Tomorrow isn't always a day away[1]

Carolyn Jane ANDERSON — *University of Massachusetts, Amherst*

Abstract. This paper explores non-utterance time readings of *tomorrow*, which, though unexpected under the standard pure indexical view of *tomorrow*, are attested in American English: e.g., "Last week, UPS said that the package would be delivered tomorrow." This example has two readings: an utterance time reading, which can be felicitously followed by "I hope it arrives on time!", and a non-utterance time reading, which can be followed by "But it never showed up!" I present experimental evidence that (1) non-utterance time readings of *tomorrow* are acceptable for many American English speakers and (2) not due to Free Indirect Discourse or indexical shift. Instead, I propose an analysis of *tomorrow* as anaphoric to a salient perspective.

Keywords: temporal adverbials, context sensitivity, experimental semantics, perspective.

1. Introduction

Indexicality and anaphoricity are two types of context sensitivity. Indexicals refer relative to the context of utterance, while anaphoric reference is relative to discourse-given referents. While some expressions are purely indexical or purely anaphoric, a growing number of expressions have been found to have both anaphoric and indexical uses. In some cases, anaphoric uses have been miscategorized as indexical uses because the referent is something complex, like a result state in the case of the temporal indexical *now* (Altshuler and Stojnić, 2015).

One case where anaphoric reference is particularly difficult to distinguish from indexicality is perspective-anaphoric expressions. For instance, although *come* has been analyzed as indexical (Oshima, 2006), Barlew (2017) has argued convincingly that it is in fact anaphoric to salient perspectives. Its anaphoricity had been overlooked because the speaker and addressee's utterance-time perspectives are always salient, leading to an indexicality illusion.

I argue that *tomorrow*, long thought to be a pure indexical (Kamp, 1971; Kaplan, 1989), is in fact anaphoric to perspectives. Because the salient perspective is so often the speaker's, and the speaker's perspective is so often the same as the utterance context, *tomorrow*'s perspectival nature has been obscured by its identical behavior to pure indexicals in most environments. I explore the interpretation of *tomorrow* in environments in which the salient 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 (non-UT) time readings of *tomorrow*, as in (1).

(1)   Two weeks ago, Jane said that the package would arrive tomorrow, but it never came.

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 ungrammatical, they are easily found in corpus data, and I present experimental evidence that they are accepted by many American English speakers.

The existence of non-UT readings does not disprove that *tomorrow* is a pure indexical, since non-UT indexical readings can arise if the context parameter has been manipulated, such as in

Free Indirect Discourse (FID) or indexical shift. In order to investigate non-UT interpretations of *tomorrow*, I present a series of studies that narrow the hypothesis space for such readings, considering three main possibilities: that non-UT readings of *tomorrow* arise from FID effects; that they are like indexical shift reported in other languages (Schlenker, 2003); or that they are anaphoric.

I establish a baseline of acceptability for non-UT readings through a sentence acceptability judgment task, and then test the predictions that the FID and indexical shift accounts make about pronoun use and embedding under attitude verbs. On the basis of the results, I argue that non-UT readings are not indexical; however, on the basis of quantificational binding data, I also argue that *tomorrow* is not anaphoric to just any time. I posit instead that *tomorrow* can be anaphoric to salient perspectives, and that this behavior has been overlooked because the salient perspective is almost always identical to the context of utterance.

This suggests 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-UT 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; it is used *anaphorically* when it refers relative to a reference point given by the discourse context. In addition, Kaplan (1989) defines a class of *pure indexicals*, expressions whose interpretation is always indexical.

In Kaplan (1989), *tomorrow* is described a pure indexical, which makes non-UT interpretations surprising unless they occur only when the context parameter has been manipulated in some way. Alternatively, these uses of *tomorrow* may be anaphoric, in which case *tomorrow* is not a pure indexical.[2] We can distinguish two types of analyses for *tomorrow*: indexical accounts, which maintain the pure indexical status of *tomorrow*, and anaphoric accounts, which posit that non-UT readings arise through anaphoric reference.

I explore two analyses that could maintain the pure indexical status of *tomorrow*: a Free Indirect Discourse account, and an indexical shift account.

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

Temporal adverbials like *tomorrow* are known to have non-UT interpretations in Free Indirect Discourse, 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 like deictic motion verbs are used relative to the protagonist's perspective. In FID, temporal adverbials refer relative to the protagonist's *now*, which is event time (ET), rather than the narrator's, which is utterance time (UT).

---

[2]There are views of modeling context in which anaphoric expressions are treated as pure indexicals (Stojnić, 2016).

Most analyses of FID maintain the pure indexical status of protagonist-oriented indexicals. Rather than changing the lexical semantics of such expressions, they propose a modification of Kaplanian contexts. For instance, Eckardt (2014), building on Doron (1991)[3] proposes that there are 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. In direct speech, only the external context is available, but FID makes an internal context available. When there are two context parameters, shiftable indexicals refer relative to the internal context, while rigid indexicals remain fixed to the external context.

This account posits that non-UT readings arise when *tomorrow* is interpreted relative to an internal context. The FID account, and others that use multiple context parameters or context-overwriting, is consistent with a pure indexical view of *tomorrow*, since even when receiving protagonist-oriented interpretation, *tomorrow* would refer to a context parameter.

A FID view of non-UT *tomorrow* makes several predictions. First, we would expect that other items that are protagonist-oriented in FID contexts would receive non-UT interpretations in the same contexts as *tomorrow*. We would expect uniform behavior across temporal and locative indexicals and expressives in contexts where non-UT interpretations of *tomorrow* occur.

Second, non-UT interpretations of *tomorrow* would not need to be embedded under a speech verb, since FID often does not involve an explicit speech or attitude verb.

Third, the FID explanation predicts that non-UT readings should not be possible in narrator-oriented clauses, in other words, clauses where a first-person subject is reporting their own thoughts. First-person pronouns in FID always refer back to the narrator, rather than the protagonist, and the narrator's time is utterance time (Banfield, 1982).

2.2. Analysis 2: *tomorrow* is a shifty indexical

Another analysis consistent with a pure indexical view of *tomorrow* is that non-UT interpretations of *tomorrow* are due to indexical shift. Indexical shift is a phenomena found in many languages in which indexicals embedded under a speech or attitude verb are interpreted relative to the embedded context rather than the matrix context.[4]

In the Korean example in (2), 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).

(2)   Context: It is January 8th.
          cinan   cwu-ey   Mary-ka   nwuka   nayil   ttenanta-ko   malhayss-ni?
          last   week-in Mary-NOM   who-NOM   tomorrow   leave-C   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, 1999; Anand and Nevins, 2014; Deal, 2014). Under such analyses, the indexicals involved are still considered pure indexicals, because they are evaluated relative to a context parameter. Supporting evidence for

---

[3]I present the analysis from Eckardt (2014) because it is one of the most complete treatments of temporal phenomena, but see Schlenker (2004), Sharvit (2008), and Maier (2015) for other analyses.

[4]For a list of languages with shifty indexicals and a more in-depth discussion of the phenomena, see Deal (2017).

this treatment comes from the fact that shifty indexicals do not allow quantificational binding.

One plausible explanation for non-UT interpretations of *tomorrow*, then, is that they are instances of indexical shift. If this is the case, *tomorrow* may still be a pure indexical, since the dominant analyses of indexical shift posit that the original context parameter is overwritten by a context parameter representing the embedded speech context. The lexical semantics of *tomorrow* do not need to change: non-UT interpretations are created by a covert shift operator that manipulates the temporal parameter of the context tuple.

This account predicts that *tomorrow* should be infelicitous in quantificational binding environments, as reported for other shifty indexicals. In addition, it predicts that non-UT interpretations should only arise when *tomorrow* is embedded under a speech or attitude verb.

2.3. Analysis 3: *tomorrow* is anaphoric

If neither of the two accounts sketched above capture the behavior of *tomorrow*, then perhaps *tomorrow* is not a pure indexical, and its non-UT uses are anaphoric. Saying that non-UT interpretations of *tomorrow* are anaphoric is only a partial account, since in order to understand such readings, we also need to know what kind of object *tomorrow* can be anaphoric to. As work on *now* highlights, seemingly temporal expressions are not always anaphoric to any salient time in the discourse context; they may be anaphoric to something more complex, such as a result state (Altshuler and Stojnić, 2015; Altshuler, ming). Even if *tomorrow* has anaphoric uses, therefore, it may not be able to take any salient time as its reference time. For now, I set aside this question and sketch out the predictions that all anaphoric accounts make.

First, an anaphoric account predicts that non-UT readings could arise even when there is no manipulation of the context parameter as there is in FID or indexical shift. Because of this, an embedding speech verb is not predicted to be obligatory.

Second, the anaphoric account does not necessarily predict that other indexicals will behave similarly in the same contexts, since it posits a different lexical semantics for *tomorrow*, one that allows anaphoric selection of reference times. We might expect *tomorrow* to behave similarly to *the next day*, since *the next day* is also anaphoric; however, if *tomorrow* is anaphoric to a different kind of referent, this will not necessarily be the case.

2.4. Evaluating analyses of non-utterance time *tomorrow*

I have outlined three kinds of analyses for non-UT *tomorrow*: two consistent with pure indexicality, and one that takes an anaphoric approach. These accounts make different predictions about the availability of non-UT readings in a number of environments, summarized in Fig. 1.

Table 1: Overlap of analysis predictions

Subject	Unembedded	Embedded
1st-person		Indexical
2nd-person		shift
3rd-person	FID	

Indexical shift hypothesis =
Anaphoric hypothesis =
FID hypothesis =

Using experimental methods, we can evaluate these accounts by testing the availability of non-UT readings in contexts for which their predictions differ. In Experiment 1, I establish a baseline of acceptability of non-UT *tomorrow*. In Experiments 2 and 3, I test the predictions of the indexical accounts by manipulating the environment in which *tomorrow* occurs. Last, I report results from a debriefing task that bears upon the anaphoric hypothesis.

3. Experiment 1: establishing a baseline

Experiment 1 establishes a baseline of acceptability for non-UT interpretations of *tomorrow*. The grammatical environment used is one that all three analyses predict should license non-UT readings of *tomorrow*: embedding under a speech verb with a third-person subject.

3.1. Method

Data on the acceptability of non-UT interpretations of *tomorrow* in American English was collected through a comic-captioning task where participants rated captions for three-panel 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; we expect true captions to receive higher ratings than false captions. If participants rate the *tomorrow* items higher than the false baseline and close to *the next day*, then we confirm that non-UT readings of *tomorrow* are possible.

3.1.1. Participants

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.[5] This left 72 participants. These exclusion criteria were preregistered through the Open Science Foundation.

3.1.2. Materials

20 critical items were developed and distributed across four Latin Squared lists. Each list was combined with the same set of 10 fillers. Each item included a comic strip and a sentence below it (Fig. 1). In the comic's first panel, one character promises to 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. Participants were asked to judge the sentence as a caption for the third panel.

---

[5]This high rate may be due to the fact that there was no training item that required participants to read the day-of-week labels for the comics.

Figure 1: Experiment 1 example stimulus

Kevin is angry because Kate said that she would water his plants $\left\{ \begin{array}{l} \text{tomorrow} \\ \text{the next day} \\ \text{Friday} \\ \text{Saturday} \end{array} \right\}$.

Four conditions were created by manipulating the final 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 day of week name of the second panel, a factually correct baseline. Whether or not speakers allow non-UT readings of *tomorrow*, they were expected to interpret the *tomorrow* in the first panel as referring to the second day shown in the panel. The captions containing *tomorrow* are felicitous only if *tomorrow* is given a non-UT reading, since the first character promises to do the action on the second day depicted by the comic, and the other character would not have grounds for anger on the UT reading.

Table 2: Experiment 1 predictions

Temporal expression	Truth	Predicted ratings
False control	False	Low
True control	True	High
the next day	True	High
tomorrow	If participant allows non-UT reading, True	High
	If not, False	Low

Three kinds of fillers were used: bad fillers, which had captions that were obviously incorrect; good fillers, which had captions that were correct and required no temporal reference; and medium fillers, which were factually correct, but under- or over-informative.

We also collected basic demographic information about the participants, in order to explore whether any sociolinguistic factors such as age or geographic location affected their ratings; none of these factors were found to be informative, so discussion of them will be omitted.

3.1.3. Procedure

Stimuli were displayed and responses collected using the Ibex platform for web-based experiments. Each experimental session began with an informed consent form and a demographic survey. Next, participants read that they would see a comic strip with a sentence below it and be asked to rate the acceptability of the sentence as a caption for the third panel of the comic. Participants were given 3 items in order to train them in scale use: a true item, which they were told most participants would rate at 7; a false item, which they were told most participants would rate at 1; and a medium filler, which they were told most participants would rate at 4.

3.1.4. Regression analysis

An analysis using paired t-tests was preregistered, but after discussion with colleagues, we decided to use a mixed effects ordinal regression model.[6] The maximal random effects structure was used for all models: we included random intercepts and slopes for all fixed-effects predictors, for participants, and for items. 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.

In addition to the model described above, we ran a model that included the pragmatically suboptimal fillers as a fixed-effect predictor, in order to compare the ratings for these fillers to the *tomorrow* condition. This comparison was not preregistered.

3.2. 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* items (Table 3).

Figure 2: Experiment 1 participant means by condition

---
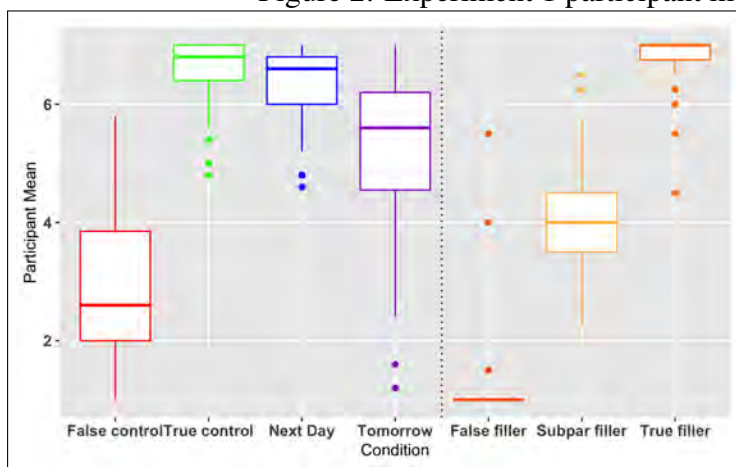
[6]The comparisons of interest were the same under both analyses.

Table 3: Experiment 1 results

Condition	Mean ratings	95%CI
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

In the regression analysis, all three coefficients were reliable effects at $p < 0.00001$. This shows that the ratings in the *tomorrow* condition were significantly different from the false baseline, but also from *the next day*. Participants accept non-UT readings of *tomorrow*, but to a lesser degree than *the next day*.

Table 4: Experiment 1 mixed effects regression analysis, fixed effects (N=1440)

	$\hat{\beta}$	z	p
False control	-3.36(+/- 0.36)	-9.3	< 0.00001
True control	2.47(+/- 0.38)	6.45	<0.00001
next day	1.53(+/- 0.29)	5.35	<0.00001

We might ask whether *tomorrow* items have high ratings not because they are grammatical, but because they can be accommodated through semantic coercion. The ratings of the pragmatically sub-optimal fillers suggest otherwise; participants rate the *tomorrow* items more highly than these items, which required a small amount of accommodation in order to fit the context.

To test whether the *tomorrow* scores were significantly different from those of the pragmatically subpar fillers, a second regression model that included the medium fillers was run.

Table 5: Experiment 1 mixed effects regression analysis including medium fillers, fixed effects (N=1728)

	$\hat{\beta}$	z	p
False control	-2.78(+/- 0.29)	-9.5	< 0.00001
True control	2.25(+/- 0.33)	6.74	<0.00001
next day	1.34(+/- 0.25)	5.45	<0.00001
Medium fillers	-1.53(+/-0.27)	-5.59	<0.00001

The coefficient for the medium filler condition was significant, indicating that *tomorrow* is rated significantly higher than the medium fillers. The difference in ratings between the medium fillers and the *tomorrow* items provides more evidence that non-UT *tomorrow* is accepted.

3.2.1. Interspeaker variation

There is interspeaker variation in the acceptability of the non-UT *tomorrow* items. Fig. 3 shows the means for each condition by participant in order of increasing *tomorrow* means. While most mean *tomorrow* ratings group together above the scores for the bad baseline, some speaker means for *tomorrow* are just as low as the bad baseline scores. This suggests that the mean for *tomorrow* is lower than *the next day* not because all participants consistently give it medium ratings, but because there is a small group of participants who rate it very low.
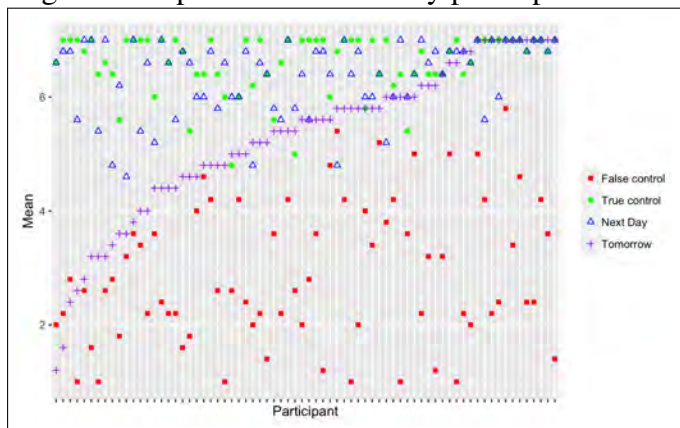
Figure 3: Experiment 1 means by participant

Table 6: Overlap of analysis predictions

Subject	Unembedded	Embedded
1st		Indexical
2nd		shift
3rd	FID	

Indexical shift hypothesis =
Anaphoric hypothesis =
FID hypothesis =

3.3. Discussion

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

Having established this baseline of acceptability, I turn to evaluating the three analyses outlined in Section 2 (Fig. 5). Experiment 2 manipulates pronoun use to explore whether non-UT interpretations of *tomorrow* can be explained by FID effects.

4. Experiment 2: exploring Free Indirect Discourse as an analysis

One hypothesis about non-UT readings of *tomorrow* is that they arise from FID effects. Experiment 2 tests this hypothesis using first-person narration, which, as discussed in Section 2.1, blocks protagonist-oriented readings. If participants rate *tomorrow* items comparably in Experiment 1 and 2, the non-UT readings are not due to FID effects alone.

4.1. Methods

52 participants were recruited through Amazon Mechanical Turk.[7] 4 failed to meet the inclusion criteria and were removed, leaving 48 participants balanced across experimental lists.[8]

---

[7] Analysis of the effect size from Experiment 1 suggested that the number of participants could be reduced.

[8] A training item was added that highlighted the day-of-week labels, reducing the participant removal rate.

The stimuli and fillers from Experiment 1 were modified to use first-person narration: captions were changed to use first-person pronouns, and the promiser was labeled as the narrator (Fig. 4).

The experimental procedure was as in Experiment 1, except that participants were instructed that each comic represented a diary written by one of the characters on the day it describes.

Figure 4: Experiment 2 example stimulus
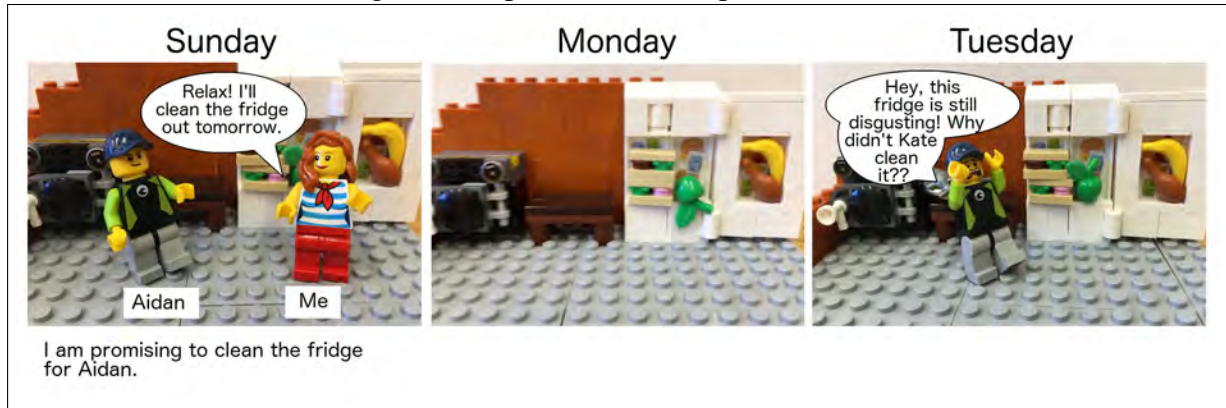
Aidan is angry because I said that I would clean the fridge $\left\{ \begin{array}{l} \text{tomorrow} \\ \text{the next day} \\ \text{Sunday} \\ \text{Monday} \end{array} \right\}$.

4.2. Results

The results of Experiment 2 were very similar to Experiment 1. Participants rated the *tomorrow* items lower than *the next day*, but above the pragmatically suboptimal fillers, the false fillers, and the false control items (Table 7). As in Experiment 1, there was a small group of participants who gave *tomorrow* items low ratings, but the majority gave high ratings to *tomorrow* items.

Table 7: Participant means by condition for Experiment 1 and Experiment 2

Condition	Experiment 1	Experiment 2
False control	2.9	2.2
Tomorrow	5.3	5.6
The next day	6.4	6.5
True control	6.6	6.6

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

4.3. Discussion

Experiment 2 replicated Experiment 1. There was no significant difference between the *tomorrow* ratings in the two experiments. Contrary to the predictions of the FID account, participants rated the *tomorrow* items with first-person subjects similarly to the Experiment 1 items. This suggests that non-UT interpretations of *tomorrow* are not limited to FID contexts.

Neither Experiment 1 or Experiment 2 provide evidence for or against the other two hypotheses: the indexical shift and the anaphoric accounts. Experiment 3 tests the predictions of the indexical shift account.

Figure 5: Attested pattern of non-UT *tomorrow*

Subject	Unembedded	Embedded
1st-person		✓
2nd-person		
3rd-person		✓

Indexical shift hypothesis =
Anaphoric hypothesis =
FID hypothesis =

## 5. Experiment 3: testing the indexical shift account

Experiment 1 showed that non-UT readings of *tomorrow* are possible under speech-verb embedding. Experiment 3 tests whether such embedding is necessary, as predicted by the indexical shift account. In Experiment 3, we remove the embedding verb from the critical sentences. The indexical shift account predicts that non-UT readings should not arise in this context, because the context parameter has not been shifted. Anaphoric accounts, by contrast, predict that this should not affect the availability of non-UT readings, so long as a non-UT perspective is salient.

If the results show that non-UT 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 against the two pure indexicality-consistent accounts.

5.1. Methods

53 participants were recruited through Amazon Mechanical Turk. 5 failed to meet the exclusion criteria and were removed, leaving 48 participants balanced across experimental lists. The same experimental methods were used as in Experiment 1 and 2, but the embedding speech verbs were removed from the captions. To provide a salient previous time, the conversation between the characters is mentioned in the caption, but without an embedding speech verb (Fig. 6).

Figure 6: Experiment 3 example stimulus
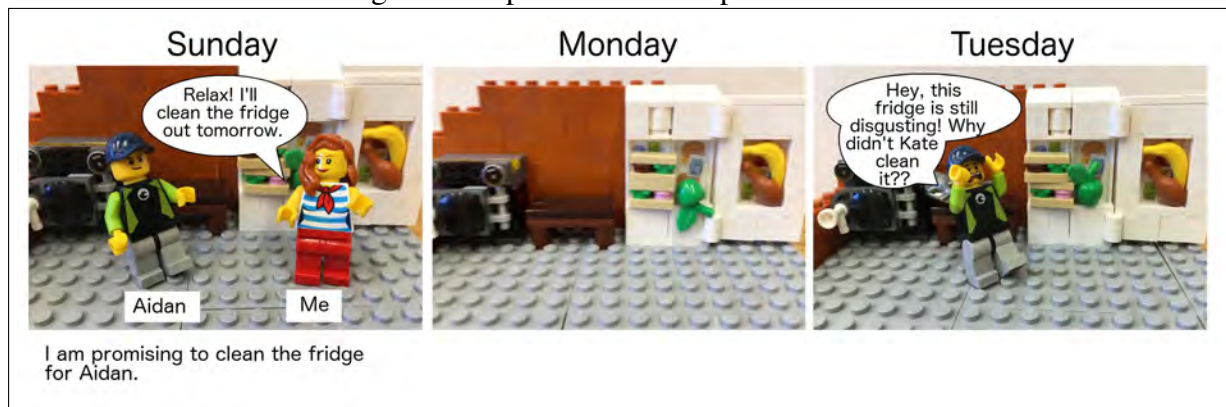
It was such a simple task to clean the fridge $\begin{Bmatrix} \text{tomorrow} \\ \text{the next day} \\ \text{Friday} \\ \text{Saturday} \end{Bmatrix}$! I can't believe I forgot.

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. The medium fillers and the false fillers were rated higher than in previous experiments.

Table 8: Comparison of mean ratings across experiments

Condition	Experiment 1	Experiment 2	Experiment 3
False control	2.9	2.2	2.1
Tomorrow	5.3	5.6	4.1
The next day	6.4	6.5	5.9
True control	6.6	6.6	6.3
False fillers	1.1	1.2	1.4
Medium fillers	4.0	3.9	5.0
True fillers	6.8	6.7	6.4

Figure 7: Experiment 3 main condition ratings
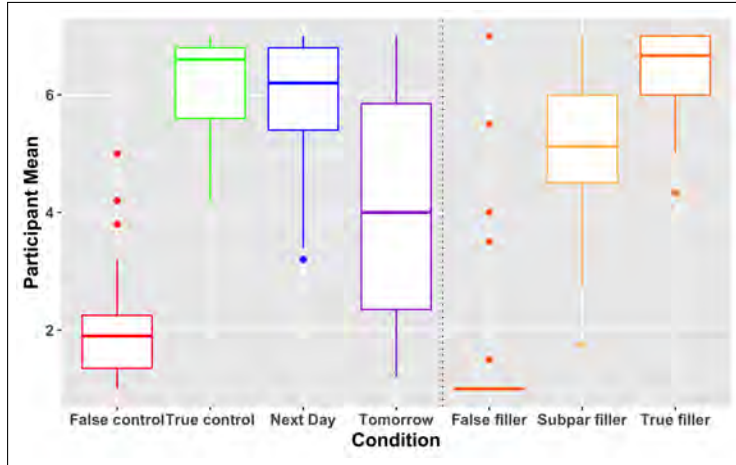
In the mixed-effects ordinal regression model, all three coefficients were reliable effects at p < 0.00001. Thus, despite the lower *tomorrow* scores in this experiment, the *tomorrow* condition was still significantly different than the false control condition.

Table 9: Experiment 3 mixed effects regression analysis, fixed effects (N=960)

	$\hat{\beta}$	z	p
False control	-3.41(+/- 0.50)	-6.81	< 0.00001
True control	3.20(+/- 0.48)	6.63	<0.00001
next day	2.54(+/- 0.50)	5.08	<0.00001

Although the main comparisons in Experiment 3 were similar to Experiment 2, a mixed-effects ordinal regression model comparing Experiment 1 and 3 finds significant interactions between experiment and *tomorrow* and between experiment and *the next day*, indicating that the differences between the Experiment 1 and Experiment 3 ratings for the *tomorrow* and *the next day* conditions were significant.

Table 10: Experiment 1 and Experiment 3 comparison mixed-effects regression analysis, fixed effects and interactions (N=2400)

Condition	$\hat{\beta}$	z	p
False control	-3.44(+/-0.36)	-9.44	< 0.00001
True control	2.31(+/-0.36)	6.47	<0.00001
the next day	1.50(+/-0.33)	4.52	<0.00001
Experiment 2	-1.72(+/-0.47)	-3.6	0.0002
False control * Experiment 2	0.35(+/-0.53)	0.65	0.52
True control * Experiment 2	0.87(+/-0.52)	1.68	0.09
the next day * Experiment 2	0.97(+/-0.47)	2.06	0.039

The 95% confidence intervals for participant means were wider than in Experiment 1 and 2; for *tomorrow* items, the 95% CI was larger than 1 Likert scale point.

Table 11: Experiment 3 results

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
True control	6.3	6.1-6.5
Bad fillers	1.4	3.9-5.4
Medium fillers	5.0	4.7-5.4
Good fillers	6.4	6.3-6.6

5.2.1. Interspeaker variation

There was a higher amount of interspeaker variation in *tomorrow* ratings in this experiment, along with greater variance in ratings for other conditions. The distribution of *tomorrow* scores in Experiment 3 is almost bimodal: one group of participants rates *tomorrow* near *the next day*, while another group rates it near the bad baseline. In Fig. 8, which plots *tomorrow* ratings by participant, some participants have scores clustered at the bottom, and others have them clustered at the top. Both groups occasionally give a rating at the opposite end of the scale (shown by a long line connecting the participant's lowest and highest ratings).

Figure 8: Experiment 3 *tomorrow* ratings (left) and *the next day* ratings (right) by participant, in order of increasing difference between highest and lowest rating
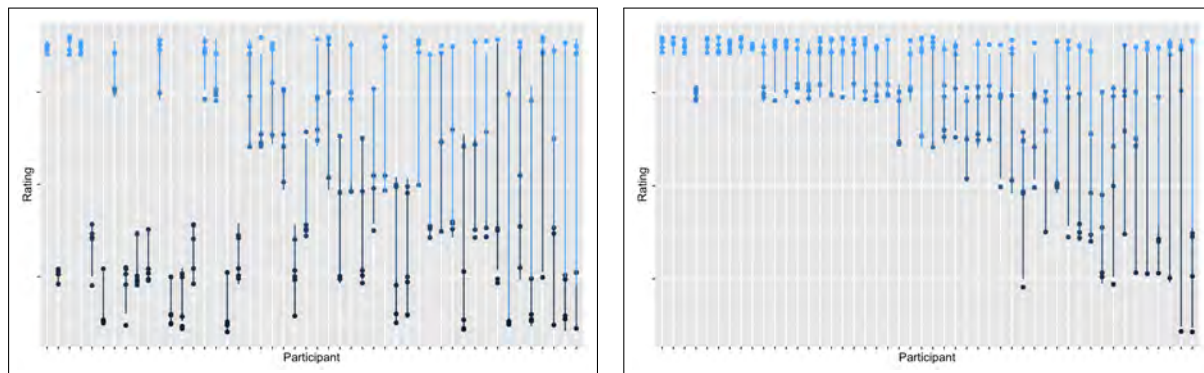
The question is whether the participants who give *tomorrow* low ratings in this experiment are the same as those in Experiment 1 and 2. It may be that Experiment 3 sampled more heavily from this population; or it may be that removing the embedding verb has a real effect on the acceptability, and that these participants would have accepted the *tomorrow* items in Experiment 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 pragmatically sub-optimal fillers, though they were still significantly above the bad baseline. In addition, there was more interspeaker variation in *tomorrow* ratings than in previous experiments.

The lower *tomorrow* ratings are problematic. One possibility is that Experiment 3 sampled more heavily from the speaker group that does not ever accept non-UT readings of *tomorrow*. Another (non-exclusive) explanation is that *tomorrow* is anaphoric, but to something more difficult to process than a simple time. This could explain why *tomorrow* is rated lower than *the next day*, and, if the attitude verb facilitates the anaphora resolution, why the ratings are lower in Experiment 3.

Figure 9: Attested pattern of non-UT *tomorrow*

Subject	Unembedded	Embedded
1st-person	✓	✓
2nd-person		
3rd-person		✓

Indexical shift hypothesis =
Anaphoric hypothesis =
FID hypothesis =

Although one group of participants gave the *tomorrow* items mostly low ratings, a substantial group gave them consistently high ratings. These results argue against an indexical shift account, since it predicts that non-UT *tomorrow* should never occur outside of embedding. I proceed with an analysis of *tomorrow* that focuses on the group of speakers who accept *tomorrow* in Experiment 3, since the results suggest that for a substantial group of speakers, non-UT readings of *tomorrow* are possible in non-embedded contexts.

6. Anaphoric accounts of *tomorrow*

Experiments 1, 2, and 3 test the predictions of three accounts: the FID account, the indexical shift account, and the anaphoric account. Experiment 2 showed that contrary to the FID account, non-UT interpretations of *tomorrow* occur in narrator-anchored environments. Experiment 3 showed that unlike indexical shift, they also occur outside of embedded contexts for one group of speakers. Under these accounts, *tomorrow* could be a pure indexical; their failure suggests that non-UT readings are anaphoric, and *tomorrow* is not a pure indexical. However, this is a partial account: we also need to determine to what kind of referent *tomorrow* can refer.

6.1. Is *tomorrow* anaphoric to salient times?

Pure indexicals are infelicitous in quantificational binding contexts, because their referent does not covary with the quantifier (as it is fixed to the utterance context). Anaphoric expressions, on the other hand, can appear in quantificational binding contexts so long as there is a salient referent for them that covaries with the quantifier. The anaphoric expression *the next day* is felicitous in (3), because the quantification over time provides a salient referent.

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

Since *tomorrow* is a temporal expression, an intuitive anaphoric account is that it is anaphoric to a salient time in the preceding discourse: *tomorrow* is akin to *the next day*, except that *tomorrow* allows indexical interpretations. This account predicts that anaphoric uses of *tomorrow* should be possible in quantificational binding contexts where times are quantified over.

However, this account is not the only possible anaphoric account. As work on *now* has illustrated (Altshuler and Stojnić, 2015; Altshuler, ming), temporal expressions are not necessarily anaphoric to any salient time in the discourse context. Moreover, the fact that *tomorrow* items received lower ratings than *the next day* suggests that they do not pattern exactly alike.

6.1.1. Evaluating the time anaphoric account

The prediction that *tomorrow* should be felicitous under quantification over times was tested in a postexperiment task following Experiment 2. The task measured acceptability of *tomorrow* in two quantificational binding contexts: quantification over times (4) and over speech contexts (5).

(4)   Every time { I/Kevin } wash my car, it rains { tomorrow/the next day }.

(5)   Every time the UPS person says that { my/the } package has been delivered, it doesn't show up until { tomorrow/the next day }.

Participants rated 1 time quantification item and 2 speech-context quantification items in each of the 4 conditions produced by crossing temporal adverbial (*tomorrow* or *the next day*) with person (1st or 3rd), for a total of 12 items, presented using a Latin squared design.

6.1.2. Results

Quantificationally bound *tomorrow* items received low ratings in all conditions. An ANOVA of the *tomorrow* items shows no significant difference by condition (0.55(3,188)= 0.67; $p > 0.05$).

Table 12: Experiment 2 binding task results

Condition	Mean *tomorrow* scores [95%CI for part. means]	Mean *next day* scores [95%CI for part. means]
Embedded 1st-person	2.7 [2.3-3.0]	5.9 [5.6-6.2]
Unembedded 1st-person	2.4 [1.8-2.9]	6.2 [5.8-6.6]
Embedded 3rd-person	2.7 [2.4-3.0]	5.8 [5.4-6.1]
Unembedded 3rd-person	2.5 [2.0-2.9]	6.1 [5.7-6.6]

This quantificational binding task disconfirms the predictions of the time-anaphoric account: *tomorrow* is not anaphoric to just any salient time in the preceding discourse.

6.2. A perspectival view of non-utterance time *tomorrow*

The experimental data suggests that for some speakers, *tomorrow* can refer anaphorically, but not to just any time. I propose that it is anaphoric to something more complex: a perspective.

6.2.1. Perspectival anaphoric reference

In proposing that *tomorrow* is anaphoric to salient perspectives, I draw upon Barlew (2017)'s analysis of deictic motion verbs as anaphoric to salient perspectives. The motion verb *come* can refer relative to the speaker's location at UT; to the speaker's location at ET (6); to the addressee's location at UT or ET; or to that of an attitude-holder (7).

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

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

Working in a dynamic semantics framework, Barlew (2017) proposes that *come* presupposes that a salient perspective-holder in the Common Ground is located at the destination of motion. He notes that an indexical analysis of *come* is ruled out by the fact that *come* can appear in quantificational binding contexts, so long as the perspective-holder covaries with the quantifier. In (8), each instantiation of the woman provides a salient perspective as an anchor for *come*; *was glad* acts like an attitude verb in heightening the salience of each woman's perspective.

(8)   Every woman was glad that her wayward child came to Christmas dinner. (Barlew, 2017: 52)

Barlew (2017) follows Roberts (2015) in taking a perspective to be a set of centered worlds: the worlds in which the agent's beliefs are true and the center is their self-ascribed location. Regardless of how perspective is formalized, his data reveals an important point: perspectives must be time-indexed in order to distinguish between the ET and UT licensing of *come*.

6.2.2. Towards a perspectival account of *tomorrow*

I propose that, like perspectival motion verbs, *tomorrow* is anaphoric to salient perspectives: it receives non-UT interpretations when it is anchored to a non-UT perspective, and UT interpretations when it is anchored to an UT perspective. In particular, I propose that *tomorrow* takes the time parameter of a salient perspective as its reference time.

This proposal captures the pattern of acceptability in the experimental data. In Experiment 1 and 2, a salient non-UT perspective is provided by the subject of the attitude verb. In Experiment 3, the event-time perspective of the narrator is made salient by focusing on the ET evaluation of the task ('such a simple task!'). If ET perspectives are less salient than attitude-verb introduced perspectives, it would explain the lower ratings for *tomorrow* in Experiment 3; the pattern could also be explained by interspeaker variation in the accessibility of ET perspectives.

Furthermore, a perspectival view of *tomorrow* explains why its anaphoric uses have been overlooked: the time-index of a perspective is often identical to that of the utterance context. It is only in cases where the perspective is an event-time perspective, or where the perspective-holder is mistaken about their temporal location (assuming perspective is *de se* as in Roberts, 2015), that temporal perspectival anaphora and (shifty) indexicals will diverge.

6.2.3. Quantificational binding revisited

So far, the experimental data is compatible with a perspectival view of *tomorrow*, but does not provide positive evidence for it. Another look at quantificational binding contexts could provide this: under the perspectival account, *tomorrow* should be felicitous under quantificational binding if the perspective covaries with the quantifier (9).

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

First-person binding examples should also become more acceptable when the saliency of the ET perspectives is increased by additional perspectival context, such as expressives or epithets.

(10)   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.

(11)   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.

I find these examples somewhat better than the ones tested experimentally, but full testing of the quantificational binding examples is left for future work.

7. Conclusion

There is an emerging consensus that many expressions thought to be purely dependent on the context parameter for their meaning are not (Harris and Potts, 2009; Altshuler and Stojnić, 2015; Maier, 2017). I have argued that *tomorrow* has anaphoric uses, at least for a substantial portion of speakers, and therefore is not a pure indexical. In a series of experiments, I have shown that non-UT readings of *tomorrow* are not limited to FID environments or to attitude reports, as is indexical shift in other languages. In addition, I present quantificational binding evidence suggesting that *tomorrow* is not anaphoric to times. I argue instead that *tomorrow* is anaphoric to perspectives and that non-UT readings arise whenever the time-index of the perspective does not match that of the utterance context.

Although I have presented experimental evidence against several possible explanations, open questions remain. Future work should investigate whether *tomorrow* can appear in quantificational binding situations where the perspective covaries with the quantifier. In addition, 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 the environments in which indexical shift occurs. When

instances of indexical shift are being investigated, care should be taken to test the environments in which perspective shift, but not indexical shift, is predicted to occur.

The non-UT interpretations of *tomorrow* discussed here add to the growing list of expressions that are sensitive to perspective: the temporal and spatial self-location of an agent in a world. This work suggests an elegant parallel between perspectival items in the spatial domain, like *come*, and perspectival items in the temporal domain, like *tomorrow*. By probing into seemingly exceptional uses of expressions like *tomorrow* experimentally, we can begin to understand the role of perspective and context-sensitivity in spatio-temporal reference.

References

Altshuler, D. (Forthcoming). Tense and temporal adverbs ("I learned last week that there would now be an earthquake"). In L. Matthewson, C. Meier, H. Rullman, and T. E. Zimmermann (Eds.), *The Blackwell Companion to Semantics*. Wiley.

Altshuler, D. and U. Stojnić (2015). The attention-coherence model of prominence: A look at 'now'. Paper presented at the International Conference on Prominence in Language, University of Cologne.

Anand, P. and A. Nevins (2014). Shifty operators in changing contexts. In R. Young (Ed.), *The Proceedings of Semantics and Linguistic Theory*, Volume 24, Ithaca, NY, pp. 20–37. Cornell University.

Banfield, A. (1982). *Unspeakable Sentences: Narration and Representation in the Language of Fiction*. New York, NY: Routledge.

Barlew, J. (2017). *The semantics and pragmatics of perspectival expressions in English and Bulu: The case of deictic motion verbs*. Doctoral Dissertation, The Ohio State University.

Deal, A. R. (2014). Nez Perce embedded indexicals. In H. Greene (Ed.), *Proceedings of Semantics of Under-represented Languages of the Americas*, Volume 7. University of Massachusetts, Amherst Graduate Linguistics Student Association.

Deal, A. R. (2017). Shifty asymmetries: universals and variation in shifty indexicality. Ms. The University of California, Berkeley.

Doron, E. (1991). Point of view as a factor of content. In S. K. Moore and A. Z. Wyner (Eds.), *The Proceedings of Semantics and Linguistic Theory*, Volume 1, Ithaca, NY, pp. 51–64. Cornell University.

Eckardt, R. (2014). *The semantics of free indirect discourse: how texts allow us to mind-read and eavesdrop*, Volume 31 of *Current Research in the Semantics / Pragmatics Interface*. Leiden, The Netherlands: Brill.

Harris, J. and C. Potts (2009). Perspective-shifting with appositives and expressives. *Linguistics & Philosophy 32*(6), 523–552.

Kamp, H. (1971). Formal properties of 'now'. *Theoria 37*(3), 227–273.

Kaplan, D. (1989). Demonstratives. In J. Almog, J. Perry, and H. Wettstein (Eds.), *Themes from Kaplan*, pp. 481–564. Oxford: Oxford University Press.

Maier, E. (2015). Quotation and unquotation in free indirect discourse. *Mind & Language 30*(3), 345–373.

Maier, E. (2017). A good indexical is hard to find. Paper presented at Anaphora Resolution in Sign and Spoken Languages, University of Göttingen.

Oshima, D. (2006). Motion deixis, indexicality, and presupposition. In M. Gibson and J. Howell (Eds.), *The Proceedings of Semantics and Linguistic Theory*, Volume 16, Tokyo, Japan,

pp. 172–189. University of Tokyo.

Park, Y. (2018). *Attitudes de se and Logophoricity*. Doctoral Dissertation, University of Massachusetts, Amherst.

Roberts, C. (2015). Indexicality: de se semantics and pragmatics. Ms. The Ohio State University.

Schlenker, P. (1999). *Propositional attitudes and indexicality: a cross-categorical approach*. Doctoral Dissertation, MIT.

Schlenker, P. (2003). A plea for monsters. *Linguistics and Philosophy 26*, 29–120.

Schlenker, P. (2004). Context of thought and context of utterance. a note on free indirect discourse and the historical present. *Mind & Language 19*(3), 279–304.

Sharvit, Y. (2008). The puzzle of free indirect discourse. *Linguistics and Philosophy 31*, 353–395.

Stojnić, U. (2016). *Context-sensitivity in a Coherent Discourse*. Doctoral Dissertation, Rutgers University.