participant allows non-UT reading	High
	False otherwise	Low

Table 1 Exp. 1 predictions

3.1.1 Participants

For Experiment 1a, 126 participants were recruited through Amazon's Mechanical Turk platform. 4 participants were excluded because English was not the language of their childhood household; 50 participants were removed because their mean ratings for the good baseline condition were not at least 1 point higher than for the bad baseline.[10] This left 72 monolingual English participants residing in the US. These exclusion criteria were pre-registered through the Open Science Foundation.

For Experiment 1b, 93 participants were recruited through Amazon's Mechanical Turk platform. 13 participants were excluded for failing the baseline criteria. The remaining 80 participants were evenly balanced across experimental lists.

3.1.2 Materials

20 critical items were developed and distributed across four Latin square lists. Each list was combined with the same set of 10 fillers. Each item included a three-

10 The high rate of failure of the baseline measure is likely due to the fact that none of the training items required participants to read the day-of-week labels for the comics. A training item highlighting the day-of-week labels was added in subsequent experiments, reducing the participant removal rate.

panel comic strip and a sentence below it. The same comic strips were used for all experiments, though the text was edited between experiments.

In the first panel of each comic, two characters are shown, and one of them says that they will do something the following day. Nothing happens in the second panel, indicating that they did not follow through. In the third panel, the other character expresses frustration with the first character's lack of action (Fig. 2).

Kevin is angry because Kate said that she would water his plants { tomorrow / the next day / Friday / Saturday }.

Figure 2 Exp. 1a example stimulus

The target sentence was shown below the comic strip, and participants were asked to judge the sentence as a caption for the third panel.

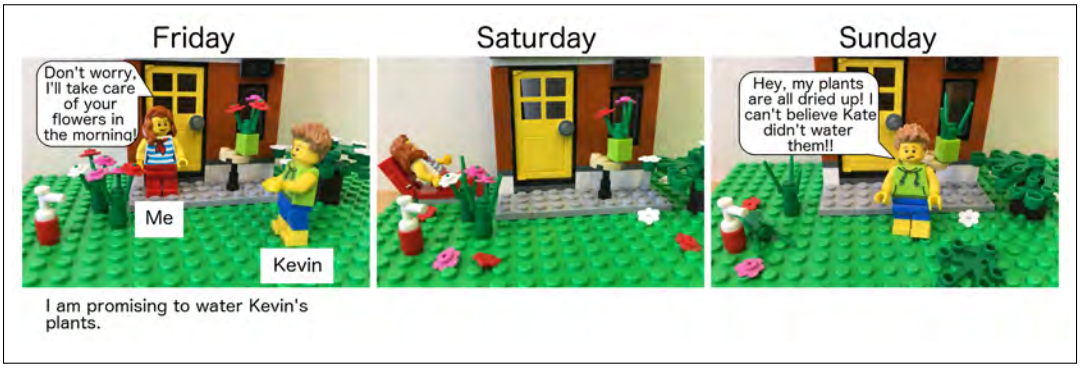
Four conditions were created by manipulating the temporal expression in the caption: *tomorrow*, the critical condition; *the next day*, the anaphoric condition; the day-of-week name of the first panel, a factually incorrect baseline; and the second day-of-week name, a factually correct baseline. Whether or not speakers allow non-utterance time readings of *tomorrow*, they are expected to interpret *tomorrow* in the first panel as referring to the second day. The *tomorrow* captions are felicitous only under a non-utterance time reading of *tomorrow*, since the first character promises to act on the second day; the other character has no grounds for anger on a UT reading.

In Experiments 1a and 1b, the target adverbials were embedded under the speech verb *say*, followed by the overt complementizer *that* to block a quotative interpretation; the verb forms in the embedded clause are also inconsistent with quotation.

In Experiment 1a, the items were as shown in Fig. 2. Experiment 1b is a replication of Exp. 1a that addresses the possibility of interpreting the critical *tomorrow* captions as mixed quotation (Maier 2014).[11] The speech bubbles were edited so that

11 Thank you to Kristen Syrett for pointing out this possibility.

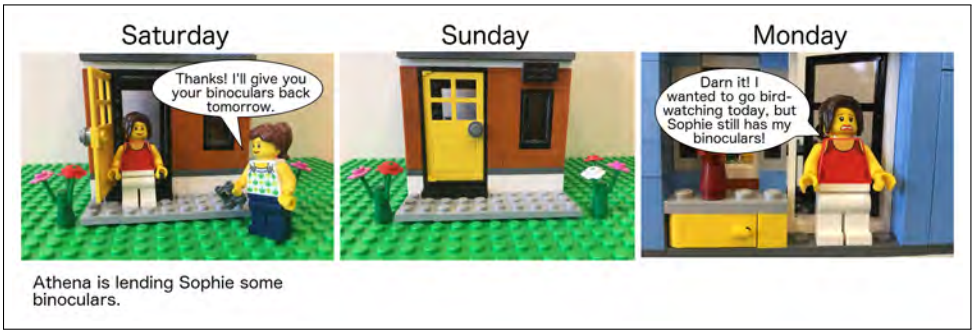
tomorrow did not appear in the comic strip itself, and the predicates were re-worded so that they were not identical to the predicates used in the critical caption (Fig. 3).



Kevin is angry because I said that I would water his plants { tomorrow / the next day / Friday / Saturday}.

Figure 3 Exp. 1b example stimulus

Three kinds of fillers were used: bad fillers, which had captions that were obviously incorrect; good fillers, which had correct captions; and pragmatically subpar medium fillers, which were factually correct, but under- or over-informative. In the filler shown in Fig. 4, for instance, the participant must infer that Athena is bored because she would have gone birdwatching had Sophie returned her binoculars.



Athena is frustrated because she is bored.

Figure 4 Example medium filler

3.1.3 Procedure

Stimuli were displayed and responses collected using the Ibex Farm platform for web-based experiments (Drummond 2013). Each experimental session began with

an informed consent form and a demographic survey.[12] Next, participants were given instructions and practiced on 3 training items: a true item, which they were told most people would rate at 7; a false item, which they were told most people would rate at 1; and a medium filler, which they were told most people would rate at 4. After Exp. 1a, an additional training item calling attention to the day-of-week labels was added.

3.1.4 Regression analysis

An analysis using paired t-tests was preregistered, but after discussion with colleagues, I decided to use a mixed effects ordinal regression model.[13] The maximal random effects structure was used for all models: random intercepts and slopes were included for all fixed-effects predictors, for participants, and for items. All models were fitted using the ordinal package in R (Christensen 2019).

Treatment coding was used, treating *tomorrow* as the baseline condition. This resulted in the following fixed-effects contrasts: *tomorrow*, 1 for *tomorrow* and 0 otherwise; *the next day*, 1 for *the next day* and 0 otherwise; false control, 1 for the false control and 0 otherwise; and true control, 1 for the true control and 0 otherwise. Treating *tomorrow* as the baseline allows us to interpret the intercepts as indicating how far the mean scores in other conditions differ from those in the *tomorrow* condition. A model that included the medium fillers as a fixed-effect predictor was also run. This comparison was not preregistered.

3.2 Exp. 1a results

The results showed that participants rated the *tomorrow* items much higher than the false baseline items, but somewhat lower than the true baseline and *the next day*.

Condition	Mean ratings	95%CI for part. means
False control	2.9	2.6-3.2
Tomorrow	5.3	5.0-5.6
The next day	6.4	6.2-6.5
True control	6.6	6.5-6.7
Bad fillers	1.1	0.9-1.3
Medium fillers	4.0	3.8-4.2
Good fillers	6.8	6.7-6.7

Table 2 Exp. 1a results

12 Demographic information was collected in order to explore whether there was any significant sociolinguistic variation by age or geographic region, but these factors were not found to be informative.

13 The comparisons of interest were the same under both analyses.

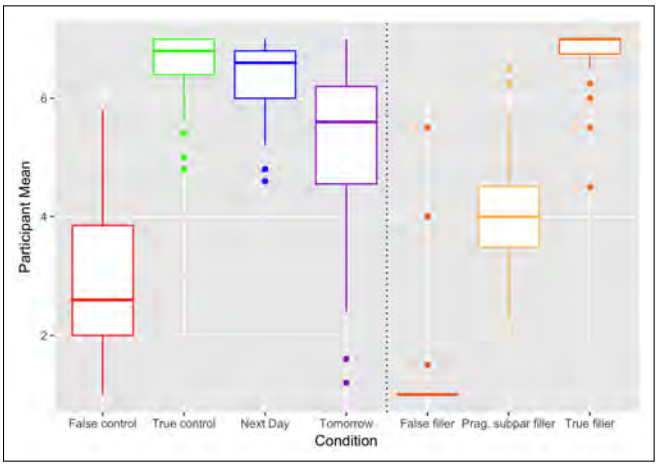


Figure 5 Exp. 1a participant means by condition

In the regression analysis, all three coefficients were reliable effects at $p < 0.0001$. This shows that the ratings in the *tomorrow* condition were significantly different from the false baseline, but also from *the next day*. Participants accepted non-utterance time readings of *tomorrow*, but found them worse than *the next day*.

	$\widehat{\beta}$	z	p
False control	-3.36(+/- 0.36)	-9.3	< 0.0001
True control	2.47(+/- 0.38)	6.45	<0.0001
next day	1.53(+/- 0.29)	5.35	<0.0001

Table 3 Exp. 1a mixed effects regression analysis, fixed effects (N=1440)

Because the *tomorrow* items were rated lower than *the next day*, it could be the case that they are not truly grammatical, but that speakers can interpret them as true through semantic coercion. However, the ratings of the pragmatically sub-optimal fillers argue against this: these items require only a small amount of accommodation to fit the context, yet participants rate the *tomorrow* items more highly than them.

A second regression model that included the pragmatically subpar fillers was run. The coefficient for the medium filler condition was significant, indicating that *tomorrow* was rated significantly higher than the medium fillers. This strengthens the claim that non-utterance time *tomorrow* is accepted.

	$\widehat{\beta}$	z	p
False control	-2.78(+/- 0.29)	-9.5	< 0.0001
True control	2.25(+/- 0.33)	6.74	<0.0001
next day	1.34(+/- 0.25)	5.45	<0.0001
Medium fillers	-1.53(+/-0.27)	-5.59	<0.0001

Table 4 Exp. 1a mixed effects regression analysis including medium fillers, fixed effects (N=1728)

3.2.1 Interspeaker variation

There is interspeaker variation in the acceptability of non-utterance time *tomorrow*. Fig. 6 shows participant means in each condition in order of increasing *tomorrow* means. While the *tomorrow* means for most participants are close to their *the next day* means, the *tomorrow* means of some participants are just as low as their bad baseline means.

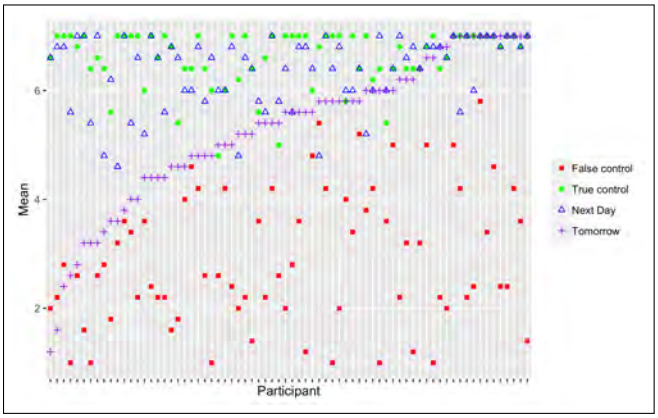


Figure 6 Exp. 1a means by participant

This suggests that there is a small group of participants who do not accept non-utterance time uses of *tomorrow*, whose low ratings for *tomorrow* items depress the overall mean for the condition.

3.3 Exp. 1b results

Experiment 1b is a replication in which the items were paraphrased to rule out partial quotation interpretations. Experiment 1b replicated the findings of Experiment 1a. The mean for the *tomorrow* condition was slightly lower, but still significantly above

that of the bad baseline and medium filler conditions. Ratings in all of the main conditions were lower in Exp. 1b compared to Exp. 1a.

Condition	Mean	95%CI
false control	2.1	[1.9-2.3]
Tomorrow	4.9	[4.5-5.2]
The next day	6.1	[5.9-6.3]
True control	6.4	[6.3-6.6]
Bad fillers	1.3	[1.1-1.5]
Medium fillers	4.2	[4.0-4.3]
Good fillers	6.7	[6.5-6.8]

Table 5 Exp. 1b results

Condition	Exp. 1a	Exp. 1b
False control	2.9	2.1
Tomorrow	5.3	4.9
The next day	6.4	6.1
True control	6.6	6.4
Bad fillers	1.1	1.3
Medium fillers	4.0	4.2
Good fillers	6.8	6.7

Table 6 Exp. 1a and 1b mean ratings

All significance tests reported in Experiment 1a replicated. The only significant difference between ratings in each condition in Experiments 1a and 1b was for the true control condition, which received lower ratings in the Exp. 1b.

Condition	Result
Exp. 1a *tomorrow* v. Exp. 1b *tomorrow*	1.3(148.7) = 0.19, p > 0.0125[14]
Exp. 1a *the next day* v. Exp. 1b *the next day*	0.8(146.2) = 0.44, p > 0.0125
Exp. 1a true control v. Exp. 1b true control	4.4(136.4) = 2.2e-05, p < 0.0001
Exp. 1a false control v. Exp. 1b false control	-0.04(149.1) = 0.96, p > 0.0125

Table 7 Paired t-tests between main conditions in Exp. 1a and 1b using z-transformed means

Since Exp. 1b replicates Exp. 1a, the ratings for the non-utterance time readings of *tomorrow* reported in Exp. 1a are not due to quotative readings of *tomorrow*.

3.4 Discussion

Experiments 1a and 1b show that participants rate non-utterance time interpretations of *tomorrow* lower than *the next day*, but well above the false control items and the pragmatically subpar fillers, establishing that non-utterance time readings of *tomorrow* are accepted by a large group of American English speakers.

Having established this baseline, I turn to evaluating the analyses outlined in Section 2 (Fig. 7). Experiment 2a and 2b explore whether non-utterance time interpretations of *tomorrow* can be explained by Free Indirect Discourse effects.

14 p-values are reported using Bonferroni-corrected thresholds to account for multiple comparisons.

Hypothesis	Quant. binding	UT-oriented epithets	No embedding	Attitude reports
Anaphoric	★	★	★	✓
Shifty indexical		★		✓
FID			★	✓

★ = prediction; *checkmark* = experimental verification

Figure 7 Attested versus predicted availability of non-UT readings of *tomorrow*

4 Experiments 2a and 2b: exploring Free Indirect Discourse as an analysis

One hypothesis about non-utterance time readings of *tomorrow* is that they arise from Free Indirect Discourse effects. Experiments 2a and 2b assess whether participants accept non-utterance time readings of *tomorrow* outside of FID environments.

Experiment 2a manipulates pronoun use to control for FID effects, under the assumption that first-person pronouns block FID (Banfield 1982). While it is true that in FID, the protagonist is usually referred to in the third-person, the exception is when the protagonist's perspective is that of the narrator at some earlier time.[15]

Since this is an interpretative possibility for the comic strip items, Experiment 2b uses narrator-oriented epithets to block FID readings. In FID environments, expressives and epithets are interpreted according to the protagonist's perspective (Eckardt 2014); including narrator-oriented epithets should block FID readings.

If the Exp. 2 *tomorrow* ratings are similar to those in Experiments 1a and 1b, this is evidence against the view that non-utterance time interpretations of *tomorrow* arise due to context shift in Free Indirect Discourse environments.

Condition	Truth	Predicted ratings
tomorrow	True if non-UT *tomorrow* allowed outside of FID	High
	False otherwise	Low

Table 8 Exp. 2 predictions

4.1 Methods

The methods used were the same as in Experiments 1a and 1b, with the exception of the critical manipulations described below and the sample size.

15 Thank you to Amy Rose Deal and Vincent Homer for discussion on this point.

4.1.1 Participants

In Experiment 2a, 52 participants were recruited on Amazon's Mechanical Turk platform. 4 participants were excluded based on the baseline criteria, leaving 48 participants balanced across experimental lists.[16] In Experiment 2b, 56 participants were recruited on the Prolific platform. 8 were excluded based on the baseline criteria, leaving 48 participants balanced across experimental lists.

4.1.2 Materials

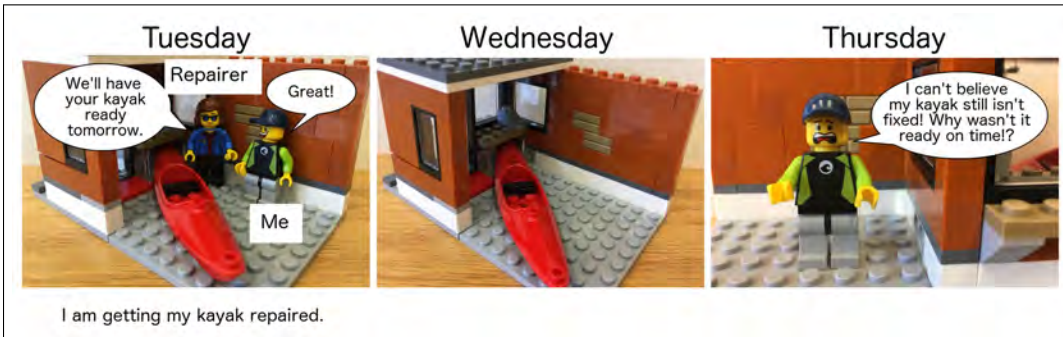
In Experiment 2a, the stimuli from Exp. 1a were modified to use first-person narration. The captions were changed to include first-person subjects, and the promise-maker was labeled as the narrator ("Me") (Fig. 8). The experimental procedure was as in Exp. 1a and 1b, except that participants were instructed that the captions represented diary entries written by the "Me" character on the day they describe.

Aidan is angry because I said that I would clean the fridge { tomorrow / the next day / Sunday / Monday }.

Figure 8 Exp. 2a example stimulus

In Experiment 2b, the stimuli were modified to include a narrator-oriented epithet. The character who was promised something was labeled as the narrator, and a narrator-oriented epithet was added to the critical captions. For example, in Fig. 9, the epithet "that liar" cannot be interpreted as reporting Kevin's Day 1 thoughts, since he has no grounds for thinking that the other character is a liar until Day 3.

16 Analysis of the effect size in Exp. 1a suggested that a smaller number of participants could be used. Large effect sizes were found for the difference between *tomorrow* and the medium fillers (Cohen's d = 0.82), and between *tomorrow* and *the next day* (Cohen's d = 0.81). If the observed effect sizes were reliable indicators of the population effect sizes, power of 0.9 could be achieved with 20 participants. Given that post-hoc effect size analyses overestimate, subsequent experiments used 48 participants.



I am frustrated because that liar at the shop said that my boat would be ready { tomorrow / the next day / Tuesday / Wednesday }.

Figure 9 Exp. 2b example stimulus

4.2 Exp. 2a results

The results of Exp. 2a were similar to Exp. 1a. Participants rated the *tomorrow* items higher than the medium fillers and false items, but lower than *the next day* (Table 9).

Condition	Mean ratings	95%CI for part. means
False control	2.2	1.9-2.4
Tomorrow	5.6	5.2-6.0
The next day	6.5	6.3-6.7
True control	6.6	6.4-6.7
False fillers	1.2	1.0-1.5
Medium fillers	3.9	3.6-4.2
True fillers	6.7	6.5-6.9

Table 9 Exp. 2a mean ratings by condition

In the mixed-effects ordinal regression model, all three coefficients were reliable effects at $p < 0.001$. Thus, ratings for the *tomorrow* condition were significantly different from both the false control condition and *the next day* condition.

	$\widehat{\beta}$	z	p
False control	-5.48 (+/- 0.56)	-9.8	< 0.0001
True control	1.78 (+/- 0.48)	3.7	<0.001
next day	1.56 (+/- 0.45)	3.5	<0.001

Table 10 Exp. 2a mixed effects regression analysis, fixed effects (N=960)

As in previous experiments, there was a small group of participants who gave *tomorrow* items low ratings.

4.2.1 Exp. 2a comparison with Exp. 1a

A mixed-effects ordinal regression model was run to compare Experiments 1a and 2a. The Experiment 2a coefficient was not significant, indicating that the *tomorrow* scores did not differ significantly between experiments.

Condition	$\widehat{\beta}$	z	p
False control	-3.48(+/-0.36)	-9.63	< 0.0001
True control	2.45(+/-0.36)	6.78	<0.0001
the next day	1.58(+/-0.30)	5.22	<0.0001
Exp. 2	0.58(+/-0.47)	1.24	0.21
False control * Exp. 2	-1.51(+/-0.53)	-2.83	0.005
True control * Exp. 2	-0.71(+/-0.52)	-1.37	0.17
the next day * Exp. 2	-0.18(+/-0.44)	-0.42	0.67

Table 11 Exp. 1a and Exp. 2a comparison mixed-effects regression analysis, fixed effects and interactions (N=2400)

The false control scores were lower in Exp. 2a, as reflected by the significant coefficient for the interaction between experiment and the false control condition.

4.3 Exp. 2b results

Experiment 2b replicated Experiment 2a using epithets rather than first-person pronouns to control for FID effects. The results of Exp. 2b were similar to previous experiments. Participants rated the *tomorrow* items lower than *the next day*, but above the pragmatically suboptimal fillers and the false control items (Table 12).

Condition	Mean ratings	95% CI for part. means
False control	2.5	[2.2;2.8]
Tomorrow	5.2	[4.7;5.7]
The next day	6.5	[6.3;6.7]
True control	6.7	[6.6;6.8]
Bad fillers	1.2	[1.0;1.3]]
Medium fillers	4.2	[3.9;4.5]
Good fillers	6.5	[6.3;6.8

Table 12 Exp. 2b results

In the mixed-effects ordinal regression model, all three coefficients were reliable effects at $p < 0.001$. Thus, ratings for the *tomorrow* condition were significantly different from both the false control condition and *the next day* condition (Table 13).

	$\widehat{\beta}$	z	p
False control	-3.9	-9.4	< 0.0001
True control	4.1	145.5	<0.0001
next day	3.4	5.8	<0.0001

Table 13 Exp. 2b mixed-effects analysis, fixed effects (N=960)

	$\widehat{\beta}$	z	p
False control	-3.4	-8.0	< 0.0001
True control	2.8	4.7	<0.0001
next day	2.5	4.5	<0.0001
Medium filler	-1.4	-3.2	<0.01

Table 14 Exp. 2b mixed-effects analysis with medium fillers, fixed effects (N=960)

The coefficient for the medium filler condition was also significant in the regression model that included the medium fillers (Table 14).

4.3.1 Exp. 2b comparison with Exp. 1a

Mean participant ratings in each condition were similar in Experiment 1a and Experiment 2b. A mixed-effects ordinal regression model was run to compare the results of Experiments 1a and 2b. The Experiment 2b coefficient was not significant, indicating that the *tomorrow* scores did not differ significantly between experiments.

Condition	$\widehat{\beta}$	z	p
False control	-3.41(+/-0.37)	-9.22	< 0.0001
True control	2.47(+/-0.41)	6.00	<0.0001
the next day	1.59(+/-0.32)	4.83	<0.0001
Exp. 2	-0.12(+/-0.48)	-0.25	0.81
False control * Exp. 2	-0.52(+/-0.53)	-0.99	0.33
True control * Exp. 2	0.51(+/-0.60)	-0.85	0.40
the next day * Exp. 2	0.75(+/-0.50)	1.48	0.14

Table 15 Exp. 1a and Exp. 2b comparison mixed-effects regression analysis, fixed effects and interactions (N=2400)

4.4 Discussion

Exp. 2a and 2b replicated Exp. 1a. There was no significant difference between the *tomorrow* ratings in either Exp. 2a or Exp. 2b when compared to Exp. 1a. Contrary to the predictions of the FID account, participants rated the *tomorrow* items with narrator-oriented epithets similarly to the Exp. 1a items. This suggests that non-utterance time interpretations of *tomorrow* are not eliminated when Free Indirect Discourse interpretations are blocked.

The experiments so far have not provided evidence for or against the other two hypotheses: the indexical shift and the anaphoric accounts. Exp. 3 tests the predictions of the indexical shift account.

Hypothesis	Quant. binding	UT-oriented epithets	No embedding	Attitude reports
Anaphoric	★	✓	★	✓
Shifty indexical		✓		✓
FID		💣	★	✓

★ = prediction; *checkmark* = experimental verification; 💣= falsified prediction

Figure 10 Attested versus predicted availability of non-UT readings of *tomorrow*

5 Experiment 3: testing the indexical shift account

The previous experiments have established that non-utterance time readings of *tomorrow* are possible under speech-verb embedding; Experiment 3 tests whether such embedding is necessary. The indexical shift account predicts that non-utterance time readings should only arise when *tomorrow* is embedded under speech / attitude

predicates. Anaphoric accounts, by contrast, predict that this should not affect the availability of non-utterance time readings if there is still a prominent referent.

If the results show that non-utterance time readings of *tomorrow* do not arise, it will be strong evidence in favor of an indexical shift account. On the other hand, if the results show that such readings arise in unembedded contexts, the anaphoric account will be the most promising, since we will have evidence that the readings arise outside of the environments for which context shift has been proposed.

5.1 Methods

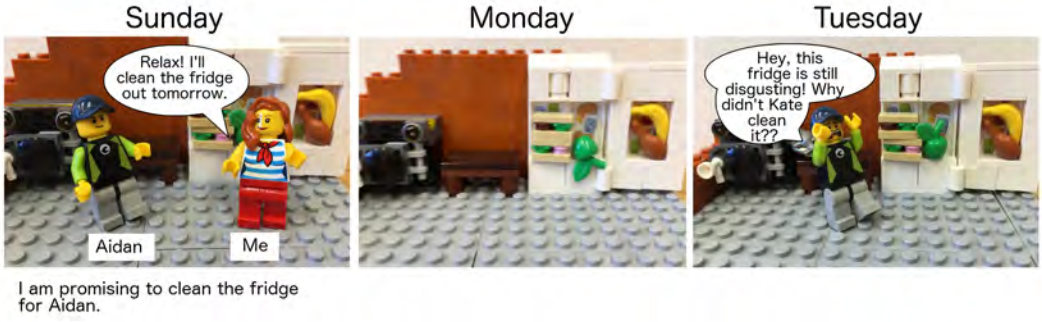
The methods for Exp. 3 were the same as in previous experiments, except for the manipulation of the critical caption described below.

5.1.1 Participants

53 participants were recruited on Amazon Mechanical Turk. 5 failed the baseline criteria and were removed, leaving 48 balanced across experimental lists.

5.1.2 Materials

In Exp. 3, the embedding speech verbs were removed from the captions. To provide a prominent previous time for anaphoric reference, the conversation between the characters is mentioned in the caption (Fig. 11).



It was such a simple task to clean the fridge { tomorrow / the next day / Friday / Saturday}! I can't believe I forgot.

Figure 11 Exp. 3 example stimulus

Under an anaphoric account, we expect similar ratings for *tomorrow* as in previous experiments, since the mention of the Day 1 conversation should provide

a prominent time different than utterance time for anaphoric reference. Under an indexical shift account, however, we expect unavailability of non-utterance time readings of *tomorrow*. Thus, if the *tomorrow* ratings are unaffected by removing the attitude predicates, it is evidence that non-utterance time interpretations of *tomorrow* do not arise solely from context shift in attitude reports.

Condition	Truth	Predicted ratings
tomorrow	False if speech embedding needed	Low
	True otherwise	High

Table 16 Exp. 3 predictions

5.2 Results

The mean ratings for *tomorrow* items were lower than in previous experiments, but still significantly higher than the false control items and the false fillers. The true controls, the false controls, the true fillers, *the next day*, and *tomorrow* were all rated lower than in previous experiments, while the medium fillers and the false fillers were rated higher than in previous experiments.

Condition	Exp. 1a	Exp. 1b	Exp. 2a	Exp. 2b	Exp. 3
False control	2.9	2.1	2.2	2.5	2.1
Tomorrow	**5.3**	**4.9**	**5.6**	**5.2**	**4.1**
The next day	6.4	6.1	6.5	6.5	5.9
True control	6.6	6.4	6.6	6.7	6.3
False fillers	1.1	1.3	1.2	1.2	1.4
Medium fillers	4.0	4.2	3.9	4.2	5.0
True fillers	6.8	6.7	6.7	6.5	6.4

Table 17 Comparison of mean ratings across experiments

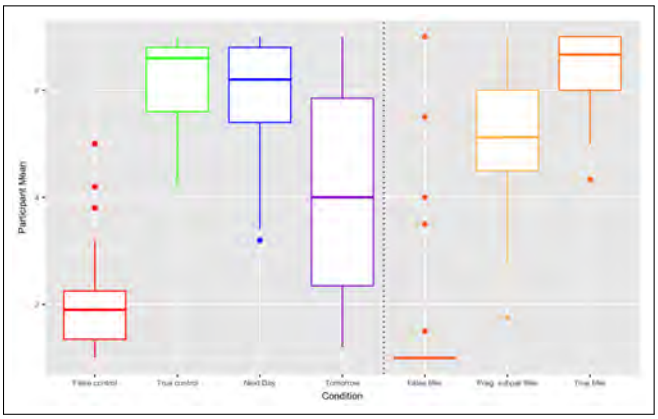


Figure 12 Exp. 3 main condition ratings

In the mixed-effects ordinal regression model, all three coefficients were reliable effects at $p < 0.0001$. Thus, despite the lower *tomorrow* scores in this experiment, the *tomorrow* condition was still significantly different than the false control condition.

	$\widehat{\beta}$	z	p
False control	-3.41(+/- 0.50)	-6.81	< 0.0001
True control	3.20(+/- 0.48)	6.63	<0.0001
next day	2.54(+/- 0.50)	5.08	<0.0001

Table 18 Exp. 3 mixed effects regression analysis, fixed effects (N=960)

In the model that included the medium fillers, however, the coefficient for the medium fillers was not significant at $p = 0.0125$, indicating that the difference between the *tomorrow* condition and the pragmatically subpar fillers was not statistically significant.

	$\widehat{\beta}$	z	p
False control	-3.16(+/- 0.46)	-6.84	< 0.0001
True control	3.03(+/- 0.45)	6.69	<0.0001
next day	2.39(+/- 0.47)	5.08	<0.0001
Medium fillers	0.98(+/-0.45)	2.16	0.03

Table 19 Exp. 3 mixed effects regression analysis with medium fillers, fixed effects (N=960)

5.2.1 Comparison with previous experiments

Although the main comparisons in Exp. 3 were similar to Exp. 1a, a mixed-effects ordinal regression model comparing Exp. 1a and 3 finds significant interactions between experiment and *tomorrow* and between experiment and *the next day*, indicating that the differences between the Exp. 1a and Exp. 3 ratings for the *tomorrow* and *the next day* conditions were significant.

Condition	$\widehat{\beta}$	z	p
False control	-3.44(+/-0.36)	-9.44	< 0.0001
True control	2.31(+/-0.36)	6.47	<0.0001
the next day	1.50(+/-0.33)	4.52	<0.0001
Exp. 2	-1.72(+/-0.47)	-3.6	<0.001
False control * Exp. 2	0.35(+/-0.53)	0.65	0.52
True control * Exp. 2	0.87(+/-0.52)	1.68	0.09
the next day * Exp. 2	0.97(+/-0.47)	2.06	0.039

Table 20 Exp. 1a and Exp. 3 comparison mixed-effects regression analysis, fixed effects and interactions (N=2400)

The 95% confidence intervals for participant means were wider than in previous experiments; although the *tomorrow* interval is widest in all experiments, in Exp. 3 it was wider than 1 Likert scale point.

Condition	Mean rating	95%CI for part. means
False control	2.1	1.8-2.3
Tomorrow	4.1	3.5-4.6
The next day	5.9	5.6-6.2

Table 21 Exp. 3 results

5.2.2 Interspeaker variation

There was a higher amount of interspeaker variation in *tomorrow* ratings in this experiment. While in all experiments there was a group of participants who gave *tomorrow* items consistently low ratings, the distribution of *tomorrow* scores in Exp. 3 is almost bimodal. In Fig. 13, which plots *tomorrow* ratings for each item by participant, we see two distinct groups: one which consistently rates *tomorrow* near *the next day*, and one which rates it near the bad baseline. Both groups occasionally

produce ratings at the opposite end of the scale (indicated by a long line connecting the participant's lowest and highest ratings).

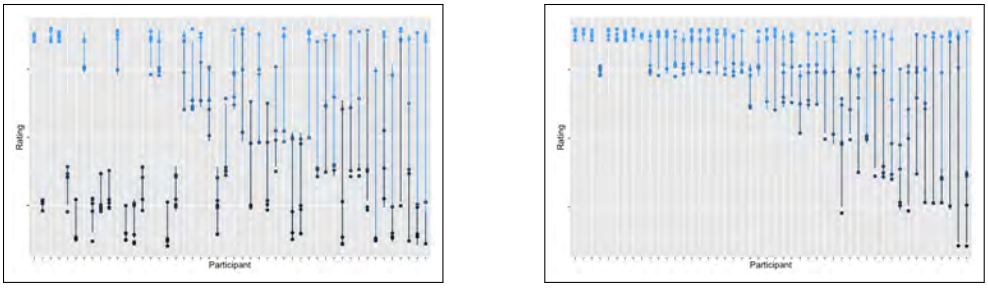


Figure 13 Exp. 3 *tomorrow* ratings (left) and *the next day* ratings (right) by participant, in order of difference between highest and lowest rating

The question is whether the participants who give *tomorrow* low ratings in this experiment are the same as those in previous experiments. It may be that Exp. 3 sampled more heavily from this population; or it may be that removing the embedding verb has a real effect on the acceptability, and that these participants would have accepted the *tomorrow* items in Exps. 1 and 2.

5.3 Discussion

Ratings for *tomorrow* items were lower in Experiment 3 than in previous experiments. They were not significantly different than the ratings for the medium fillers, though they were significantly above the bad baseline. In addition, there was more interspeaker variation in *tomorrow* ratings than in previous experiments.

Hypothesis	Quant. binding	UT-oriented epithets	No embedding	Attitude reports
Anaphoric	★	✓	✓	✓
Shifty indexical		✓	💣	✓
FID		💣	✓	✓

★ = prediction; *checkmark* = experimental verification; 💣= falsified prediction

Figure 14 Attested versus predicted availability of non-UT readings of *tomorrow*

Although a larger group of participants rejected non-utterance time *tomorrow* in Experiment 3 compared to other experiments, another substantial group gave the *tomorrow* items high ratings. The existence of this second group is inconsistent

with the indexical shift account, since it predicts that non-utterance time *tomorrow* should never occur outside of embedding. In the rest of this paper, I proceed with an analysis of *tomorrow* for the group of speakers who accept *tomorrow* in Exp. 3. Whether the difference between these two groups stems from different lexical entries for *tomorrow* or from processing differences remains a topic for future work.

6 A perspective anaphoric account of *tomorrow*

Experiments 2a, 2b, and 3 test the predictions of two context-shift accounts: the Free Indirect Discourse account and the indexical shift account. Experiment 2b showed that contrary to the FID account, non-utterance time interpretations of *tomorrow* occur alongside narrator-oriented epithets. Experiment 3 showed that unlike shifted interpretations of shifty indexicals, non-utterance time readings of *tomorrow* are available outside of embedded contexts for many speakers. Taken together, these experiments establish the availability of non-utterance time readings of *tomorrow* outside the environments in which context shift has been proposed to occur.

Under context shift accounts, *tomorrow* could be a pure indexical; the evidence that non-utterance time readings occur outside of documented context shift environments, though not eliminating a context shift account if context shift is found to occur more widely than currently thought, lends support to an anaphoric view.

Although the intuitive anaphoric account for a temporal adverbial is that it is anaphoric to any prominent time, I develop a different account for *tomorrow*: that it is anaphoric to the time index on a prominent perspective. This may not be as intuitive as a simple time anaphoric account, but previous work on temporal adverbials has shown that the most obvious anaphoric account is not always the correct one (Stojnić & Altshuler 2019); in addition, the fact that *tomorrow* receives lower ratings than *the next day* suggests that they are not anaphoric to exactly the same objects.

My proposal does not rest on these points alone, however: in Section 7, I present evidence from quantificational binding that *tomorrow* cannot be anaphoric to just any prominent time; and that perspectival contexts heighten the acceptability of *tomorrow* in quantificational binding contexts.

6.1 Perspectival anaphoric reference

In proposing that *tomorrow* is anaphoric to prominent perspectives, I draw upon Barlew (2017)'s analysis of *come* as anaphoric to prominent perspectives. The verb *come* describes motion relative to the location of a perspective-holder at event time or utterance time. As shown below, the destination of motion can be the speaker's location at utterance time (12); the speaker's location at event time (13);

the addressee's location at utterance time or event time; or the location of an attitude-holder (14).

(12) Come here, please!

(13) Context: Speaker is currently in her office.
When I got to the bar it was empty, but by the time Mark came, it was full.

(14) Susan said to come there when we're done.

In Barlew (2017)'s account, *come* carries a presupposition that a prominent perspective-holder is located at the destination of motion. He argues that an indexical analysis of *come* (Oshima 2006) is inconsistent with the fact that *come* can appear in quantificational binding contexts, so long as the perspective co-varies with the quantifier. In (15), each instantiation of the woman provides a prominent perspective as an anchor for *come*; *was glad* heightens the prominence of each woman's perspective.

(15) Every woman was glad that her wayward child came to Christmas dinner. (Barlew 2017)

Barlew (2017) follows Roberts (2015) in proposing that perspectives, which he treats as sets of centered worlds, are tracked in the Common Ground. In his system, speaker and listener perspectives are always available, and additional perspectives, like those of attitude-holders, can be introduced during the discourse (Roberts 2015).

6.2 Perspectives track temporal information

The perspectival view of deictic motion verbs outlined above requires that perspectives encode temporal as well as spatial information. This is needed in order to account for the difference in event time and utterance time licensing of *come* shown in (12) and (13). One way to distinguish utterance time and event time perspectives would be to model each perspective as a tuple containing a set of centered worlds and a time index. To illustrate the semantics of *come* below, I include a space index as well.[17] The event-time versus utterance-time licensing of *come* can then be explained by optionality in whether the time index of the perspective matches the event time or the time index of the context parameter (utterance time).[18]

17 The inclusion of both a time and space index is a matter of notational convenience; we can instead omit the space index and require that in each world, the perspective-holder self-locates at location l at the given time.

18 I elide the careful consideration of motion paths in Barlew (2017), as *come* is not my focus.

(16) $[[\text{come}]]^{c,g} = \lambda e.\exists p.\text{MOVE}(e) \wedge \text{DEST}(e, p_s) \wedge (\tau(e) \subset p_t \vee c_t \subset p_t)$
where p is a prominent perspective in the Common Ground with time index p_t and space index p_s and c_t is the time index of the context parameter c

If there were no way of obtaining the temporal location of the perspective-holder from the perspective, then we would predict that *come* could be licensed by the self-location of a perspective-holder at the destination at any time, not just event-time or utterance-time. This is not the case; having once visited a location does not allow you to forever describe motion to that place using *come*.

I adopt this notation for perspectives in what follows; however, my proposal does not rest on this way of formalizing perspective, so long as there is a linguistic object representing a perspective that is time-indexed.

6.3 Towards a perspectival account of *tomorrow*

Given that perspectives must carry temporal information, I propose that the temporal adverbial *tomorrow* makes use of this information. I posit that, like the motion verb *come*, *tomorrow* is anaphoric to perspectives: it receives non-utterance time interpretations when it is anchored to a non-utterance time perspective, and utterance time interpretations when it is anchored to an utterance time perspective. In particular, *tomorrow* takes the time index of a prominent perspective as its reference time.

Formally, the lexical semantics of *tomorrow* take an event and anchor it temporally to the day after the time index of a perspective in the Common Ground.

(17) Perspectival semantics for *tomorrow*:
$[[\text{tomorrow}]]^{c,g} = \lambda Q.\lambda e.\exists p.\tau(e) \subset \iota t.\text{DAY-AFTER}(t, p_t) \wedge Q(e)$
where p is a prominent perspective and p_t is its temporal index

I will not propose a specific way for the perspective variable shown in (17) to receive its value. My proposal is consistent with several methods of valuing this perspectival variable, including Barlew (2017)'s dynamic semantics approach in which perspectives are tracked in the Common Ground; Sudo (2018)'s proposal that there is a perspective parameter in addition to the context parameter; or an account where it is bound by an abstract perspectival head. What my proposal for *tomorrow* minimally requires is some linguistic object that represents a perspective, including its temporal information, to which *tomorrow* can refer anaphorically.

The perspective-anaphoric semantics for *tomorrow* captures the pattern of acceptability in the experimental data. In Experiments 1a, 1b, 2a, and 2b, a prominent non-utterance time perspective is provided by the subject of the attitude verb. In Experiment 3, the prominent perspective is that of the speaker, since there is no attitude-holder. The phrases like 'such a simple task!' heighten the saliency of the

speaker's event-time perspective by focusing on the event-time evaluation of the task. The lower ratings for *tomorrow* in this experiment might be accounted for if event-time perspectives are less salient than attitude-verb introduced perspectives; or if there is interspeaker variation in the accessibility of event-time perspectives.

In addition to accounting for the experimental data, the perspectival view of *tomorrow* explains why *tomorrow* behaves like a pure indexical in so many contexts. Because the most prominent perspective is generally that of the speaker at utterance-time, the time index of the context parameter and the time index of the prominent perspective are often identical. It is only in cases where the perspective is an event-time perspective, or where the perspective-holder is mistaken about their temporal location, that temporal perspectival anaphora will behave differently from indexicals.

The overlap in licensing conditions between indexicals and perspectival expressions makes it difficult to tell the two apart. In the majority of cases the most prominent perspective encodes identical information as the context parameter; and many of the environments in which other perspectives are made prominent, such as under attitude predicate embedding, are also environments in which context shift occurs. In order to distinguish perspectival expressions from (shiftable) indexicals, it is necessary to create environments that block context shift but make a non-utterance time perspective prominent, as in Experiments 2-3; or test quantificational binding contexts. I turn to evidence from quantificational binding in the next section.

7 Quantificational binding

Pure indexicals are infelicitous in quantificational binding contexts, because their referent is fixed to the utterance context and cannot co-vary with the quantifier. Anaphoric expressions, on the other hand, can appear in quantificational binding contexts so long as there is a prominent referent that co-varies with the quantifier. The anaphoric expression *the next day* is felicitous in (18), because the quantification over time provides a prominent referent for each instantiation of the quantifier.

(18) [[Whenever I wash my car, it rains the next day]]^{c,g} =
$\forall ts.t$.WASH(my car, I) –> $\exists e.\tau(e) \subset \iota t'$.DAY-AFTER$(t',t) \wedge$RAIN$(e)$

Under a pure indexical analysis of *tomorrow*, *tomorrow* should be infelicitous in these environments, since the time index of the context parameter does not vary with the quantifier. A pure indexical semantics for *tomorrow* (19), would mean that whenever the speaker washes their car, it rains the day after the sentence is spoken.

(19) [[Whenever I wash my car, it rains tomorrow]]^{c,g} =
$\forall t$ s.t. WASH(my car, I) –> $\exists e.\tau(e) \subset \iota t'$.DAY-AFTER$(t',c_t) \wedge$RAIN$(e)$

The perspectival account of *tomorrow* that I have sketched makes a different prediction about quantificational binding than either the pure indexical semantics shown in (19) or the plain time anaphoric semantics for *the next day* shown in (18). If *tomorrow* is anaphoric to perspectives, then it should be felicitous in quantificational binding contexts when the prominent perspectives co-vary with the quantifier.

A perspectival account of *tomorrow* predicts felicity for an adapted form of Barlew's quantificational binding example for *come* (15), repeated as (20).

(20) Every woman was glad that her wayward child came to Christmas dinner. (Barlew 2017)

(21) On Christmas Eve, every little girl stays awake for hours wondering what she will find under the Christmas tree tomorrow morning.

Example (21) should be felicitous since each girl provides a perspective for *tomorrow*; the prominence of each of their perspectives is heightened by *wonder*.

Another way that perspectives can co-vary with quantification is when the quantification is over times and each time provides a prominent event time perspective. In (22) and (23), the expressive content should increase the saliency of the event time perspectives of the listener and speaker, respectively.

(22) Every time you have to kick a drunk idiot out of the bar, you get to gloat about how hungover the jerk will be tomorrow.

(23) My coworker is such a brat. Every time the jerk thinks it'll be sunny tomorrow, he calls in 'sick' and I have to cover his shift.

The perspectival view of *tomorrow* predicts that *tomorrow* should be felicitous in some but not all quantificational binding environments. In particular, if the acceptability of *tomorrow* in quantificational binding contexts is improved when the perspectives that co-vary are made more prominent, as in the examples given above, this will be evidence in support of the perspectival view that I have sketched.

7.1 Quantificational Binding Task 1: time and speech context quantification

A post-experiment debriefing task on quantificational binding was included following Experiment 2a.[19] The task probed the acceptability of *tomorrow* and *the next day* in two quantificational contexts: quantification over times and over speech contexts (Fig. 15). Participants rated 2 time quantification and 4 speech-context quantification items for each adverbial, for a total of 12 items, presented in a Latin square design.

19 A less extensive task quantificational binding task was included after Exp. 1a. The results were similar and are therefore omitted here for brevity.

Time quantification: Every time I wash my car, it rains { tomorrow / the next day}.

Speech context quantification: Every time the UPS person says that the package has been delivered, it doesn't show up until {tomorrow / the next day }.

Figure 15 Quant. Binding Task 1 example stimuli

7.1.1 Results

Quantificationally bound *tomorrow* items received low ratings in all conditions. An ANOVA of the *tomorrow* items shows no significant difference by condition (1.77(1,382)= 0.18; p > 0.05).

Adverbial	Time quant. mean [95% CI]	Speech quant. mean [95% CI]
tomorrow	2.4 [2.0-2.9]	2.7 [2.4-3.0]
the next day	6.2 [5.9-6.5]	5.8 [5.6-6.1]

Table 22 Exp. 2a binding task results

The fact that *tomorrow* receives very low ratings in quantificational binding contexts where there is a prominent time available at each instantiation of the quantifier suggests that *tomorrow*, unlike *the next day*, is not anaphoric to just any prominent time in the preceding discourse.

Hypothesis	Quant. binding	UT-oriented epithets	No embedding	Attitude reports
Time-anaphoric	💣	✓	✓	✓
Shifty indexical	X	✓	💣	✓
FID	X	💣	✓	✓

★ = prediction; *checkmark* = experimental verification; 💣= falsified prediction

Figure 16 Attested versus predicted availability of non-UT readings of *tomorrow*

7.2 Quantificational Binding Task 2: perspectival contexts

A preliminary investigation of the quantificational binding cases predicted to be felicitous by the perspectival account was made in second quantificational binding task. This task was included after the main task in Experiment 2b.

(24) I try to start winding down by 9pm so that I have time to tidy up the kitchen, think through what I'm going to wear tomorrow, and generally get my things together. That way I can sleep in as long as possible and still get to work on time!

(25) One of my professors wears a different hat to work each day. He even has a website where you can pick a hat for him to wear tomorrow.

(26) I set up an app on my phone that notifies me every time tomorrow's forecast calls for snow.

Figure 17 Quantificational Binding Task 2 naturally occurring *tomorrow* items

4 kinds of quantificational binding items were included: 3 *tomorrow* items from the previous quantificational binding task, expected to receive low ratings on the basis of previous results; 3 *the next day* items from the previous task, expected to receive high ratings; 3 naturally-occurring *tomorrow* items collected from corpora (Fig. 17); and 3 examples constructed to test the perspectival predictions (Fig. 18). Participants rated these 12 items after the main task in Experiment 2b.

(27) On Christmas Eve, every little girl stays awake for hours wondering what she will find under the Christmas tree tomorrow morning.

(28) Every time you have to kick a drunk idiot out of the bar, you get to gloat about how hungover the jerk will be tomorrow.

(29) My coworker is such a brat. Every time the jerk thinks it'll be sunny tomorrow, he calls in "sick" and I have to cover his shift.

Figure 18 Quantificational Binding Task 2 perspectival *tomorrow* items

7.2.1 Results from Quantificational Binding Task 2

The results of this task differ from those of Quantificational Binding Task 1. In this task, the bound *tomorrow* items did not uniformly receive low ratings. Although the *tomorrow* items presented in the previous task, labeled 'prev-*tomorrow*' in Fig. 23, again received low ratings (although not as low as in the previous task), the two new conditions, 'nat-*tomorrow*', the naturally-occurring corpus examples, and 'persp-*tomorrow*', the constructed perspectival items, received higher ratings.

Differences among ratings for quantificational binding items were investigated using a cumulative link mixed effect model. The *the next day* items were treated

as the baseline, and three fixed effects were included: tomorrow, coded as 1 if the adverbial used was *tomorrow* and 0 otherwise; perspectival, coded as 1 if the item was a perspectival item and 0 otherwise; and natural, coded as 1 if the item was a naturally-occurring quantificational binding instance and 0 otherwise.

Condition	Mean rating	95%CI for part. means
Prev-*the next day*	5.6	[5.2;6.0]
Prev-*tomorrow*	3.7	[3.2;4.1]
Persp-*tomorrow*	4.8	[4.4;5.3]
Nat-*tomorrow*	5.2	[4.8;5.6]

Table 23 Quantificational Binding Task 2 results

All three fixed effects were reliable at $p < 0.0001$ (Table 7.2.1), indicating that while the use of *tomorrow* decreases participants' ratings, the ratings for the perspectival and natural conditions have a significant positive effect on participants' ratings compared to the *tomorrow* items from Quantificational Binding Task 1.

	$\widehat{\beta}$	z	p
Perspectival	1.2(+/- 0.25)	5.02	< 0.0001
Natural	1.7(+/- 0.37)	6.79	<0.0001
Tomorrow	-2.3(+/- 0.26)	-6.05	<0.0001

Table 24 Quant. Binding Task 2 mixed effects regression analysis, fixed effects (N=576)

In addition, there was a moderate negative correlation between participants' perspectival *tomorrow* scores and the difference between their mean *tomorrow* and *the next day* scores in the main task ($r=-0.45$; $p<0.01$). The closer that participants rated *tomorrow* and *the next day*, the higher they rated perspectival *tomorrow*; this is expected if acceptance of *tomorrow* in the perspectival quantificational binding task is contingent on acceptance of non-utterance time readings of *tomorrow*.

The higher ratings for the *tomorrow* items in the perspectival condition support the perspectival view of *tomorrow* proposed above. Instances of bound *tomorrow* in which there are prominent perspectives that co-vary with the quantifier, whether the perspective-holder herself varies, as in the Christmas example, or the time-index on the perspective varies, as in the expressive examples, are rated higher than instances of bound *tomorrow* under ordinary time or speech situation quantification.

The results also provide further evidence against a pure indexical view of *tomorrow*, which predicts infelicity in all quantificational binding environments.

Hypothesis	Time quant.	Persp. quant.	UT-oriented epithets	No embedding	Attitude reports
Time-anaph.	💣	✓	✓	✓	✓
Persp.-anaph.	X	✓	✓	✓	✓
Shifty indexical	X	💣	✓	💣	✓
FID	X	💣	💣	✓	✓

★ = prediction; *checkmark* = experimental verification; 💣= falsified prediction

Figure 19 Attested versus predicted availability of non-UT readings of *tomorrow*

The higher ratings of the naturally-occurring *tomorrow* items are more difficult to explain, since two of the three seem to involve quantification over times similar to the items from Quantificational Binding Task 1, and the other seems to involve quantification over thought situations (24). This example also has a habitual flavor, and it may be that modality is playing a role in the relatively high ratings.

7.3 Analysis

Experiments 1-3 established that a substantial group of American English speakers (though not all) allow non-utterance time readings of *tomorrow* outside of the environments in which context shift has been proposed to occur. Results from two quantificational binding tasks further suggests that *tomorrow* is not anaphoric to simple times, but that it is felicitous in quantificational binding contexts in where there are perspectives that co-vary with the quantifier. In light of this data, I have proposed that *tomorrow* refers anaphorically to the time index of a perspective.

8 Conclusion

There is an emerging consensus that many expressions thought to depend purely on the context parameter for their meaning do not (Stojnić & Altshuler 2019, Maier 2017). I have argued that for a substantial group of speakers, *tomorrow* can be used anaphorically and therefore is not a pure indexical. In a series of experiments, I have shown that non-utterance time readings of *tomorrow* are not limited to context shift environments such as Free Indirect Discourse or attitude reports, and that *tomorrow* can appear in some, but not all, quantificational binding environments.

I have proposed that *tomorrow* is anaphoric to the time index of a perspective, and that non-utterance time readings arise whenever there is a mismatch between utterance time and this time index. I have also shown that one prediction of this account is borne out by the data: the felicity of *tomorrow* in quantificational binding contexts improves when there are perspectives that co-vary with the quantifier.

Although I have presented experimental evidence in favor of a perspective-anaphoric view of *tomorrow*, open questions remain. The nature of the interspeaker variation observed across experiments merits further study, as does the question of whether other temporal adverbials, such as *yesterday*, behave similarly.

This work highlights the difficulty in differentiating between perspective shift and indexical shift. The environments in which perspectival items receive non-utterance-context interpretations overlap to a great extent with those in which indexical shift occurs. When indexicals participate in context shift, particular care must be taken to test the environments in which perspective shift, but not context shift, is predicted.

The non-utterance time readings of *tomorrow* discussed here add to the growing list of expressions that are sensitive to perspective. In addition, they reveal a tidy parallel between perspectival items in the spatial domain, like *come*, and perspectival items in the temporal domain, like *tomorrow*. Investigating seemingly exceptional uses of context-sensitive expressions like *tomorrow* can shed light on the multiplicity of strategies for context-sensitive reference in the spatiotemporal domain.

References

Abrusán, Márta. 2018. Protagonist projection. *Semantics and Linguistic Theory* 28.

Altshuler, Daniel. 2017. Tense and temporal adverbs ("I learned last week that there would now be an earthquake"). In Lisa Matthewson, Cécile Meier, Hotze Rullman & Thomas Ede Zimmermann (eds.), *The Blackwell Companion to Semantics*, Wiley.

Altshuler, Daniel & Una Stojnić. 2015. The attention-coherence model of prominence: A look at 'now'. Paper presented at the International Conference on Prominence in Language.

Anand, Pranav & Andrew Nevins. 2014. Shifty operators in changing contexts. *Semantics and Linguistic Theory* 24.

Banfield, Ann. 1982. *Unspeakable sentences: Narration and representation in the language of fiction*. Routledge.

Barlew, Jefferson. 2017. *The semantics and pragmatics of perspectival expressions in english and bulu: The case of deictic motion verbs*. Doctoral Dissertation, Ohio State University.

Christensen, R. H. B. 2019. Ordinal: Regression models for ordinal data. R package version 2019.3-9. http://www.cran.r-project.org/package=ordinal/.

Deal, Amy Rose. 2014. Nez Perce embedded indexicals. In Hannah Greene (ed.), *Proceedings of semantics of under-represented languages of the americas*, vol. 7, University of Massachusetts, Amherst Graduate Linguistics Student Association.

Deal, Amy Rose. 2017. Shifty asymmetries: universals and variation in shifty indexicality. Ms.

Doron, Edit. 1991. Point of view as a factor of content. *Semantics and Linguistic Theory* 1.
Drummond, Alex. 2013. Ibex Farm. http://spellout.net/ibexfarm.
Eckardt, Regina. 2014. *The semantics of free indirect discourse: how texts allow us to mind-read and eavesdrop*.
Harris, Jesse & Christopher Potts. 2009. Perspective-shifting with appositives and expressives. *Linguistics & Philosophy* 32(6).
Kamp, Hans. 1971. Formal properties of 'now'. *Theoria* 37(3).
Kaplan, David. 1989. *Demonstratives*.
Maier, Emar. 2014. Mixed quotation: the grammar of apparently transparent opacity. *Semantics & Pragmatics* 7(7).
Maier, Emar. 2015. Quotation and unquotation in free indirect discourse. *Mind & Language* 30(3).
Maier, Emar. 2017. A good indexical is hard to find. Paper presented at Anaphora Resolution in Sign and Spoken Languages.
Oshima, David. 2006. Motion deixis, indexicality, and presupposition. *Semantics and Linguistic Theory* 16.
Park, Yangsook. 2018. *Attitudes de se and logophoricity*. Doctoral Dissertation, University of Massachusetts, Amherst.
Quer, Joseph. 2005. Context shift and indexical variables in sign language. *Semantics and Linguistic Theory* 15.
Roberts, Craige. 2015. Indexicality: de se semantics and pragmatics. Ms.
Schlenker, Philippe. 2003. A plea for monsters. *Linguistics and Philosophy* 26.
Schlenker, Philippe. 2004. Context of thought and context of utterance. a note on free indirect discourse and the historical present. *Mind & Language* 19(3).
Sharvit, Yael. 2008. The puzzle of free indirect discourse. *Linguistics and Philosophy* 31.
Stojnić, Una. 2016. *Context-sensitivity in a coherent discourse*. Doctoral Dissertation, Rutgers University.
Stojnić, Una & Daniel Altshuler. 2019. Formal properties of 'now' revisited. Ms, Columbia University and Hampshire College.
Stokke, Andreas. 2013. Protagonist projection. *Mind & Language* 28(2). 204–232.
Sudo, Yasutada. 2018. Come vs. Go and Perspective Shift. Ms.

Carolyn Jane Anderson
650 North Pleasant Street
Amherst, MA 01003
carolynander@umass.edu