Featured paper

Minding the clock

Kathryn Bock,[a,*] David E. Irwin,[a] Douglas J. Davidson,[a] and W.J.M. Levelt[b]

[a] *Beckman Institute, University of Illinois, Urbana, IL 61801, USA*
[b] *Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands*

Abstract

Telling time is an exercise in coordinating language production with visual perception. By coupling different ways of saying times with different ways of seeing them, the performance of time-telling can be used to track cognitive transformations from visual to verbal information in connected speech. To accomplish this, we used eyetracking measures along with measures of speech timing during the production of time expressions. Our findings suggest that an effective interface between what has been seen and what is to be said can be constructed within 300 ms. This interface underpins a preverbal plan or *message* that appears to guide a comparatively slow, strongly incremental formulation of phrases. The results begin to trace the divide between seeing and saying—or thinking and speaking—that must be bridged during the creation of even the most prosaic utterances of a language.
© 2003 Elsevier Science (USA). All rights reserved.

Keywords: Language production; Eyetracking; Incrementality; Naming; Time-telling; Idioms

Perception and language are our portals to the world. Perception is the stuff of raw experience. Language takes us beyond raw experience to an understanding of things never immediately perceived, to other places, other times, and other minds. For this to work, language must be grounded in and shaped by experiences that allow us to induce the categories and relationships in events and states of affairs that we may never directly encounter. At the root of this grounding is the ability to communicate what we perceive. The research we report aims at understanding how this ability works for the deceptively simple process of telling the time.

Time-telling is anchored by the physical devices that are used to measure and mark time. These devices convey what we call *mundane* time, in contrast to the abstruse qualities of time of traditional interest to physicists (Hawking, 1998), philosophers (Le Poidevin, 1993), psychologists (Grondin, 2001), anthropologists (Gell, 1992), and linguists (ter Meulen, 1995). Mundane

time constitutes a widely used framework for conceptualizing, marking, and communicating the landmarks in a culturally shared timescape and our passage through it.

Among the basic tools for charting mundane time are calendars and clocks. For purposes of communication, these tools have associated with them the expressions that we use to talk about time. Though commonplace, these expressions and the circumstances in which they are used conjure up unexpected mysteries for cognitive psychology and psycholinguistics. Some of these mysteries lurk in the idiomaticity and cognitive complexity of time-telling.

Like prototypical idioms (along the lines of *kick the bucket*), time expressions frustrate rational analysis. Apart from reliably mentioning hour and minute marks, the forms of reference to hours, minutes, and the relationships between them can be unpredictable, literally opaque (what does *five of nine* mean? why is it said when the big hand points at 11?), and logic-defying: A British English speaker's *half six* means 6:30, a Dutch speaker's corresponding expression (*half zes*) means 5:30, and an

American stands a good chance of missing both appointments. In turnabout, British speakers are unsure whether an American's *quarter of* is a quarter before or a quarter after the hour. Yet unlike prototypical, nondecomposable idioms, time expressions are compositional and productive. Such expressions have been called *formulas* (Kuiper, 1996). What makes them idiomatic is the arbitrariness of their construction with respect to the reference domain.

This arbitrariness can be seen in the variations among basic expressions for telling time in even closely related languages. Table 1 gives a small sample from a few Indo-European languages in literal translation to English. The table illustrates some striking differences in the use of hourly reference points. In some languages and some kinds of time expressions, the hour of reference is the hour just past (as in *two-ten*) and it remains the same throughout the hour. The Italian examples are like this. In Czech and Russian, throughout an hour's passage the hour of reference can be what an English speaker would regard as the coming hour (as in *ten minutes out of the third hour*). In other languages and types of expression, the hour of reference changes at the half-hour from the hour just past to the coming hour, as in the English examples in the table.

The number of reference points also varies. In addition to the hourly reference point, languages may have a secondary reference point at the half-hour (in some Germanic languages) or the quarter hour (in Czech). The consequence is that what an English speaker calls *twenty after two* is likely to be called the equivalent of *ten before half three* in Dutch and *a quarter out of three and five minutes* in Czech. These secondary reference points can come into play at different points in the hour, depending on the language and dialect. In many Dutch dialects the secondary reference point operates between ten minutes before and ten minutes after the half-hour, but in German it is customary to use the secondary reference point between five before and five after the half-hour.

The cognitive complexity of mundane time is evident to children struggling to master it and to adults trying to teach it. For most of the rest of us the system is so familiar that its intricacies go unnoticed. To begin to appreciate what is involved, consider the perceptual and cognitive transformations required to produce a time expression in Dutch. A Dutch speaker viewing a digital clock that reads 2:40 is likely to express the time in a way that translates to English as *ten past half three*. Considering that there is no ten, no half, and no three to be found anywhere in the display, it is implausible to attribute the performance to a simple read-off or naming of the numbers. A rough, partial outline of the logic behind into the production of just a subset of time expressions is shown in Table 2 for English and Dutch.

Interest in clock-reading among psychologists has been confined mostly to work in neuropsychology and human factors (e.g., Cahn & Diesfeldt, 1977; Kartsounis & Crewes, 1994; Miller & Penningroth, 1997), with occasional forays into the perceptual and cognitive processes underlying performance in adults (Goolkasian & Park, 1980; Paivio, 1978) and children (Siegler & McGilly, 1989). As a target for psycholinguistic investigation the topic seems at best unlikely, though Shanon

Table 1
Common relative time expressions for 2:20, 2:25, and 2:35 in eleven Indo-European languages from three language families

Language family and language	English renditions (literal translations) of relative time names		
	2:20	2:25	2:35
Germanic			
English	Twenty past two	Twenty-five past two	Twenty-five to three
Dutch	Ten before half three	Five before half three	Five after half three
German	Twenty after two	Five before half three	Five after half three
Danish	Twenty minutes after two	Twenty-five minutes after two	Five minutes after half three
Swedish	Twenty after two	Five before half three	Five after half three
Italic (Romance)			
French	Two hours twenty	Two hours twenty-five	Three hours minus twenty-five
Spanish	Two and twenty	Two and twenty-five	Three minus twenty-five
Italian	Two and twenty	Two and twenty-five	Two and thirty-five
Slavic			
Polish	Twenty past two	Twenty-five past two	Twenty-five to three
Czech	Quarter out of three and five minutes	In five minutes half out of three	In ten minutes three-quarters out of three
Russian	Twenty minutes of the third [= third hour]	Twenty-five minutes of the third	Without twenty-five minutes three

Table 2
Telling five-minute times between five and twenty-five past the hour from an analog clock, using relative time expressions (e.g., *ten after seven*) in English and Dutch: A logical decomposition

English	Dutch
1. Name the minute (e.g., "ten")	
a. Find the minute hand	a. Find the minute hand.
b. Assess the angle between the 12 o'clock mark and the minute hand.	b. Assess the angle between the 12 o'clock mark and the minute hand.
c. The minute number-name is the size of the angle divided by 6 ($30° = 5$ min)	c. If the angle is between $30°$ and $90°$, the minute number-name is the size of the angle divided by 6 ($30° = 5$ min)
• unless the angle is $90°$, when the minute number-name can be *quarter*	• unless the angle is $90°$, when the minute number-name can be *quarter*
	d. If the angle is $120°$ or $150°$ from the 12 o'clock mark, the minute number-name is the size of the angle *relative to the 6 o'clock mark* divided by 6 ($30° = 5$ min)
2. Name the relation (e.g., "after")	
If the minute hand is between $30°$ and $150°$, the relation is "after"	a. If the minute hand is between $30°$ and $90°$, the relation is "after".
	b. If the minute hand is at $120°$ or $150°$, the relation is "before".
3. Name the reference point (e.g., "seven")	
a. Find the hour hand.	a. Find the hour hand.
b. The name of the reference point is the number closest to the hour hand	b. If the minute hand is between $30°$ and $90°$, the name of the reference point is the number closest to the hour hand ($=$ value *b*).
	c. If the minute hand is at $120°$ or $150°$, the name of the reference point is *half* followed by value $b + 1$.

(1985) and Van der Henst, Carles, and Sperber (2002) have explored questions about the construction and pragmatics of time-telling expressions.

Our goal in understanding the cognitive underpinnings of time-telling is different, and we come at it from the perspective of issues in language production. We envision a production process that runs roughly along the lines depicted in Fig. 1, where something grandly called Perception and Thought evolves into movements of the articulators. The goal of research on production is to understand what happens in between thinking and speaking. Time-telling offers an unusually promising domain in which to explore this problem.

Scientific understanding of how people convey ideas linguistically is limited by the difficulty of gaining access to whatever it is that constitutes an idea, by the risk of distorting and trivializing the demands on production that comes from experimental efforts to specify both the idea and the utterance, and by the diversity of expression that characterizes normal communication even in carefully controlled conditions (Bock, 1996; Bock & Griffin, 2000; Ferreira & Swets, 2002). The telling of time overcomes many of these limitations.

Time-telling offers the advantages of a precise definition of what is to be conveyed linguistically (the "idea" is the time itself) and a repertoire of time-telling expressions that is available to almost all adults. The perceptual displays that give rise to synonymous time ideas can differ dramatically (e.g., between analog and
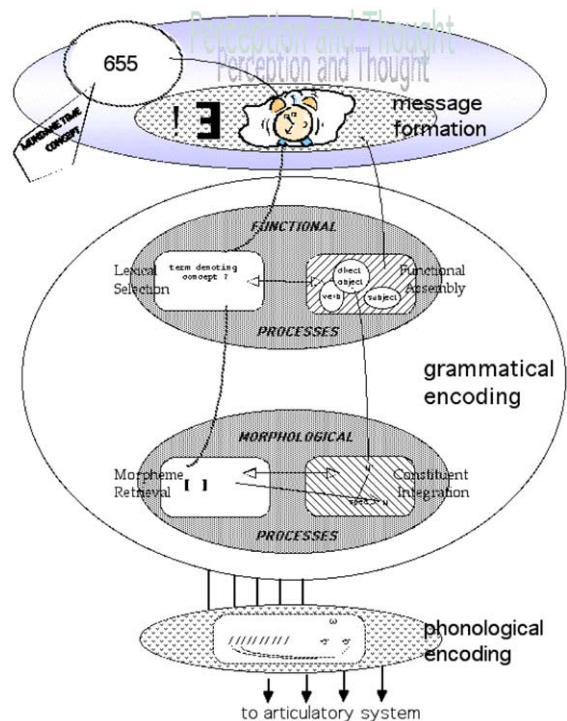

Fig. 1. Component processes in language production (adapted from Bock, Eberhard, Cutting, Meyer, & Schriefers, 2001).
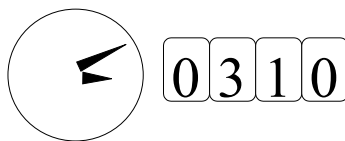
Fig. 2. Synonymous analog (left) and digital (right) clock displays.

digital displays, as in Fig. 2) and the range of time expressions, though limited, offers options (compare *one-twenty* and *twenty after one*). The idiomaticity of the expressions means that they are fairly rigid in form and content. They are nevertheless compositional. Consequently, their felicitous production requires speakers to map discrete elements of perceptual or conceptual information to variable elements and relations in phrasal expressions. Time-telling thereby provides a tractable approach to exploring and explaining how ideas become phrases, and to examining how the process changes when perceptions, expressions, and languages change.

Mundane-time expressions can be constructed in either of two familiar, viable, but contrasting systems, which we call *relative* and *absolute*. The relative system includes such expressions as *ten past two* and *quarter (or fifteen) to four*; the absolute system includes the corresponding expressions of *two-ten* and *three forty-five*. Table 3 illustrates the contrast between relative and absolute expressions for five-minute times in English and Dutch.

There are three important differences between the expression types. The first is that prototypical relative expressions contain an overt mention of the relationship between the hourly reference point and the minute (e.g., *past*, *after*; *to*, *till*, *of*); absolute expressions do not. The second is that in relative expressions, the reference point

changes during a whole hour's transit from the passing hour to the upcoming hour; in absolute expressions, the reference point remains the same. The third is that relative expressions name the minute before the hour, while absolute expressions name the hour before the minute.

The availability of these two familiar but different ways of expressing time and two familiar but different ways of displaying time opens a promising way for exploring how a time idea develops into a time utterance. One can readily examine changes in the temporal progress of production when the display format changes but the expression format remains the same, and when the display format remains the same but the expression format changes. In short, one can achieve the elusive goals of manipulating and controlling the input and the output of a fairly intricate, compositional production process.

To begin to trace the perceptual and cognitive underpinnings of time-telling, we exploited the possibilities of headband-mounted eyetracking. Because this technology permits unobstructed speaking while correcting for head movements that accompany speech, it allows the eyes to serve as meters of the mental processes that support language use. The value of this tool for the study of language production has been demonstrated in work that shows a tight link between the timing of eye movements and the timing of word production (Meyer, Sleiderink, & Levelt, 1998; Meyer & van der Meulen, 2000).

The time-locking between eye movements and the production of words can be observed even for fluent, spontaneously produced sentences (Griffin & Bock, 2000). Griffin and Bock showed speakers pictures of simple events (e.g., a dog chasing a mailman; a horse kicking a cow) in counterbalanced arrangements to control for perceptual and conceptual properties. Speakers described the events extemporaneously, in single sentences, with the pictures remaining in view

Table 3
Relative and absolute expressions for five-minute times in English and Dutch

Time on digital display	Time expression			
	Relative		Absolute	
	English: *It's . . .*	Dutch: *Het is . . .*	English: *It's . . .*	Dutch: *Het is . . .*
H:00	*H o'clock*	*H uur*	*H*	*H uur*
H:05	*five past H*	*vijf over H*	*H oh-five*	*H uur vijf*
H:10	*ten past H*	*tien over H*	*H ten*	*H uur tien*
H:15	*quarter past H*	*kwart over H*	*H fifteen*	*H uur vijftien*
H:20	*twenty past H*	*tien voor half H + 1*	*H twenty*	*H uur twintig*
H:25	*twenty-five past H*	*vijf voor half H + 1*	*H twenty-five*	*H uu vijfentwintig*
H:30	*half past H*	*half H + 1*	*H thirty*	*H uur dertig*
H:35	*twenty-five to H + 1*	*vijf over half H + 1*	*H thirty-five*	*H uur vijfendertig*
H:40	*twenty to H + 1*	*tien over half H + 1*	*H forty*	*H uur veertig*
H:45	*quarter to H + 1*	*kwart voor H + 1*	*H forty-five*	*H uur vijfenveertig*
H:50	*ten to H + 1*	*tien voor H + 1*	*H fifty*	*H uur vijftig*
H:55	*five to H + 1*	*vijf voor H + 1*	*H fifty-five*	*H uur vijfenvijftig*

Note. English speakers use *before*, *till*, and *until*, as well as *to*, and *after* as well *past*.

while eye movements were monitored. The results of this and several control tasks (involving delayed production, event comprehension, and mere viewing), motivated a general decomposition of the language production process into nonlinguistic *apprehension* and linguistic *formulation*. Respectively, apprehension and formulation roughly correspond to the initiation of message formation and the unfolding of grammatical encoding, in Fig. 1.

The hypothesized apprehension process involves an initial, rapid uptake of information that includes at least perception and event comprehension. It establishes the coarse conceptual structure of a perceived event, including some of the event's causal and aspectual features. The initial uptake may occur in little more than a single glimpse, in under 300–400 ms. Despite its speed, apprehension can be sufficient to identify a viable starting point for linguistic formulation. Griffin and Bock found that speakers' earliest saccades tended to move the eyes toward the entity that would serve as the subject of an upcoming sentence, although neither the role of the entity (e.g., as an agent or patient of the action) nor its perceptual prominence, taken alone, could objectively predict what the subject would be. That is, the selection of a subject appeared to depend on apprehending something about the event as a whole. Supporting this, event comprehenders' eye movements picked out the patient of an event as rapidly as apprehension occurred, a performance that is possible only with a general grasp of an event's causal structure. This points to a fast, parallel scene extraction and analysis process, along lines suggested in work by Biederman (1981), Gordon (1999), Loftus and Mackworth (1978), and Potter (1975).

Subsequent eye movements over the elements of the scenes in Griffin and Bock's (2000) study suggested a linguistic formulation process that is slower and more serial, or *incremental*. The eyes moved successively to the referents of the upcoming subjects and objects of sentences, in advance of the onset of corresponding words. These eye movements were well coordinated with the contents of accompanying fluent speech. On average, the eye reached each object approximately 900 ms prior to the onset of the word that named it. In light of the findings by Meyer and colleagues (Meyer et al., 1998; Meyer & van der Meulen, 2000), it is probably not coincidental that this eye-voice span is virtually identical to the average time needed for naming pictured objects (Snodgrass & Yuditsky, 1996).

The performance of time-telling poses stricter tests for the hypothesized distinction between nonlinguistic apprehension and linguistic formulation. Time expressions differ markedly from the utterances elicited in Griffin and Bock's experiment, where in all probability speakers described events they had never before seen (e.g., a truck hitting a nurse) with sentences they had never before said (e.g., "A nurse is getting hit by a truck"). In the present experiments, speakers viewed familiar displays and produced familiar expressions that are exceptionally well practiced and highly idiomatic (e.g., "ten twenty-five"). Likewise, the perception and comprehension of a clock display may have different properties than the perception and comprehension of a simple event: Whereas evolution has prepared the human species to rapidly understand who is doing what to whom and perhaps to talk about it, the telling of time has a history much too short for adaptations in the human cognitive system to arise. Accordingly, the perception of clock displays might be more linguistically saturated than event perception, in the sense that the linguistic properties of an upcoming time expression might drive visual processing of the display. Alternatively, it could be that the perception of clock displays is more rigidly perceptual and less flexibly interfaced with the linguistic formulation system, reflecting the nature of clocks as artifacts.

These considerations are reflected in the two goals of our work. The first has to do with assessing whether the linguistic formulation of time expressions differs from more productive expressions. A speaker's familiarity with clock displays and with the accompanying expressions for time might eliminate the ocular accompaniments of the methodical, incremental linguistic formulation process observed by Griffin and Bock, or the serial order of eye fixations could lack a tight temporal link to the components of time expressions. Either way, the order of fixation might be determined by factors other than order of linguistic expression. Such factors include the perceptual properties of the display, such as the size of the hands in an analog clock or the left-to-right number array on a digital clock.

Our second goal was to examine whether and how the uptake of visual information changes to support the creation of time expressions when the linguistic reference points for time-telling differ. From an experimental perspective, English and Dutch are nearly ideal languages for doing this. They are closely related West Germanic languages, with many cognates in their vocabularies and widespread similarities in syntactic structures and mechanisms. The cultures are similar, the technologies for time-telling are similar and share the same history (Whitrow, 1988), and most of the basic expressions for telling time have near-literal translations into the other language. For example, *ten to five* is *tien voor vijf*, *ten after five* is *tien over vijf*, and so on (see Table 3). Speakers in both communities make use of relative and absolute expressions. Yet despite such widespread similarities, there is a fundamental difference in how time is told using relative expressions in the two languages.

The difference is in the reference frames for relative time expressions. English speakers tell relative time

Hour
12

Five past hour
1

Five to hour 11

Ten past
2 hour

Ten to
hour 10

Comparable expressions
in English and Dutch
with whole-hour reference point

Quarter 9
to hour

3 Quarter
past hour

Ten
after 8
half of
next hour

Dutch (glossed)
switches
to half-hour
reference point

4 Ten
before
half of
next hour

Five
after 7
half of
next hour

5 Five
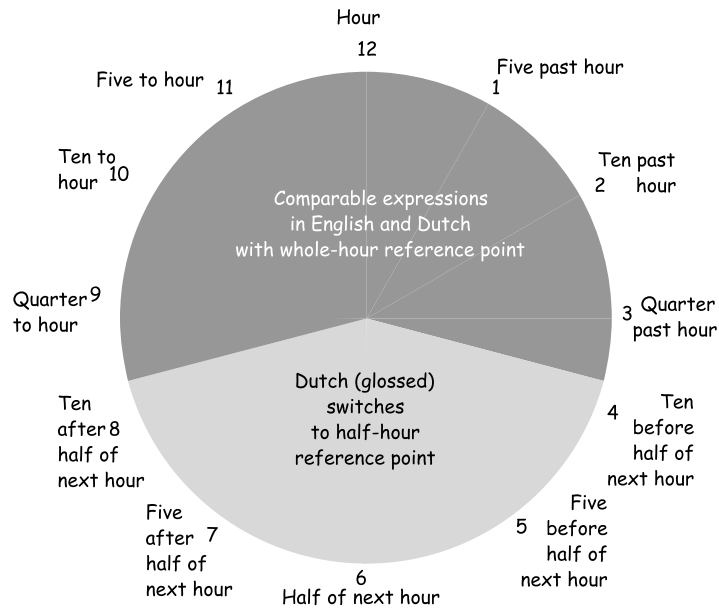before
half of
next hour

6
Half of next hour

Fig. 3. Distribution of five-minute time names in English and Dutch.

within an hourly framework; Dutch speakers tell it within an hourly framework that has a secondary reference point at the half-hour. Fig. 3 illustrates the differences on an analog clock face. The systems are identical during the first and last quarters of each hour, including the *quarter after* and *quarter to* expressions for the fifteen-minute marks. However, they diverge during the second and third quarter-hours. In the second quarter-hour, English continues to count time relative to the hour, so that at five minutes into the second quarter hour the time is 20 past the hour. In Dutch, however, the same time is counted as ten minutes prior to the half-hour (literally translated, the time is ten before the half-hour). At five minutes into the third quarter-hour, English counts the time with respect to the forthcoming whole hour (*twenty-five to*). In Dutch, the time is told as five minutes past the half-hour. Accentuating the difference, the half-hour itself is counted differently: In English, it is half past the previous whole hour, whereas in Dutch it is half of the next whole hour. Thus, what English speakers call *twenty past two* is, in Dutch, *ten before half three*.

During time-telling, these reference points may function like spatial reference frames in guiding visual attention (Logan, 1995). In terms of the information-processing operations in Table 2 above, the clear effect of the difference in reference points is that Dutch speakers have to do something more or something different than English speakers do in order to name minutes, relations, and reference points in time-telling expressions.

To summarize, we aimed to use eyetracking to prospect for the temporal signatures of apprehension and formulation in the production of time expressions. At the same time, we examined whether these signatures change in languages that have different reference point systems in their mundane time expressions.

Overview of the experiments

In the first experiment we elicited time names for analog and digital displays of clock times, to assess preferences for alternative expressions as a function of display type and language. Building on this normative information, Experiments 2 and 3 were designed to trace the broad outlines of the cognitive and linguistic underpinnings for time expressions. The experiments combined eyetracking with measures of speech timing to chart the production of time names. We varied clock formats (analog or digital), expressions produced (relative or absolute), and the duration of the clock displays (100 or 3000 ms), using number-free analog clock faces to maximize the contrast with the digital format. The manipulation of display duration was designed to determine the perceptual role played by the display itself during the production process, as well as the extent to which the production of alternative forms of time expressions depended on the continued accessibility of the display.

In Experiment 2 we tested native speakers of Dutch and American English to explore the impact of different languages and expression types across alternative clock

formats, focusing in particular on whether successive eye movements tended to reflect the perceptual properties of the displays or the linguistic properties of the utterances produced. If a display's visual features have an overriding influence on the behavior of the eyes, we expected initial fixations to go to perceptually prominent or stable locations (such as the analog minute hand or the leftmost digital time window). This would imply that changes in fixation patterns with display format depend on the format rather than on the properties of expressions used to name them. Alternatively, if the behavior of the eyes is systematically influenced by the properties of an upcoming utterance, we expected initial fixations to be directed toward the information in the display that is needed for initiating the utterance, regardless of display type. The comparison between English and Dutch allowed us to compare languages in which the information processing requirements for extracting information from a clock display may differ substantially (see Table 2).

Experiment 3 focused more closely on the effects of expression preferences on the viewing of clock displays. American English-speaking undergraduates viewed alternative displays while producing a single type of expression, as in Experiment 2, or produced alternative expressions while viewing the same type of display. We examined whether strong preferences for one manner of time expression changed the temporal relationship between eye fixations and upcoming words when using disfavored expressions.

The general questions addressed in Experiments 2 and 3 thus had to do with whether and how any observable properties of apprehension and formulation change in response to differences in display and expression format, including differences in expression format associated with different languages. One set of hypotheses emerges from the idea that the perceptible properties of clock displays are paramount in determining how the eyes move when extracting information from the displays for the purpose of language production. If so, similar displays should elicit similar trajectories and timing of eye movements regardless of what kinds of expressions accompany them. An alternative set of hypotheses accords a strong functional role in the control of eye movements to the linguistic demands of telling time. Support for the latter hypotheses would be found in changes in the trajectory and timing of eye movements accompanying changes in expression, with time-locking between the eye's trajectory and the unfolding utterance's composition.

Experiment 1: normative time-telling

Because relative (*ten past three*) and absolute (*three-ten*) expressions are both available for time-telling from

analog and digital clocks but may not be equally favored, the first experiment was designed to determine any preferences for different expressions depending on type of display and language. To assess the expressions that speakers customarily use, we collected norms from native American English and Dutch speakers. Analog or digital displays of whole five-minute times from a 12-h period were presented in a paper-and-pencil task, and participants wrote out the times in words. To prevent routinization of the responses, each participant received only two clock displays. We anticipated that the natural compatibilities between analog displays and relative expressions and between digital displays and absolute expressions would affect the distribution of the expressions that speakers used. Less obviously, there may be different expression preferences in English than in Dutch, and different expression preferences at different times around the hour.

Method

Participants. Norms were collected from 144 undergraduate volunteers at the University of Illinois at Urbana-Champaign and 144 at the Katholieke Universiteit Nijmegen, the Netherlands. The task was presented after a class or while participants were waiting for the start of an unrelated experiment.

Materials. The materials consisted of the 144 five-minute time intervals in the 12-h period between 12:00 and 11:55. All times were depicted in both analog and digital formats, as illustrated in Fig. 2. The analog displays contained no numbers or interval marks. Booklets were assembled by selecting two times at random from the analog displays for one booklet, and using the corresponding times from the digital displays in a separate booklet.

All booklets consisted of a page of instructions and two clock displays, each on a separate page. Beneath each display was printed *What time is it? It's _____* or the equivalent Dutch expression [*Hoe laat is het? Het is_____*].

Procedure. Each participant received one booklet. The instructions indicated that only five-minute time intervals would be shown, and nothing shorter. Participants were asked to tell the time shown in as natural a manner as possible, and to write out their answers in words (to avoid responses such as 3:10).

Scoring. Totals of 288 times were named from analog displays (144 in English and 144 in Dutch) and 287 from digital (144 English and 143 Dutch). There was one missing response.

The responses were scored as *relative* when the minute number or the words *quarter* [*kwart* in Dutch] or *half* preceded the hour number, separated by an appropriate preposition. Responses were scored as *absolute* when the hour number preceded the minute number without an

intervening preposition or, in Dutch absolute expressions, the word for *hour* [*uur*] separated the hour and minute. Scoring was carried out without regard for the accuracy of responses with respect to the depicted times, although the vast majority of responses were correct.

The same criteria were applied in both languages, with one exception. A few of the Dutch participants used English-style relative expressions when naming the times flanking the half-hour, saying the Dutch equivalents of, for example, *twenty past three* and *twenty to four* instead of the more typical *ten before half four* and *ten past half four*. There were 15 of these responses in all, 7 to analog and 8 to digital displays. These were omitted from the scoring.

Virtually all of the whole hours were called [*number*] *o' clock* or [*number*] and their Dutch equivalents. Because these are uninformative about the preference for relative or absolute expressions, all whole-hour responses were excluded from the data.

There were no responses using a 24-hour system (e.g., *twenty-three hours twenty minutes* for a clock showing 11:20). Although 24-hour times are commonly used for official scheduling purposes in the Netherlands, corresponding time expressions are rarely encountered in everyday speech.

Results

Fig. 4 shows the percentages of relative expressions used by Dutch and English speakers for analog and digital clock displays. There was an overwhelming difference between the two languages in the preference for relative expressions, with Dutch speakers tending to use them preponderantly for both types of displays, and English speakers tending to avoid them. However, in both languages there was a small but consistent increase in the use of relative expressions with analog displays.
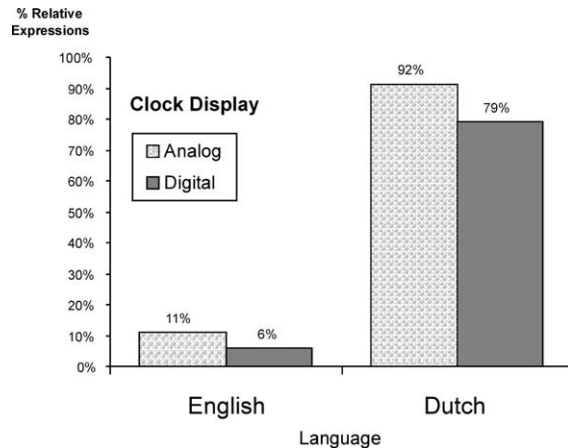
An analysis of variance was performed on these data using language (English or Dutch), display (analog or digital), and five-minute interval (the 11 intervals between :05 and :55) as the independent variables. The use of relative expressions was the dependent variable. Hour time was treated as the random factor, and missing observations were coded as zero. Effects were treated as significant when their associated probabilities were less than or equal to .05.

These analyses confirmed the trends evident in Fig. 4. There were significant main effects of language $(F(1, 11) = 430.3)$ and clock $(F(1, 11) = 41.7)$, as well as a significant interaction between language and clock $(F(1, 11) = 10.8)$. There was also a significant main effect of five-minute time $(F(10, 110) = 3.9)$ without interactions involving language or clock type (all $Fs < 1$). Cardinal times (the whole, half, and quarter hours) were named more consistently than other times in both languages, eliciting predominantly the modal expressions for each language. So in English, absolute expressions were used for 97% of the cardinal times. In Dutch, relative expressions analogous to *quarter past* and *half past* were used 89% of the time, with the only deviations from such expressions occurring to digital displays.

For the remaining five-minute times, Fig. 5 illustrates the distributions of time expression use. In addition to the overall disparities in the use of relative expressions for the two languages and clock types, the figure reveals

Fig. 4. Percentages of relative expressions (e.g., *ten past three*) used when naming times from analog and digital clock displays.
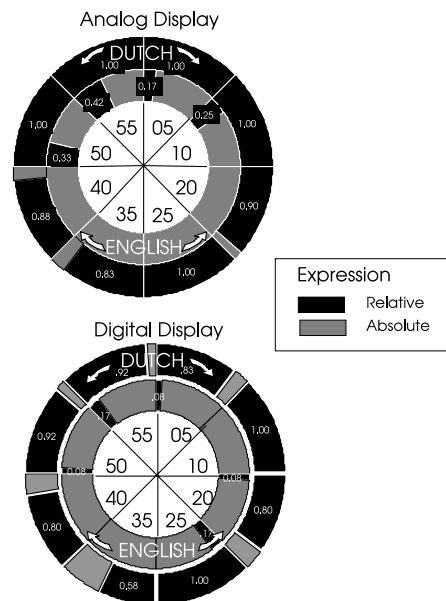

Fig. 5. Proportions of relative time expressions (in black) to absolute time expressions (in gray) used in Dutch and English when naming noncardinal times at five-minute intervals (5, 10, 20, 25, 35, 40, 50, 55) around the hour from analog (top) and digital (bottom) clock displays.

a tendency to deviate from modal expressions most often for five-minute times in the immediate neighborhood of reference points, primarily flanking the 12 o'clock region in English and the 6 o'clock region in Dutch. Note that these are the points where the hour number changes, at the hour itself in absolute expressions (the English modal expression) or ten minutes before the half-hour in Dutch relative expressions (the Dutch modal expression). We averaged the deviations that occurred within plus or minus 10 min of the reference points and compared them to the deviations that occurred elsewhere (excluding the cardinal times). Overall, nonmodal expressions were used on .14 of the trials in the vicinity of a reference point change and on .04 of the trials elsewhere.

Discussion

The norms disclosed three clear trends in how time is told in Dutch and English. In both languages, analog displays elicited more relative expressions (e.g., *ten to two*) than digital displays did. This reflects the compatibility between the physical stimulus and the preferred response format. Analog displays provide a degree of perceptual grounding for relative expressions in terms of the spatial geometry and direction of motion of the hands, giving interpretations to prepositions such as *till*, *to*, *before*, *after*, and *past* and their Dutch equivalents *voor* and *over*. Digital displays do not. Still, the effects of this compatibility were fairly weak.

There was also an effect of reference points. Excluding the cardinal times, which were named consistently in both languages, five-minute times in the vicinity of the points where the reference hour changes (at the hour in absolute expressions and the half-hour in relative expressions) tended to elicit more variable expressions than other times. This occurred with both display formats and in both languages, despite the fact that Dutch and English expression preferences were opposed and the relevant reference points differed. This suggests that the effect is neither simply perceptual nor simply linguistic, but arises in the mapping between seeing and saying.

The third effect was large and unexpected. American undergraduates used absolute expressions almost exclusively; Dutch undergraduates somewhat less exclusively used relative expressions. Casual observations prompted by this finding confirmed that American undergraduates indeed rely on absolute expressions regardless of the circumstances. We can only speculate at the reason for this disparity. Perhaps it is the ubiquity of digital clocks and watches in the United States compared to the Netherlands, or differences in the teaching of time expressions in the two countries. Regardless, it is clear that college-age speakers of the languages have opposite preferences for time expressions. What is fre-

quent in Dutch time-telling is overwhelmingly different from what is frequent in contemporary American English.

Experiment 2

In Experiment 2 we turned to the perceptual and linguistic dynamics associated with the production of time expressions. Dutch and American English speakers told all of the 144 five-minute times from a twelve-hour clock while their eye movements were monitored. Each speaker told all the times once from an analog and once from a digital display, in a blocked arrangement, to assess how display changes affected the coupling of eye movements with speaking. Because the outcomes of the time-norming study indicated that it would be impracticable to allow speakers to freely name the times, half the speakers of each language were encouraged to use relative and the other half absolute expressions.

The displays were shown to speakers for either 100 or 3000 ms, with different speakers receiving just one duration in the course of the experiment. The purpose of this manipulation was to diagnose the role played by eye movements during time telling. Because 100 ms is too brief for initiating a saccade, and the first saccade masks the information in iconic memory (Irwin, 1992), the apprehension of a visual clock time should be seriously disrupted to the degree that the uptake of information depends on successive fixations to the elements of the display.

A comparison between different types of expressions (relative and absolute) and different languages (Dutch and English) was made between speakers. This allowed an assessment of the roles played by different orders of mention of time elements and different reference points. Relative and absolute expressions reverse the order in which minutes and hours are named, so that changes in eye movements and associated changes in the timing of speech as a function of linguistic factors can be easily explored. The existence of a different mundane-time framework in Dutch, with a secondary reference point at the half-hour in relative expressions, made it possible to explore how reference points affect how clocks are viewed.

The main questions had to do with the stability of eye movements over different types of displays when speakers used the same expressions with both. Visually, the angular geometry of analog clocks is incommensurate with the left-to-right number display of digital clocks. If display properties matter more to the systematic behavior of the eyes than the psycholinguistic processes associated with the production of language, we expected to observe changes in the relationship between eye movements and speech accompanying different clock displays.

Method

Participants. The participants were eight native Dutch speakers and eight native American English speakers recruited from staff and visitors of the Max Planck Institute for Psycholinguistics in Nijmegen, the Netherlands. All were volunteers.

Materials. The displays were analog and digital clock faces similar to those used in Experiment 1, but without frames around the digital number windows. All 144 five-minute times between 12:00 and 11:55 were presented in both clock formats.

Equipment and procedure. The clock faces were displayed on a 39 cm ViewSonic 17PS screen while participants' eye movements were monitored with an SMI Eyelink-HiSpeed 2D headband-mounted eyetracker. The eyetracker was controlled by a Pentium 90 MHz computer. The experiment was run under the Nijmegen Experiment Setup (NESU) with a NESU button box on a Pentium 400 MHz computer. Participants' utterances were recorded over a Sennheiser ME400 microphone to a Sony DTC55 digital audio tape recorder for later transcription.

Participants were seated in front of the monitor, outfitted with the eyetracker headband, at a viewing distance from which the analog and digital clock faces subtended 35° and 27° of visual angle, respectively, and the digits on the digital clock were 7.5° tall. The eyetracker was calibrated at the beginning of the experimental session. Drift correction occurred automatically after every 12 trials. At the onset of each trial a fixation point was displayed in the center of the display for 1 s. In the analog clock conditions the center of the display was the midpoint of the clock face; in the digital conditions the center was offset slightly into the minute region of the clock face.

Fig. 6 illustrates the sequence of trial events. At the disappearance of the fixation point the clock face was displayed for 100 or 3000 ms, and participants named the time as instructed. Over both display durations, eye movements were monitored beginning with the appearance of the fixation point and continuing through
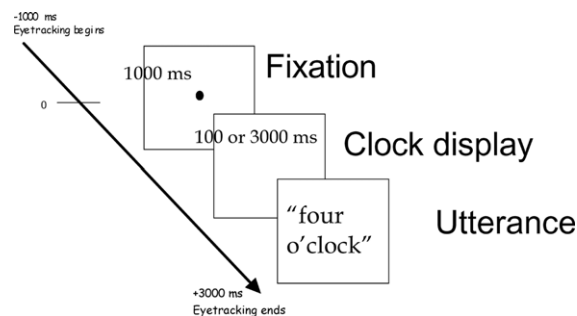

Fig. 6. Sequence and timing of trial events in Experiments 2 and 3 (English example).

3000 ms after its disappearance. Four seconds after the onset of the display the message *Press the button to proceed* appeared, signaling the participants that they could go on to the next trial when ready, after pressing a button to initiate the next fixation point.

Display duration was a between-participants variable, with half of the participants receiving the 100 ms view and the other half receiving the 3000 ms view. Expression format (relative or absolute) was also varied between participants. Type of display (analog or digital) was blocked within participants, with the order of blocks counterbalanced across participants. Within each block all 144 of the five-minute times were displayed once, with the orders of presentation randomized. Four different randomizations of the 144 times were created and applied both to the analog and to the digital clocks. Each participant received a different randomization of the analog and digital clock faces.

Participants read a set of instructions prior to each block. Before the first block, they were told to focus on the fixation point until the clock appeared, and then to say what time was displayed. They were asked to try to produce fluent utterances and to avoid inadvertent vocal noises (like lip-smacking and swallowing).

The participants were also instructed about the types of expressions to use and the types of expressions to avoid. Two examples were given, with two clock faces in the format to be presented in the upcoming block. One clock showed 9:55 and the other 11:15. The instructions for relative expressions said "for the clocks shown here, you would say 'five to ten' and 'quarter past eleven' and not 'nine fifty-five' or 'eleven-fifteen.'" The instructions for absolute expressions were just the reverse, to "say 'nine fifty-five' and 'eleven-fifteen' and not 'five to ten' or 'quarter past eleven.'" Before the second block, participants were told that the times would be displayed differently, and the same two example times were shown in the new clock format. The instructions were to use the same expressions as in the first part of the experiment, and the positive examples were given again. In the relative condition these were *five to ten* and *quarter past eleven* and in the absolute condition *nine fifty-five* and *eleven-fifteen*. The negative examples were not repeated.

Equivalent instructions were given in English and in Dutch. After the instructions a series of 12 practice items was presented, drawn equally from the cardinal times (with three instances each of whole hours, half hours, quarter hours, and three-quarter hours).

Design. Each of the 16 participants viewed all 144 five-minute times twice, once on analog and once on digital clock faces. Half of the participants spoke English and half spoke Dutch. Within the languages, half of the speakers produced relative and half produced absolute expressions, and within expression types, equal numbers of speakers viewed the clock displays for 100 and 3000 ms.

Eye-movement analysis. Regions were defined separately for the analog and digital clock faces. On the analog clocks there were 12 time regions made up of 30° slices centered on the radius connecting each five-minute point to the clock center. For example, the boundaries of the :05 region were (a) the radius between the clock center and the clock edge at 15° (2 min 30 s) and (b) the radius between the clock center and the clock edge at 45° (7 min 30 s). On each analog trial, the time-relevant hour region was the region in which the hour hand occurred and the time-relevant minute region was the region in which the minute hand occurred. For example, when an analog clock read 2:20, the time-relevant regions were those encompassing the canonical locations of the 2 (between 45° and 75°) and the 4 (between 105° and 135°) on a normal numbered clock face. Fixations in the 12 o'clock and 6 o'clock regions were also tabulated for times when neither hand occupied these regions. Fixations within 2° of the clock center (in the immediate vicinity of the fixation point) were classified as *Center* and not included in the time regions, and fixations in regions outside of the time-relevant ones were classified as *other*.

On the digital clocks the four regions consisted of the four number windows, which were equal in size. From left to right, these windows were the hour 10 s window, the hour 1 s window, the minute 10 s window, and the minute 1 s window. All four regions were considered time-relevant. Fixations were excluded from analysis when they fell outside a bounding box that subtended 27° horizontally and 16.5° vertically.

The eyetracking system's software identified saccades and measured the durations of fixations between them. The system defined a saccade as a change in eye position that covered more than .15° at a velocity greater than 30°/s with acceleration greater than 8000°/(s²).

Speech analysis. Speech onset latencies for all utterances were measured by the NESU voicekey during the experimental sessions. The forms of the expressions were coded as relative or absolute according to the criteria described in Experiment 1. Responses were excluded when they were too fast or too slow (less than 300 ms or greater than 3000 ms), erroneous, disfluent, or deviant. Errors consisted of mistaken minute names. Disfluent utterances contained filled pauses, noticeable hesitations, or noises that erroneously triggered the voicekey. Deviant responses departed from the specified relative or absolute format. We also excluded the whole-hour trials and trials for the 12 times that have overlapping hour and minute hands on the analog display (e.g., 1:05, 2:10, 3:15, etc.). The former times lack a minute expression and the latter lack separable visual targets on analog clocks.

Results

The principal findings emerged from analyses of the speech and eye-movement data separately and then jointly. We present the results from speech onset latencies, from the patterning of eye fixations on time-relevant regions, and from the relative timing of speech onsets with respect to eye movements. Analyses of variance were carried out on measures of latency, fixation locations and timing, and the relative timing between eye movements and speech. Except as noted, the designs for the analyses treated participants as the random factor and display (analog or digital) as a within-participants factor crossed with the between-participants factors of display duration (100 or 3000 ms), expression (relative or absolute), and language (English or Dutch). Here and in Experiment 3, effects were regarded as significant when their associated probabilities were less than or equal to .05 and marginal when greater than .05 but less than or equal to .10.

How fast and how accurate is time naming? To assess how display and expression properties matter to time naming, we examined the latencies to begin producing relative and absolute time expressions from analog and digital clock displays. Table 4 shows that responses were faster to digital clocks in all cases, but the size of the digital advantage changed as a function of response

Table 4
Latencies to produce relative and absolute expressions from analog and digital clocks in Dutch and American English (Experiment 2)

Expression and clock format	Language				Means
	Dutch		English		
	100 ms display	3000 ms display	100 ms display	3000 ms display	
Relative expressions					
Analog clock	1259 (81)	1337 (91)	1143 (70)	1642 (76)	1345
Digital clock	856 (90)	895 (97)	885 (81)	1074 (93)	928
Absolute expressions					
Analog clock	1705 (75)	1782 (70)	1410 (55)	1573 (64)	1618
Digital clock	608 (98)	591 (99)	580 (97)	617 (96)	599
Means	1107	1151	1004	1226	

Note. Shown in parentheses are the percentages of trials in each condition (out of 242 possible) included in all analyses.

format. In particular, the advantage for digital clocks was larger for absolute than for relative expressions. With analog clocks, relative expressions were produced faster than absolute. In line with these trends, an analysis of variance on the latencies showed significant effects of display format ($F(1, 8) = 267.4$) and a significant interaction between display format and expression type ($F(1, 8) = 46.8$). A general tendency to respond faster in the 100 ms condition was also significant ($F(1, 8) = 6.7$). There were no notable effects of language, although the interaction with expression type was marginal ($F(1, 8) = 4.8$), reflecting Americans' somewhat faster production of absolute expressions. The language difference was more in evidence for analog than for digital clocks, but the interaction with display was far from significant ($F(1, 8) = 1.9$).

Table 4 also shows the percentages of the observations in each condition that were included in the analyses. In total, 16.7% of the trials were excluded, 9.3% because of clock-reading errors, an additional 5.7% because of disfluency or deviation from the expected expression formats, 1.3% for responses that were too slow or too fast, and the remainder for fixations falling outside the defined clock regions. Analyses of variance on the numbers of acceptable responses were conducted to evaluate whether they patterned in the same way as the onset latencies. Display and the interaction between display and expression type were both significant ($F(1, 8) = 94.9$ and $F(1, 8) = 23.0$, respectively), again reflecting the ease of the digital display and the compatibility of analog displays with relative expressions and digital displays with absolute expressions. The only significant effect that was not also present in onset latencies was due to the larger number of trials excluded for Americans than for Dutch speakers ($F(1, 8) = 11.7$). The difference was larger for analog than for digital displays, which yielded a marginally significant interaction between display and language ($F(1, 8) = 4.0$). No other differences approached significance.

Where do speakers look when telling time, and how fast do they look? The effects of display and expression differences on speech onset imply that both mattered to the performance of time telling. The next question was *how* they mattered. To address this, we examined whether display and expression properties systematically influenced the pattern of eye movements.

One picture of the uptake of visual information comes from fixations on specific time-relevant regions. These fixations revealed strong effects of display without significant modulation by language factors. Table 5 gives the mean number of eye-fixations per participant within the analog and digital clock displays for both expression types in both languages. On analog clocks, the time-relevant regions comprised (a) the region in which the hour hand occurred, (b) the region in which the minute hand occurred, and (c) the 12 o'clock and 6

o'clock regions. On digital clocks all regions (the hour and minute windows) were time-relevant. In the 100 ms conditions, the time-relevant regions were the ones that had contained the information during the display. Analyses of variance were carried out separately for the analog and digital clocks on the proportions of time-relevant fixations within regions, with region as a factor. For analog clocks the region factor had four levels (hour, minute, 6 o'clock, and 12 o'clock) and for digital clocks two levels (hour and minute).

On analog clocks, 24% of all fixations occurred on the hour hand and 16% on the minute hand. This produced a main effect of region ($F(3, 24) = 67.7$). The proportion of time-relevant fixations was greater at 3000 ms than at 100 ms ($F(1, 8) = 78.2$) and this interacted with region: The increase in the proportion of fixations at 3000 ms was confined to the hour and minute regions ($F(3, 24) = 26.5$), and did not occur at 12 and 6 o'clock. On the digital clocks 98% of the fixations were within the four time windows (.995 at 3000 ms and .97 at 100 ms), with 96% to the two inner windows (the hour 1 s and the minute 10 s) and 2% to the outer windows. This was the only significant effect in the distribution of fixations over windows ($F(3, 24) = 47.2$). None of these effects interacted significantly with language or expression type, and the main effects of language and expression were likewise not significant.

Fixations varied not only in where they were but in when they happened. Taking *when* into account revealed that the timing of fixations was more sensitive to variations in expression than to variations in display. For the 3000 ms displays, speakers initiated saccades to time-relevant regions very early in processing, with first fixations on the hour or minute occurring on average 359 ms after display onset. This did not vary significantly across language, display, or expression. The timing of the onset of first fixations to specific hour and minute regions did differ according to the nature of the expression used to tell the time. Table 6 gives the mean latencies of the hour and minute fixations for all trials on which hour and minute fixations both occurred. The table shows that for relative expressions, fixations to the minute tended to precede fixations to the hour, and vice-versa for absolute expressions. An analysis of variance of differences (minutes minus hours) in fixation onset times revealed a significant effect of expression, $F(1, 4) = 42.0$. For relative expressions, the onset of fixations in minute regions preceded the onset of fixations in hour regions by 267 ms, whereas for absolute expressions the onset of fixations in minute regions *followed* those in hour regions by a mean of 203 ms. Neither the effect of language nor any interactions were significant (all $Fs < 1.05$), despite one English speaker who, while producing relative expressions to analog clocks, tended to look at the hour hand before the minute hand. Due to the eccentric location of the fixa-

Table 5
Mean numbers of fixations to regions on analog and digital clocks (Experiment 2)

Display format and time-relevant region	3000 ms		100 ms	
	Dutch	English	Dutch	English
Analog				
Relative expressions				
Hour hand	456	358	81	146
Minute hand	308	237	70	75
12 o'clock	83	34	62	38
6 o'clock	11	35	4	14
Center	86	122	278	310
Other	255	293	222	251
Absolute expressions				
Hour hand	378	334	24	41
Minute hand	238	231	17	24
12 o'clock	44	10	9	20
6 o'clock	18	17	8	10
Center	87	235	235	157
Other	190	116	62	127
Digital				
Relative expressions				
Hour 10 s window	2	4	0	2
Hour 01 s window	322	343	287	154
Minute 10 s window	440	460	372	516
Minute 01 s window	16	62	1	1
Other	0	2	0	0
Absolute expressions				
Hour 10 s window	2	10	0	0
Hour 01 s window	314	378	76	220
Minute 10 s window	523	608	322	428
Minute 01 s window	10	56	0	0
Other	9	6	2	86

Table 6
Mean onset latencies (in ms) for first fixations to minute and hour regions for relative and absolute expressions with 3000 ms displays (Experiment 2)

Expression type	Language and clock format			
	Dutch		English	
	Digital	Analog	Digital	Analog
Relative expressions				
Minute (term 1)	574	618	490	700
Hour (term 2)	1060	737	974	677
Absolute expressions				
Minute (term 2)	587	940	546	756
Hour (term 1)	588	484	506	442

tion point in digital displays (i.e., within the minute region), these asynchronies varied artifactually with display ($F(1, 4) = 25.8$).

The combination of where and when fixations occurred gives a picture of the early selectivity of visual processing that we attribute to apprehension. Fig. 7 shows how fixations to the hour and minute regions diverged, plotting the proportions of fixations within complementary hour and minute areas of the 3000 ms analog and digital displays. The observations for the two languages and upcoming expression types are pooled. The figure suggests that by 400 ms after display onset
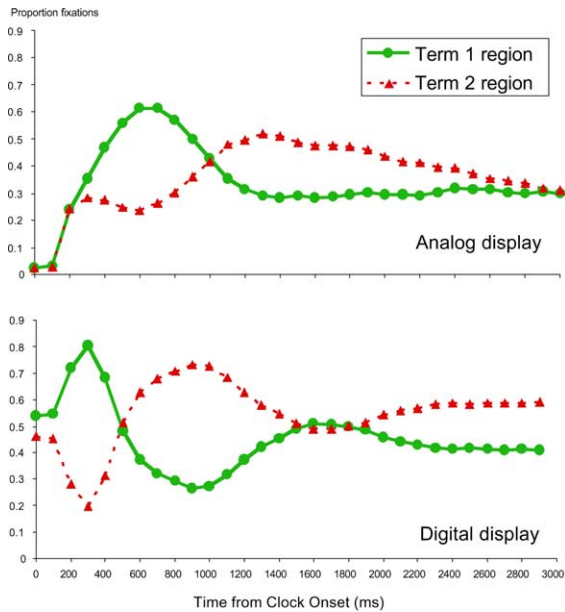
Fig. 7. Proportions of fixations to analog and digital clock regions carrying information needed for the first (term 1) and second (term 2) terms of time-telling expressions, in 100 ms increments from display onset (Experiment 2).

(and by 200 ms for digital displays), saccades tended to be directed toward the locations containing the information needed to produce the first term of the upcoming time expression. In the graphs, the term-1 region corresponds to the locations of (a) hour information for upcoming absolute expressions and (b) minute information for upcoming relative expressions, and the term-2 region to the locations of (c) minute information for upcoming absolute expressions and (d) hour information for upcoming relative expressions.

To confirm the trends suggested in Fig. 7, analyses of variance were performed on the proportions of fixations within the term-1 and term-2 regions during the first six 100 ms intervals from display onset in the 3000 ms display condition. Interval, language, display type (analog or digital), and expression (relative or absolute) were the factors. The term-1 and term-2 regions were analyzed separately because of the lack of independence between them. There were no significant effects of language or main effects of expression. However, in both analyses there were significant effects of interval and significant interactions of interval with display and expression type. Planned comparisons between the cell means at each interval were carried out using the 95% confidence limits calculated from the mean square errors for the interactions. In the term-1 regions, the proportions of fixations increased significantly between 100 and 200 ms for the digital displays, and between 200 and 400 ms for the analog displays. In the term-2 regions, the only significant effect at the same intervals was the complementary decrease in fixations on the digital display. The interactions with expression were again attributable to the off-center fixation point in digital displays.

How long do speakers look at time-relevant regions? The amount of time devoted to display processing can be assessed in terms of the number and duration of looks to clock regions, a measure analogous to first-pass reading time in research on language comprehension. In the 3000 ms condition the successive fixations and intervening saccades within a time-relevant region before leaving the region were aggregated into gazes. Before speech onset there were 1.02 gazes in hour regions and 1.13 in minute regions, on average, and the only significant difference in this measure arose in an interaction between display and time region $(F(1,4) = 13.2)$ as another artifact of the eccentric digital fixation point.

The gaze duration data are summarized in Table 7. There were significant effects of display and region but not language. Overall, first gazes were longer in analog than in digital displays (435 to 336 ms; $F(1,4) = 23.8$) and in term-1 regions than in term-2 regions (485 to 286 ms, $F(1,4) = 29.1$). The interactions of display and expression $(F(1,4) = 43.7)$ and of display, expression, and region $(F(1,4) = 34.2)$ were both significant. Display had a bigger impact on the first gazes associated with absolute than relative expressions, but this impact varied depending on whether the gazes were in the term-

Table 7

Durations of first gazes (in ms) to minute and hour regions for relative and absolute expressions (Experiment 2)

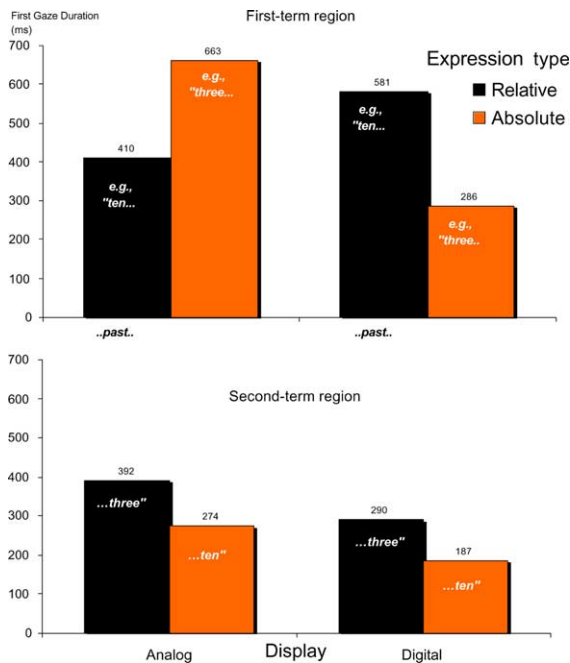| Expression type	Language and clock format			
	Dutch		English	
	Digital	Analog	Digital	Analog
Relative expressions				
Minute (term 1)	485	382	677	438
Hour (term 2)	287	415	294	368
Absolute expressions				
Minute (term 2)	200	246	174	302
Hour (term 1)	268	732	303	594

Fig. 8. Durations of first gazes to analog and digital clock regions relevant to the production of the first and second terms (illustrated with examples) in relative and absolute time expressions (Experiment 2).
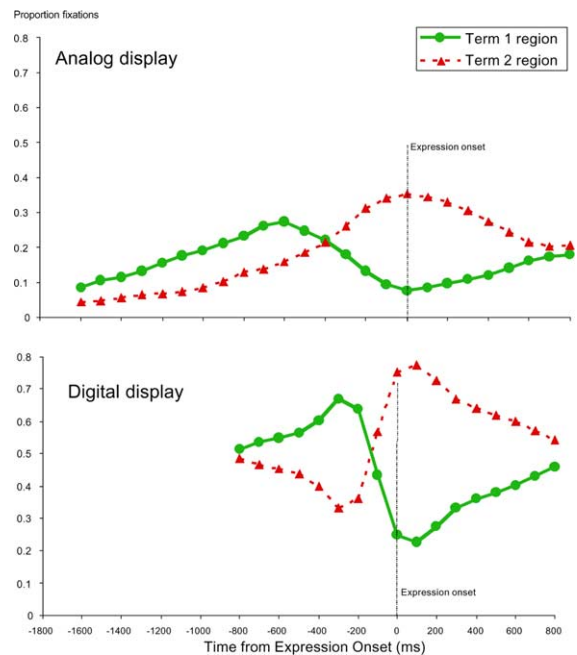

Fig. 9. Proportions of fixations to first-term-relevant and second-term-relevant clock regions before and after the onset of time expression (Experiment 2).

1 or term-2 regions. Whereas the differences in the term-2 regions were small, nonsignificant, and consistent with simple effects of display and expression type, in the term-1 regions they were substantial. Fig. 8 shows the interaction. It reveals that for analog displays, the first gazes associated with relative expressions were significantly shorter than those associated with absolute expressions. For digital displays the pattern reversed, with first gazes associated with absolute expressions significantly shorter than those associated with relative expressions. Total gaze durations (total time spent in time-relevant regions) revealed the same trends.

How is speech timed with respect to eye movements? The interplay of visual and verbal processes reflects how visual information is used when formulating time expressions. This can be captured in terms of the timing of eye movements with respect to speech. The general relationship between eye fixations and speech onsets is seen in the distributions of fixations within time-relevant regions relative to when the terms corresponding to the regions began to be named. For the 3000 ms display condition, Fig. 9 summarizes the proportions of fixations within the term-1 and term-2 regions relative to the onsets of expressions, plotted from the average onsets of the analog and digital displays. Fixations to the term-1-relevant regions peaked 600 ms before utterance initiation for analog displays and 300 ms before for digital

displays. In comparison, fixations to the term-2-relevant regions peaked at utterance initiation for analog displays and 100 ms later for digital displays.

The timing between eye fixations and speech can be quantified and analyzed in terms of eye-voice spans. The eye-voice span is the length of the interval from a fixation within a time-relevant region to the onset of a time number. For the time expressions produced in the 3000 ms display condition, Table 8 gives the eye-voice spans from first fixation onsets for all trials on which fixations to both the hour and minute regions occurred prior to expression onset.

Analyses of variance were performed on these spans with language, expression, and clock format as independent variables. For initial fixations, the mean span from fixation onset in the term-1 region to expression onset was 639 ms, but the spans were significantly shorter overall for digital (128 ms) than for analog clocks (914 ms), $F(1, 4) = 38.2$. The right-of-center location of the fixation points for the digital displays created a significant interaction between display, region, and expression ($F(1, 4) = 25.9$), but all remaining effects were nonsignificant, including those of language.

The mean eye-voice spans from final fixation offset (that is, the offset of the final fixations to time-relevant regions prior to expression onset) in the term-1 region was 313 ms. These offset spans were significantly longer for analog than for digital displays ($F(1, 4) = 138.0$).

Table 8
Mean eye-voice spans (in ms) from fixations on minute and hour regions to onset of time expressions (Experiment 2)

Fixations and expression types	Language and clock format			
	Dutch		English	
	Digital	Analog	Digital	Analog
From onset of initial fixation in minute region to onset of				
Relative (minute-hour)	320	714	590	938
Absolute (hour-minute)	2	842	70	824
From onset of initial fixation in hour region to onset of				
Relative (minute-hour)	−167	594	106	962
Absolute (hour-minute)	1	1298	112	1138
From offset of final fixation in minute region to onset of				
Relative (minute-hour)	262	478	238	439
Absolute (hour-minute)	74	125	81	130
From offset of final fixation on hour region to onset of				
Relative (minute-hour)	106	66	50	162
Absolute (hour-minute)	126	420	168	374

Note. Negative numbers indicate that the initial fixation on the region followed expression onset.

How are temporal reference points on analog clocks used during time telling? To examine performance around the reference point times for Dutch and English (see Fig. 3), we first looked at whether fixations clustered near reference points on analog displays. An analysis of variance was performed on the proportions of fixations that occurred to the top (between 10 o'clock and 2 o'clock), bottom (between 4 o'clock and 8 o'clock), and middle (between 2 and 4 and between 8 and 10) of the clocks. There were somewhat more fixations to the top (.36 in English and .41 in Dutch) than to the middle (.32 in English and .30 in Dutch) or bottom (.32 in English and .29 in Dutch), but none of these effects or their interactions with language, expression, or duration were significant.

We then repeated key analyses on subsets of the data. Specifically, we compared measures for the combined times at five past, ten past, ten before, and five before the hour (*hour-centered times*) with the combined times at ten before, five before, five past, and ten past the half-hour (*half-hour centered times*). Fig. 10 gives the speech onset latencies for the hour-centered and half-hour-centered times (pooling the data from the 100 and 3000 ms display conditions, where none of the interactions with region were significant). Table 9 gives the first-gaze durations for the term-1 and term-2 regions from the 3000 ms condition, and Fig. 11 highlights the term-1 contrasts. Analyses of variance on these data produced the same general pattern of results as in the full data set, so we report only effects that were specific to the whole- and half-hours.

For speech onset latencies and first gaze durations, slowing occurred at half-hour centered times relative to hour-centered times. This effect was significant for
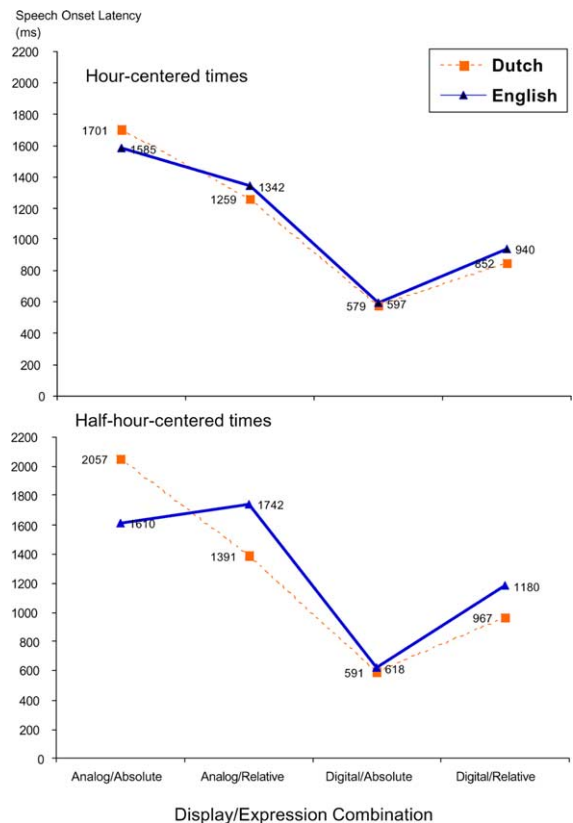

Fig. 10. Onset latencies for times flanking the hour and half-hour in Dutch and American English.

Table 9
Gaze durations in term-1 and term-2 relevant regions for hour-centered and half-hour-centered times in Dutch and American English

| Expression and clock format	Center point and language			
	Hour-centered times		Half-hour-centered times	
	Dutch	English	Dutch	English
Term 1				
Relative expressions (minute)				
Analog clock	476	500	344	470
Digital clock	460	602	528	782
Absolute expressions (hour)				
Analog clock	631	520	785	638
Digital clock	272	289	262	324
Term 2				
Relative expressions (hour)				
Analog clock	340	352	366	372
Digital clock	286	303	323	278
Absolute expressions (minute)				
Analog clock	270	355	242	300
Digital clock	200	173	204	164
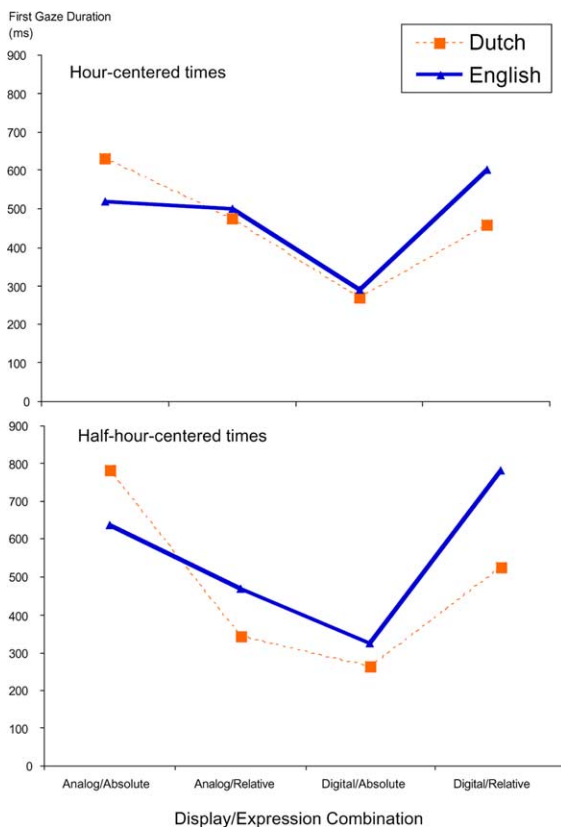
Fig. 11. First-gaze durations in term-1 regions for times flanking the hour and half-hour in Dutch and American English (3000 ms displays only).

speech onsets (1178 to 1096 ms; $F(1, 8) = 22.2$) but not first-gaze durations (517 to 468 ms, $F(1, 4) = 3.9$). The impact of incompatibility between displays and expressions (i.e., the combination of absolute expressions with analog displays and relative expressions with digital displays) was larger for the half-hour centered compared to the hour-centered times. This created a significant three-way interaction among display, expression, and time-center for first-gaze durations ($F(1, 4) = 10.2$) but not voice onsets ($F(1, 8) = .15$). For voice onsets, the interaction between time-center and expression was significant ($F(1, 8) = 22.2$), reflecting the comparatively larger difference between latencies at half-hour and hour-centered times for relative expressions (1223 ms vs. 1094 ms) than for absolute expressions (1133 ms vs. 1098 ms).

The figures suggest a somewhat greater penalty for incompatibility when each language's disfavored expression was used with an incompatible display, particularly at the half-hour. Specifically, Dutch and English speakers were both at a disadvantage when using absolute expressions with analog clocks, but Dutch speakers appeared to be at a greater disadvantage than the Americans. Likewise, both groups were at a disadvantage when using relative expressions with digital clocks, but the Americans' disadvantage was numerically greater. However, none of the interactions involving language were significant for either speech onsets or first-gaze durations in either the term-1 or term-2 regions (all Fs < 2.0).

Eye-movement trajectories for times around reference points may also reveal whether reference point infor-
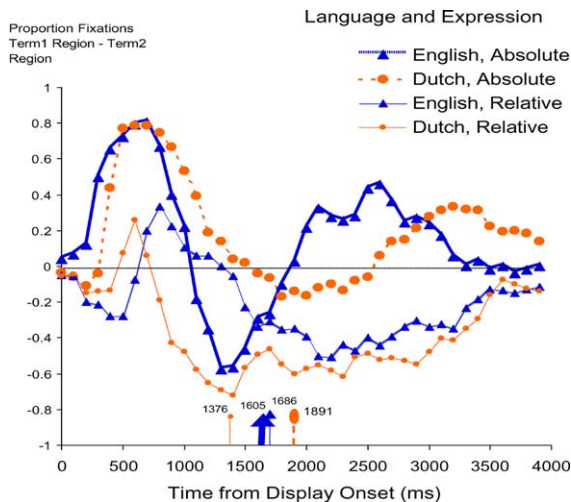
Fig. 12. Differences in proportions of fixations to term-1 and term-2 relevant regions for half-hour centered times in Dutch and American English (3000 ms displays only) in milliseconds from display onset, with mean voice onset times flagged.

mation comes into play. Fig. 12 plots the half-hour centered fixation data, where the compatibility of expressions and clocks appeared to matter most. Shown are the differences in the proportions of fixations to the term-1 and term-2 relevant regions for the 3000 ms analog displays at each 100 ms interval after display onset, depending on expression used. A positive difference indicates a preponderance of fixations in the term-1 region, and a negative difference indicates a preponderance in the term-2 regions. The mean expression-onset latencies for the conditions shown are posted along the abscissa, for reference. The figure reveals similar patterns of fixations in the two languages, including a trend toward fixating the term-2 (hour) region before the term-1 (minute) region when preparing to produce relative expressions. Inspection of the cell means for the first-fixation latencies on analog displays confirmed that this occurred only for half-hour-centered times with relative expressions. In these cases, the average latency of first fixations in the term-1 (minute) region was 752 ms compared to 682 ms in the term-2 (hour) region. In all other conditions, for both languages, both displays, both expressions, and both hour- and half-hour centered times, fixations in the term 1 region preceded fixations in the term-2 region, with mean fixation onset latencies of 516 and 807 ms, respectively.

Discussion

From this first foray into the relationship between seeing and saying clock times, it is clear that there was a strong connection between the direction pursued by the eye and the impending elements of a time expression.

The early-targeted regions of clocks corresponded not to perceptually or conceptually prominent locations, but to the locations containing the information needed to produce the first term of the upcoming expression. For the fluent utterances that we examined in this experiment, this connection was readily observable for both analog and digital displays, suggesting that it is not dependent on the perceptual format of the clock. The nature of the time expressions, however, produced a near-complete inversion in the trajectory of speakers' initial eye movements. When hours preceded minutes, as in absolute expressions, the eye first sought the area of the display containing the hour information, whether it was the hour hand or the hour window. Conversely, when minutes preceded hours, as in relative expressions, the eye first sought the area of the display containing minute information.

The strength and stability of this relationship might invite the speculation that the uptake of visual information made possible by selective eye movements is an integral part of normal time telling. To the contrary, the results from the 100 ms exposure conditions indicated that sufficient information for fairly accurate production can be extracted from the display without fixating the crucial information directly. This supports the interpretation that in more extended viewing, what controls the initial saccade can be a coarse representation of the scene. Such a representation evidently becomes available in a glimpse (Biederman, 1981; Gordon, 1999; Potter, 1975) and offers basic information from which an utterance can begin to be constructed. This is what we term *apprehension*, and it seems sufficient to direct the eye to where the details required for filling out an utterance scheme (or other action plan; Hayhoe, 2000) can be found.

When a scene is as familiar as a clock face, such details may be less essential than for novel or more variable scenes. Strikingly, however, discrete, selective eye movements to time-relevant regions were readily apparent for both clock formats. In all conditions there was a general tendency to inspect time-relevant regions in close anticipation of the elements of time expressions. This is in line with the hypothesized incrementality of linguistic formulation, and with the argument that the playing out of the formulation process is reflected in accompanying eye movements.

The first gaze in time-relevant regions seemed to play a special role, illustrated in Fig. 8. The first gaze was consistently longer than the second gaze, and its duration reflected the compatibility between the clock format and the upcoming time expression. Specifically, prior to the production of relative expressions, first gazes to the term-1 regions on analog clocks were shorter than the first gazes to the term-1 regions on digital clocks. Conversely, prior to the production of absolute expressions, first gazes to the term-1 regions on digital clocks were shorter than the first gazes to the term-1 regions on

analog clocks. The term-2 regions revealed none of these compatibility effects.

As depicted in Fig. 9, the coupling between eye movements and the elements of time expressions was orderly for both clock formats despite the incommensurate perceptual properties of analog and digital displays. The coupling was apparent for relative as well as absolute expressions and in both Dutch and English.

There were few differences worthy of note between the languages. Against the large disparity in the kinds of expressions preferred by Dutch and English speakers observed in Experiment 1, there were small, largely nonsignificant differences in the latencies to produce preferred expressions. What differences there were occurred primarily in conjunction with perceptually incompatible displays. Dutch speakers were somewhat faster than Americans to produce relative expressions to digital displays (876 to 980 ms) and Americans were somewhat faster than the Dutch to produce absolute expressions to analog displays (1492 to 1744 ms). Even for those times when the secondary reference point comes into play, around the half-hour, Dutch time-telling proceeded much like English time-telling. Dutch speakers were no more likely to look at the secondary reference point than English speakers were (in fact, in raw terms they were less likely to look at it), nor were they more likely to fixate the lower regions of the clock overall.

The absence of differences is somewhat surprising, given that Experiment 1 suggested such powerful differences between American and Dutch college students in favored time expressions. At most, there were hints that expression preferences may have modulated the effects of incompatible combinations of expressions with display, but the experiment was not designed in a way that allowed the effects of expression preferences on time-telling to be sensitively assessed. Experiment 3 aimed to resolve some of this uncertainty.

Experiment 3

The signatures of apprehension and formulation observed in Experiment 2 imply that even formulaic utterances undergo a compositional assembly process during which words are positioned in phrases during language production. In Experiment 3, the generality of these effects was put to the test by requiring speakers to change expressions midway through the experiment. All the participants were native-English-speaking college students who were expected to have the strong preference for absolute expressions found in Experiment 1, so they sometimes changed from a favored to a disfavored expression, and other times from a disfavored to a favored. Experiment 3 also examined whether the eye-voice relationships found for the onsets of time expressions were evident at the onsets of the second terms in time expressions. Such measurements help to reveal whether variations in the relative timing of speech persist beyond the first word into the continuation of utterances.

If variations in the familiarity or frequency of expressions systematically disrupt the patterns of apprehension and formulation observed in Experiment 2, different relationships between seeing and saying should emerge. In the signature of apprehension, the tendency for first eye movements to go toward expression-initial information may be different for favored compared to disfavored expressions (respectively, absolute and relative expressions in American English). Likewise, during formulation, the coordination between eye movements and expression components may differ for less-practiced expressions.

Method

Participants. The participants were 16 native American English speakers. The majority were enrolled in the departmental subject pool at the University of Illinois at Urbana-Champaign, and their participation fulfilled part of an introductory psychology course requirement. The remainder were undergraduates who were paid for their service in the experiment.

Materials, equipment, and procedure. The eyetracker was a headband mounted Eyelink (SR Research, Ltd.) controlled by a custom Pentium computer with an ASUS V3400TNT video card with 16 Mb video RAM. The experiment was run by a second custom-built Pentium computer. Participants viewed the displays on a 51 cm ViewSonic P815 monitor. They were seated at a distance from which the analog and digital clocks subtended visual angles of approximately 23° and 17°, respectively. Their utterances were recorded directly to digital soundfiles via a Creative Labs AWE64 16-bit sound card over a preamplified Shure SM10A microphone mounted on the eyetracker helmet. Button presses were made on an Eyelink hand-held button box.

The fixation point occupied the center of the analog displays. In the digital displays the position of the fixation point corresponded to the upper point in a colon that separated the hour from the minute windows. The fixation point remained on the screen during drift correction (between 500 and 1000 ms), which occurred on every trial. It was immediately replaced by the clock display for either 100 or 3000 ms. At the offset of the clock display a blank field appeared for either 3900 ms (after the 100 ms display) or 1000 ms (after the 3000 ms display), followed by a prompt to continue. In other respects the details of the procedure were the same as in Experiment 2.

For half of the participants, the conditions of the experiment and the arrangement of the materials were

identical to those for the English speakers in Experiment 2 (expression same-display changed, which we refer to as Display Change for short). The other half of the participants were in the expression changed-display same condition (Expression Change), and viewed clocks in the same format throughout the experiment, either all analog or all digital. The same random arrangements of the times were used as in Experiment 2. Within display types, participants named the times with both relative and absolute expressions. Type of expression was blocked, with the order of blocks counterbalanced over participants. Within each counterbalancing of the expression and display variables, half of the participants saw each clock for 100 ms and half for 3000 ms.

Design. Display duration (100 or 3000 ms) was a between-participants variable, as were the Display Change and Expression Change conditions. Eight participants received each of the four combinations of the duration and display/expression change factors. At each duration in the Expression Change condition, four participants saw only analog and four saw only digital displays, and all of them received 144 relative-expression and 144 absolute-expression trials. At each duration in the Display Change condition, all of the participants saw 144 analog clocks and 144 digital clocks, with four of the participants producing only relative and four producing only absolute expressions. All 16 participants named all 144 five-minute times twice.

Eye movement and speech analyses. The eye movement analyses were carried out as in Experiment 2. The speech analyses were performed using automatic speech recognition routines to identify the onsets of the minute and hour numbers in the time expressions produced. The Sphinx-II speech recognition system was used for this purpose (Huang et al., 1992; Ravishankar, 1996). The system recognizes speech using hidden Markov models to find the best path through a finite state network of triphone (Wickelphone) sequences by applying constraints from acoustic, lexical, and n-gram language model sources. The utterances from the experiment were first automatically transcribed using the recognition system and then hand-corrected by comparing the soundfiles to the machine transcriptions. Onset times for individual words were obtained using Viterbi forced alignment (see Jurafsky & Martin, 2000) of the hand-corrected transcripts to the sound files.

Results

Analyses of variance were carried out using participant-pairs as the random factor. At each duration (100 and 3000 ms), eight participant-pairs were created by combining the data from the two participants in each cell of the Expression Change and Display Change conditions to create the full factorial crossing of display type (analog or digital) with expression type (relative or absolute). This yielded four participant-pairs within the Expression Change and Display Change conditions at each duration. Unless otherwise noted, the designs for the analyses treated change condition as a between-participant-pair factor, crossed with display (analog or digital) and expression type (relative or absolute). Display duration (100 or 3000 ms) was a between-participant-pair factor. In other respects the analyses mirrored those in Experiment 2.

To streamline the presentation of results and aid comparisons with Experiment 2, Table 10 lists the effects and associated F values in Experiment 3 that replicated findings from Experiment 2. There were no disconfirmations of the Experiment 2 results, but there were findings that go beyond or substantially qualify them. We detail those in what follows.

Speed and accuracy of time naming. Mean latencies to begin producing the first and second time-terms in all conditions are shown in Table 11, along with the percentages of observations in each cell that were included in all analyses. For first terms, the results replicated Experiment 2 in showing faster responding overall with absolute compared to relative expressions, and with digital compared to analog clocks (see Table 10). There was also a significant interaction between display and expression, and except in one combination of conditions, the form of this interaction reflected the compatibility effects observed in Experiment 2: Responses to analog clocks were faster with relative expressions and responses to digital clocks were faster with absolute expressions. The exception arose for relative expressions produced with analog clocks in the 100 ms Display Change condition. Because of protracted latencies in this cell, there were significant interactions between expression and duration ($F(1, 4) = 13.0$) and among expression, duration, and change condition ($F(1, 4) = 14.4$). Change condition was associated with no other effects for onset latencies, with means of 1265 ms when expressions changed compared to 1199 ms when displays changed ($F(1, 4) = .5$).

For second terms, there were significant differences between the latencies for analog and digital displays (2314 to 1195 ms, respectively; $F(1, 4) = 66.9$) and for relative and absolute expressions (1995 to 1514 ms, respectively; $F(1, 4) = 31.0$). A larger difference in latencies between the two expression types produced to digital clocks (a 626 ms disadvantage for relative expressions) than to analog clocks (a 335 ms disadvantage for relative expressions) created a marginally significant interaction between display and expression ($F(1, 4) = 4.6$). Nothing else approached significance.

Locations and timing of fixations. Shown in Table 12 are the mean numbers of eye fixations per participant-pair to time-relevant regions in the analog and digital clock displays for both expression types, broken down by change condition. Analyses of variance were carried

Table 10
Replications in Experiment 3 of effects from Experiment 2

Measures and sources of effects	F value or 95% confidence interval in Experiment 3
Speech onset latency	
Digital displays faster than analog	$F(1,4) = 64.4$
Absolute expressions faster than relative	$F(1,4) = 13.6$
Interaction between display and expression (compatibility)	$F(1,4) = 11.9$
Distributions of eye fixations	
More fixations to 3000 ms display than 100 ms display	
Analog displays	$F(1,8) = 69.7$
Digital displays	$F(1,8) = 10.9$
More fixations to hour than minute region	
Analog displays	$F(1,8) = 25.7$
Interaction of duration and region for analog displays (no increase in looks to 12 and 6 at longer duration)	$F(3,24) = 19.8$
More fixations to top (.39) than middle (.31) or bottom (.31) of analog clock	$F(2,16) = 4.5$
First fixation latencies	
Earlier fixations to term-1 regions than term-2 regions	$F(1,3) = 11.0$
Proportion of fixations in 100 ms intervals from display onset	
On analog clocks, increased fixations to term-1 regions between 100 and 300 ms after display onset	95% confidence
On digital clocks, increased fixations to term-1 regions between 100 and 200 ms after display onset	interval = .24
First gaze durations	
Longer for analog than digital displays	$F(1,3) = 16.8$
Interaction between display and expression (longer gazes when incompatible than when compatible)	$F(1,3) = 20.3$
Interaction among display, expression, and term-1/term-2 region (compatibility effects only in term-1 region)	$F(1,3) = 9.6^{\dagger}$
Eye-voice spans from onset of fixation to onset of time expressions for 3000 ms displays	
Modal fixations in term-1 relevant regions preceded expression onset by 600 ms for analog displays and by 250 ms for digital displays	
Shorter for digital (366 ms) than analog clocks (1306 ms)	$F(1,3) = 44.9$
Eye-voice spans from offset of fixation to onset of time expression for 3000 ms displays	
Shorter for digital (114 ms) than analog clocks (289 ms)	$F(1,2) = 16.7$
Excluded observations	
More analog than digital	$F(1,4) = 16.7$

† Marginally significant.

out separately for the analog and digital clocks on the proportions of time-relevant fixations within regions, by change condition. Across the entire interval from fixation-point offset to the end of eye-movement recording, 45% of all fixations on analog clocks were within hour and minute regions and on digital clocks, 40%. The distributions of the fixations paralleled what we observed in Experiment 2, with one addition. There was a significant difference in the proportions of fixations that fell on the hour compared to the minute regions of digital clocks (.27 to .13, $F(1,8) = 25.7$). This was related to an effect of change type that occurred with digital clocks: The tendency to fixate minute regions was unaffected by change type (.13 for both conditions) but there were more fixations to the hour in the Expression

Change than in the Display Change condition (.33 to .21; $F(1,8) = 5.5$).

Table 13 gives the latencies of the first fixations to hour and minute regions for the trials on which both occurred. Fixations occurred sooner in the term-1 than in the term-2 region (510 to 807 ms). There were no significant effects of change condition in an analysis of variance (all $Fs < 1$), so a second analysis was carried out with only region (term-1 or term-2 relevant), display, and expression as factors. Region was significant (see Table 10), with marginally significant interactions between region and display ($F(1,3) = 6.3$) and among region, display, and expression ($F(1,3) = 5.7$). These interactions stemmed from a reversal in the fixations to the term-1 and term-2 regions that occurred for analog

Table 11
Latencies to produce relative and absolute expressions from analog and digital clocks in American English (Experiment 3)

Expression and clock format	Display and expression combination				Means
	Expression change		Display change		
	100 ms display	3000 ms display	100 ms display	3000 ms display	
Relative expressions					
Analog clock	(63)	(84)	(58)	(64)	
term 1 (minute)	1739	1687	2023	1402	1713
term 2 (hour)	2513	2348	2838	2228	2482
Digital clock	(92)	(87)	(85)	(78)	
term 1 (minute)	804	938	952	875	892
term 2 (hour)	1405	1537	1608	1482	1508
Absolute expressions					
Analog clock	(71)	(85)	(56)	(61)	
term 1 (hour)	1876	1862	1404	1719	1715
term 2 (minute)	2220	2148	1996	2224	2147
Digital clock	(89)	(91)	(92)	(96)	
term 1 (hour)	574	640	546	667	607
term 2 (minute)	880	923	790	936	882

Note. Shown in parentheses are the percentages of trials in each condition (out of 242 possible) included in the analyses.

displays when relative expressions were used and expressions changed. Specifically, there were unusually long latencies to fixate the minute hand of analog clocks, eliminating the difference in fixation latencies for the term-1 and term-2 regions and making the difference between the term-1 and term-2 regions for analog expressions smaller than for digital expressions. Otherwise, there were large and consistent differences in the fixation latencies to the term-1 and term-2 regions.

Fig. 13 traces the proportions of fixations occurring in the term-1 and term-2-relevant regions at each 100 ms interval after clock onset. With analog displays, fixations diverged around 400 ms after onset, compared to around 200 ms for digital displays. These are the same divergence times observed in Experiment 2. For both of the term-relevant regions, analyses of variance were performed on the proportions of fixations at each of the first six 100 ms intervals from display onset, with interval, display type, expression, and change condition as factors. Because none of the effects involving change condition approached significance (all $Fs < 1$), the data from both conditions were combined in further analyses.

For both the term-1 and the term-2 relevant regions, there were significant effects of interval and significant interactions of interval with display type. Pairwise planned comparisons of the cell means were performed as in Experiment 2, and for the term-1 regions, yielded similar results. In the term-2 regions, however, there were differences. First, for digital displays there were no significant changes in the proportions of fixations earlier than 400 ms after display onset. For analog displays, but only those accompanied by absolute expressions, fixations to the term-2 region increased

through 300 ms. This increase was followed by a significant decline at 400 ms that continued through the next 200 ms. For analog displays accompanied by relative expressions there was a significant interaction among interval, display type, and expression, pictured in Fig. 14. As suggested by the fixation latencies, there was a tendency in this condition to fixate the hour region initially before shifting to the minute region. Inspection of individual fixation patterns disclosed that a single participant, who first produced absolute and then relative expressions, strongly and consistently fixated hours and then minutes, and did so regardless of expression used.

Density and duration of gazes. Fixations in the 3000 ms display condition were aggregated into gazes and analyzed as in Experiment 2. The mean number of gazes before expression onset was .92 on minute regions and .95 on hour regions. There were marginally more gazes on analog than digital clocks (1.19 to .66; $F(1, 3) = 14.0$), and the distribution of gazes over the hour and minute regions differed for relative and absolute expressions, yielding a significant interaction between region and expression ($F(1, 3) = 20.9$). This was due to increases in the term-2 regions: For relative expressions there were more gazes in the hour than in the minute region (1.11 to .84) and vice-versa for absolute expressions (.79 to .99). No other effects approached significance.

The mean durations of first gazes are tabulated in Table 14 for all trials in which the term-1 and term-2 regions were both fixated. In a preliminary analysis there were no significant effects of change condition (all $Fs < 2.5$), so analyses were carried out on the pooled data from the two change conditions. First gazes to

Table 12

Mean numbers of fixations to regions on analog and digital clocks in American English (Experiment 3)

Display format and time-relevant region	3000 ms		100 ms	
	Expression change	Display change	Expression change	Display change
Analog				
Relative expressions				
Hour hand	348	351	54	93
Minute hand	243	270	32	64
12 o'clock	48	30	8	36
6 o'clock	10	12	6	6
Center	138	128	103	175
Other	132	208	79	200
Absolute expressions				
Hour hand	403	254	52	40
Minute hand	260	233	48	70
12 o'clock	105	31	10	22
6 o'clock	14	23	10	14
Center	296	118	124	144
Other	212	146	88	122
Digital				
Relative expressions				
Hour 10 s window	20	2	33	3
Hour 01 s window	370	274	369	112
Minute 10 s window	262	179	70	22
Minute 01 s window	5	2	4	2
Center	368	424	488	400
Other	34	11	152	7
Absolute expressions				
Hour 10 s window	28	20	10	2
Hour 01 s window	396	234	211	42
Minute 10 s window	182	202	32	45
Minute 01 s window	2	2	2	1
Center	501	444	495	522
Other	32	12	38	16

Table 13

Mean onset latencies (in ms) for first fixations to minute and hour regions for relative and absolute expressions with 3000 ms displays (Experiment 3)

Expression type	Condition and clock format			
	Expression change		Display change	
	Digital	Analog	Digital	Analog
Relative expressions				
Minute (term 1)	324	760	424	472
Hour (term 2)	808	658	1030	620
Absolute expressions				
Minute (term 2)	659	944	810	888
Hour (term 1)	581	478	544	410

analog displays were longer than to digital (586 to 316 ms), but as in Experiment 2, these effects changed significantly depending on region and expression (see Table 10). The pattern is shown in Fig. 15. In term-1 regions, gazes were shorter when display and expression were compatible (that is, when relative expressions were produced with analog displays and absolute expressions were produced with digital displays). In term-2 regions,
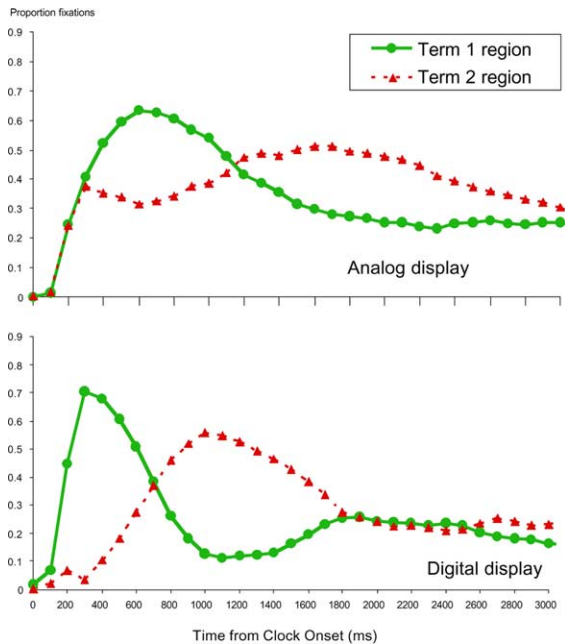
Fig. 13. Proportions of fixations to analog and digital clock regions carrying information needed for the first (term 1) and second (term 2) terms of time-telling expressions, in ms from display onset (Experiment 3).

Fig. 14. Proportions of fixations to analog clocks accompanied by the production of relative (top) and absolute (bottom) time expressions.

only display and expression complexity mattered. The same effects occurred for total gaze durations.

Speech onsets relative to eye locations. Fig. 16 displays the proportions of all fixations in the time-relevant regions relative to expression onset for the 3000 ms displays. Peak fixations in the term-1 relevant regions occurred at the same times as in Experiment 2 (see Table 10), but peak fixations in the term-2 relevant regions were 100–200 ms later, at 100 ms after expression onset for analog clocks and 300 ms after onset for digital clocks.

Table 15 contains the mean eye-voice spans from the start of fixations in the term-1 regions to first-term onset and from the start of fixations in the term-2 regions to second-term onset. Consistent with the results for gaze durations and speech onsets, eye-voice spans for first terms showed effects of the compatibility between expressions and displays, whereas eye-voice spans for second terms showed only the simple effects of expression and display complexity. In an analysis of variance, this pattern was associated with a marginally significant interaction among term, expression, and display ($F(1, 3) = 8.8$).

The remaining eye-voice span results are most easily summarized by considering how the term-1 and term-2 spans differed. The average time elapsed was 734 ms for the first term and 939 ms for the second term, but these differences varied substantially across conditions. In three of the four combinations of displays with expres-
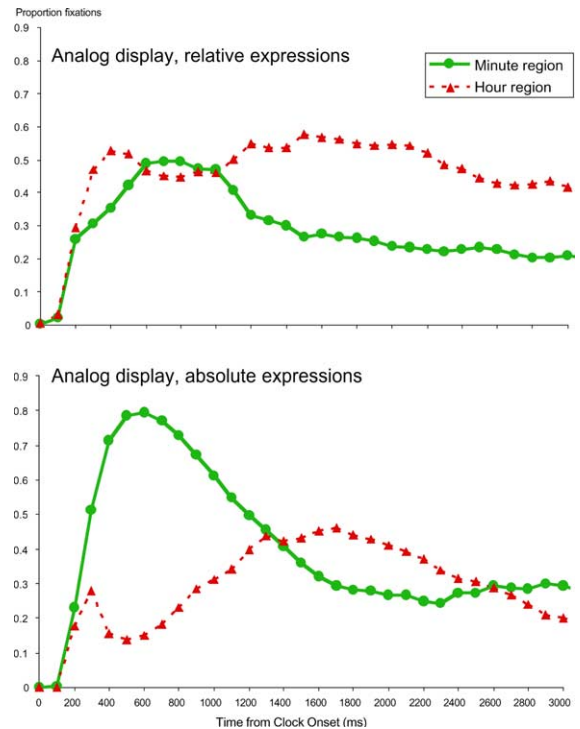
sions the absolute value of the difference between the term-1 eye-voice span and the term-2 eye-voice span averaged 87 ms. Only for relative expressions produced to analog clocks was there a large difference (716 ms, with a 937 ms eye-voice span for term 1 compared to 1653 ms for term 2). This was due to one speaker's consistent tendency to look first at the term 2 region when producing relative expressions to analog displays, and it contributed to significant interactions between term and expression ($F(1, 3) = 10.3$), term and display ($F(1, 3) = 9.9$), and expression and display ($F(1, 3) = 13.0$). The effects of term alone ($F(1, 3) = 5.0$) and expression alone ($F(1, 3) = 2.7$) were not significant. Change condition had no consistent impact for either term, yielding no Fs greater than 1 in a preliminary analysis of variance.

The eye-voice spans between the offsets of the final fixations on regions and the onsets of corresponding terms are also shown in Table 15. These final eye-voice spans were significantly longer for first than for second terms (600 to 217 ms; $F(1, 2) = 312.6$). Change condition was involved in a significant interaction with term ($F(1, 2) = 16.7$). The interaction was due to an increase in the first-term eye-voice spans from 517 ms in the Expression Change condition to 684 ms when displays changed but expressions stayed the same, while the

Table 14
Durations of first gazes (in ms) to minute and hour regions for relative and absolute expressions (Experiment 3)

Expression type	Change condition and clock format			
	Expression change		Display change	
	Digital	Analog	Digital	Analog
Relative expressions				
Minute (term 1)	412	487	459	482
Hour (term 2)	147	552	143	287
Absolute expressions				
Minute (term 2)	155	343	120	307
Hour (term 1)	204	782	244	658

second-term eye-voice spans in the same conditions decreased (from 244 to 189 ms).

Use of reference points. Fixation densities in the top, middle, and bottom of the analog display paralleled Experiment 2 (see Table 10). There was a marginal trend toward a larger regional disparity at longer durations, with .43 of the fixations to the top in the 3000 ms condition compared to .34 to the top in the 100 ms condition, and remaining fixations divided equally between the middle and bottom regions ($F(2, 16) = 3.0$). Pri-

marily responsible for this interaction was the 100 ms Expression Change condition, which created another marginally significant interaction among region, duration, and change condition ($F(2, 16) = 3.2$).

Discussion

Experiment 3 showed how one of the major results from Experiment 2 played out during the continuation of an utterance. In both Experiment 2 and Experiment 3, the eye-voice spans between first fixations in the term-1 region and expression onset reflected the compatibility of displays with expressions. For the second terms of
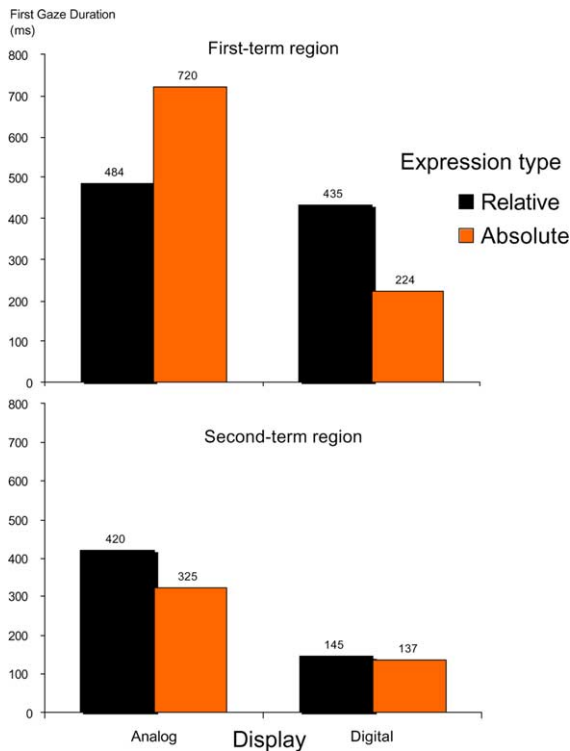

Fig. 15. Durations of first gazes to analog and digital clock regions relevant to the production of the first and second terms (illustrated with examples) in relative and absolute time expressions (Experiment 3).
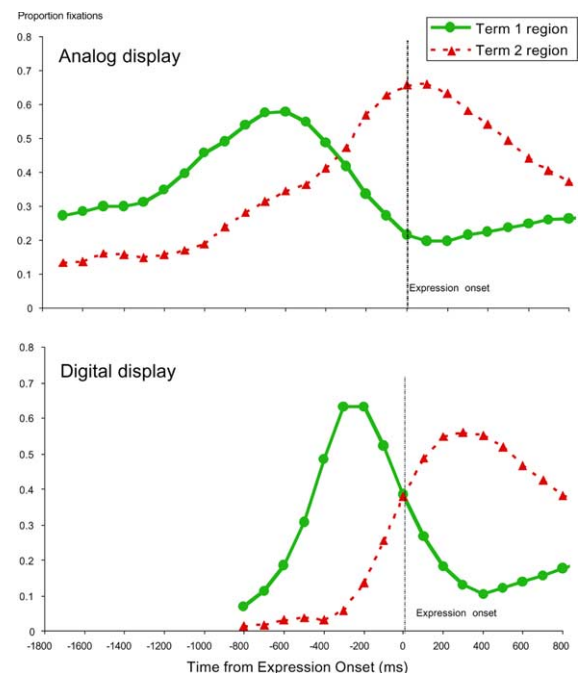

Fig. 16. Proportions of fixations to first-term-relevant and second-term-relevant clock regions before and after the onset of time expression (Experiment 3).

Table 15

Mean eye-voice spans (in ms) from fixations on minute and hour regions to onsets of minute and hour terms (Experiment 3)

Fixations and expression types	Change condition and clock format			
	Expression change		Display change	
	Digital	Analog	Digital	Analog
To first term from onset of first fixation in term-1 region				
Relative minute	622	947	466	928
Absolute hour	66	1404	129	1309
To second term from onset of first fixation in term-2 region				
Absolute minute	270	1227	133	1330
Relative hour	736	1706	508	1600
To first term from offset of final fixation in term-1 region				
Relative minute	564	706	635	796
Absolute hour	350	448	361	943
To second term from offset of final fixation in term-2 region				
Absolute minute	92	356	42	254
Relative hour	148	382	296	165

time expressions, using measures of the timing of their onsets, there was no compatibility effect. Rather, there were only simple effects of display and expression complexity. This mirrors the results from the eye movement measures, where first-fixation onsets and first-gaze durations both suggested differences between the term-1 and term-2 relevant display regions. The new result for term-2 eye-voice spans implies that, while the coordination between the eye and the voice remains throughout the time expression, there is an important change in the factors that jointly influenced the timing of speech and accompanying eye movements.

In most other respects, the results of Experiment 3 confirmed the major findings from Experiment 2. In both experiments, there were large effects of the compatibility between clock displays and time expressions, with relative expressions generally produced faster to analog clocks and absolute expressions produced faster to digital clocks than the other way around. Initial gazes disclosed a similar pattern, with longer gazes to the term-1 regions for incompatible combinations of displays and expressions than compatible combinations. First fixations on regions relevant to the first term in an upcoming expression occurred on average 510 ms after display onset, and tended to target the clock areas relevant to the first term in an upcoming time expression. Average eye-voice spans, from initial fixations on expression-relevant regions to the onset of corresponding time terms, were about 836 ms, 100 ms less or more for first and second terms, respectively. In broad summary, early viewing seems to support a coarse coding of the contents of a display, while subsequent fixations are successively targeted toward specific time-relevant regions during the generation of the elements of time expressions.

The major deviation from the results of Experiment 2 points toward an impact of speakers' expression preferences on production. The aversion of American undergraduates to relative expressions disclosed in Experiment 1 was associated with several problems in their time-telling performance in Experiment 3, especially when relative expressions were coupled with the perceptually more challenging analog clocks. First, with 100 ms display durations, there were many errors and the latencies in producing relative expressions were long, notably with changes in display. This is more likely to be due to differential experience (i.e., natural practice) than to the change itself, because there were no effects of the order in which the alternative displays were presented. Second, for one speaker producing relative expressions at the 3000 ms display duration, the first fixations on analog clocks tended to occur in the hour (term 2) regions rather than the minute (term 1) regions, reversing the usual linkage between regions and terms. This performance resembles what was seen for one English speaker who produced relative expressions in Experiment 2, but it was much more pronounced. This suggests systematic individual differences in the performance of time-telling, and perhaps in the preparation of utterances.

These problems cannot be diagnosed as mere lack of perceptual fluency with analog displays, because performance on these displays when absolute expressions were produced was comparable to that of Dutch speakers in Experiment 2. Rather, the problem seems to lie in the processes that interface perception with language. We will return to this important point in General Discussion.

Apart from the difficulties associated with relative expressions and their possible exacerbation by analog displays, speakers who were asked to change the time expressions that they were using midway through an

experimental session displayed few notable or long-lasting disruptions in performance. More precisely, any disruptions that may have occurred were sufficiently minor and short-lived that they did not substantially perturb performance. The consequence was that their performance differed little from that of speakers who used the same expression throughout.

General discussion

The questions behind these experiments centered on what happens when visual experience in a familiar domain is transformed into language. The perceptual experience came from clock faces displaying times in analog and digital formats. The language came from the standard expressions for telling time in English and Dutch. The results suggested that the development of a clock-time notion, during a process that we term apprehension, can occur within the first two to three hundred milliseconds after the exposure of a clock display. The content of this notion is often sufficient to direct the first eye movement to the clock region containing the detailed information needed for the first term in an upcoming time expression.

This performance would be unremarkable were it not that upcoming time expressions sometimes put hours first and sometimes minutes, that synonymous clock displays vary between angular geometric figures and number arrays, and that either kind of expression can be used with either kind of display. The implication is that the form of an expression that begins to be produced roughly a second after the onset of a clock display can control the first saccade to hour or minute information. In turn, this means that a perceptual display can be parsed and its components distinguished with respect to their relevance for an utterance plan within a few hundred milliseconds after display onset.

Digital displays offer fewer challenges to this performance than analog, because the hour and minute information is explicit and located in dedicated regions. On almost every measure, there was less variability associated with time-telling from digital than from analog displays, where the relevant information is implicit and distributed over a minimum of 12 regions. Even so, despite this greater perceptual complexity, the hour and minute hands of analog clocks were in most instances rapidly, selectively, and successively fixated in accordance with the needs of upcoming expressions. By the second fixation, eye movements were normally directed toward the hand carrying the information needed to begin the utterance.

This fast, reliable targeting of early eye movements is the basis of the argument for apprehension. The fixation targets consistently met linguistic requirements that varied with upcoming expression and were consistently

robust against variable perceptual properties. Although the present data do not reveal how efficient the apprehension process is from the outset of an experimental session or how it varies with experience and the familiarity of scenes, it is noteworthy that Griffin and Bock (2000) also found fast, discriminative targeting of the referents of sentence subjects in pictures of events that speakers saw only once, had never seen before, and described freely. This occurred with controls for the effects of distinctive perceptual or conceptual properties of the subject and object referents.

In exhibiting these features of apprehension, time-telling is perhaps unsurprisingly like more compositional varieties of normal language production. The apprehension of familiar visual information *should* be fast. What we had not anticipated was the selectivity and seriality of subsequent eye movements, with their tight linkage to upcoming words. Under most circumstances in the present experiments, the eyes left the region relevant to the first time-term and moved to the region relevant to the second time-term, arriving on average about a half second prior to the initiation of production and roughly 940 ms prior to the onset of the second time-term. This occurred even though fixation of the information in the individual time regions appeared to be unnecessary to produce the corresponding expressions: Speakers who saw the same clocks for only 100 ms, too little time to make an eye movement, were able to tell time without significantly reduced accuracy. On the assumption that the eyes moved to regions relevant to the preparation of an upcoming utterance when they were able to do so, this suggests that the readying of a formulaic, idiomatic expression is, at the level of grammatical encoding, at least as incremental as the preparation of a fully compositional utterance.

Apart from its implications for our understanding of how people talk, the tight, apparently natural coupling between seeing and speaking sheds light on the nature of a mechanism that operates during child language acquisition. There is compelling evidence that by around 18 months of age, children are sensitive to what an adult is looking at when the adult produces a word, and they use this information in word learning (Moore, Angelopoulos, & Bennett, 1999). Our findings and those of others, findings about both language production and language comprehension (Meyer et al., 1998; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995), illuminate why these gaze cues are available for babies to use and to learn to use (Hollich, Hirsh-Pasek, & Golinkoff, 2000), in the first place. Caregivers are not the only ones who look at what they are talking or hearing about, and doing it demands no particular social sensitivity. Apparently everyone does it, even when they are only talking to a computer.

More directly, the present findings go beyond the specifics of time telling to disclose something about the

general properties of normal language production. The findings bear on a range of classic issues about how language is produced and how language affects cognition, and we consider some of these in subsequent sections. We proceed from questions about how starting points are selected through questions about what incrementality means and whether and how the way we tell mundane time influences the way we view it, in the simplest sense.

From apprehension to a starting point

The strong correlation between initial fixation targets and initial words brings up a longstanding psycholinguistic obsession with utterance starting points for production and comprehension, or what is mentioned first. Starting points have been argued to have a special status (MacWhinney, 1977), being in various ways distinct from what comes later. This special status is most often associated with sentence subjects, but has sometimes been suggested to accrue to initial noun phrases regardless of their structural role (Gernsbacher & Hargreaves, 1988). One traditional version of these arguments in language production is that things become starting points because of the extra attention they demand or deserve by being more important, more salient perceptually or more central pragmatically (Osgood, 1971, 1980; Osgood & Bock, 1977). This predicts that speakers tend to begin their utterances with whatever naturally draws or captures their attention.

The present results and those of Griffin and Bock (2000) do not disconfirm this hypothesis, but they seriously challenge its adequacy. The subjects of sentences (the first noun phrases) produced by speakers in Griffin and Bock's study did not correspond to the elements that drew the most or the earliest fixations in a free viewing situation, nor were they straightforwardly predictable from the conceptual features of individual entities. Because event comprehension occurred within the same time frame as the selection of a starting point for production, Griffin and Bock proposed that the subject was identified on the basis of causal and aspectual coding carried out during the apprehension process. This implies that when speakers produce fluent utterances to describe events, the eye is sent not to the most salient element in a scene, but to an element already established as a suitable starting point. The eye may nominate starting points, but it does not select them. The mind does.

Put differently, the message-creating mechanisms of sentence production weigh the information provided by perception or thought within a cognitive and communicative context. Starting points are normally and naturally selected with this context in mind. In line with this, the effects of an entity's perceptual and conceptual salience on starting-point selection are fairly weak in

comparison to the effects of, say, focus and topicality in discourse (see Levelt, 1989, for review). So, the designation of a starting point can be seen as growing out of a process that identifies suitable places to start in terms of (a) the goal in force, (b) the nature of the action required to achieve it, and (c) the fit of apprehended information into this context. When an action plan involves speaking, apprehension provides a schematic coding of relations (causal relations, spatial relations, aspectual or temporal relations) within a scene or notion to frame the utterance. The conceptual fulcrum on which these relations turn defines the starting point. Our proposed alternative to a salience account, then, is this: Initial eye movements are controlled by a rudimentary utterance plan that projects relationships from a designated starting point.

Utterance initiation and incrementality

Consistent with early fixations generally occurring in regions relevant to the first terms of time expression, later fixations generally occurred in regions relevant to the second terms. The foreshadowing of individual utterance components by the locations of eye fixations in most (but not all) of the speakers argues for a sort of incrementality in the underlying linguistic formulation of an utterance, and not only in its overt articulation.

There were two notable features of the temporal coupling. One was the anticipatory character of the eye fixations. Averaging across both languages and all conditions, the onsets of fixations preceded the onsets of corresponding terms by 770 ms, implying that the eyes were pushed to a region in advance of lexical retrieval and not pulled there, for example, on the impetus of an already-selected word. Second was the apparent absence of necessity in the linkage between the eyes and utterance components. Since 100 ms exposures of clock faces provided enough information to formulate a reasonably accurate time expression, and longer exposures allowed the eye, with a single saccade, to move into an area from which any remaining information was likely to be available, the eye movements are plausibly seen as instruments of an ongoing incremental process.

We can distinguish two general ways in which incrementality in formulation might be realized. In *linear* incrementality, the preparation for one increment of an utterance would be identical to the preparation of another increment, with modulation by factors such as the complexity and frequency of the incrementation. In *hierarchical* incrementality, preparation for one increment might subsume partial preparation for another, or preparation for an initial increment might require establishing a framework that supports the production of the first increment as well as later ones.

The speech onset latencies, gaze durations, and eye-voice spans from Experiments 2 and 3 were more suggestive of hierarchical than linear incrementality.

Measures of preparation in term-1 relevant regions tended to be longer than in term-2-relevant regions and, more provocatively, were affected by different things. In term-1 regions, timing variations were associated with the compatibility of displays and expressions (analog with relative and digital with absolute), whereas term-2 regions elicited simple effects of display and expression type only.

Why are compatibility effects restricted to the term-1 regions? An attractive idea is an utterance ground-breaking effect, inversely analogous to the sentence wrap-up effects familiar from research on language comprehension (Aaronson & Scarborough, 1976; Just & Carpenter, 1980; Mitchell & Green, 1978). Just as wrap-up can be explained in terms of synthesizing and integrating the contents of a sentence with its context, groundbreaking may be explainable as an analysis or partial *dis*-integration from context performed to support the elaboration of structure and the retrieval of upcoming words. The individuation of information represented in a notion or perceptual image, along with provisional binding to indices for upcoming linguistic constituents, will proceed more or less fluently depending on natural compatibility as well as on the experience of the speaker. For example, the preparation of an absolute expression might require individuating the hour and the minute within the perceptual image and binding the individuals into an abstract structural frame, preparatory to retrieving the hour and minute names. Since a digital display conveys hours and minutes in a format that is isomorphic with the structural frame, it facilitates the binding to appropriate positions. Similarly but less transparently, analog clocks provide more perceptual support for binding information to the frames of relative expressions than digital clocks do. When a scene or notion lacks such support, it is harder to anticipate and implement the association of individuals to frame locations.

It could be objected that time expressions have no need for the kinds of structural frames assumed in this account, but that would be wrong. Even simple expressions have a structure, with requirements about the kinds of elements that occupy different positions within it. The structures carry information independent of the elements themselves. For instance, a numerical date expression might seem to be little more than a list (12-5-75), but there is a structural frame behind it. A numerical date in American English has successive slots for the month, the day, and the year (making the date above May 12th, 1975); European dating, in contrast, puts the day first (making the date December 5th, 1975). In the absence of clues about which structure is in force, more than a third of all dates are ambiguous. Time expressions, too, are simple but undeniably structural, with conventionally defined places for hours and minutes within positionally specified frames.

Language, thought, and mundane time

The choreography of the eyes and the mouth during speaking is a powerful means for investigating the relationship between perception and language. One perennially influential view of this relationship, newly energized (and newly disputed; Li & Gleitman, 2002) by studies of cognition in different languages (e.g., Berman & Slobin, 1994; Boroditsky, 2001; Bowerman & Levinson, 2001; Levinson, Kita, Haun, & Rasch, 2002; Lucy, 1992a, 1992b; Roberson, Davies, & Davidoff, 2000; Spelke & Tsivkin, 2001), is the linguistic relativity hypothesis. The most famous formulation of linguistic relativity, popularized by Benjamin Whorf, drew on examples of time reckoning in English and Hopi. Whorf wrote that in Hopi, time is reckoned not in terms of a succession of tokens of the same type, but in terms of successive reappearances of the same token (Whorf, 1956, p. 148). For example, English talks of time periods in terms of separate but consecutive days; Hopi talks of them in terms of repeated reincarnations of the same day (reminiscent of the movie *Groundhog Day*). From this, Whorf inferred that the Hopi conceive of days in roughly the way that we conceive of houseguests: A houseguest makes successive appearances during a visit, but remains the same individual. Presumably, similar arguments could be made about other units of time, minutes and hours included.

Viewed objectively, it is commonsensical to regard differences in what people say or how they say it as reflections of potential differences in what or how they think. Much more controversial is Whorf's account of how different ways of thinking arise. Linguistic determinism proposes that manners of thinking, ways of construing the world, come about *because* of one's language. Coupled with the idea that differences in language give rise to different manners of thinking and perceiving, differences in how languages work become prime candidates for explaining supposed differences in cognition and perception.

Dutch and English time expression, in particular relative time expression, is a miniature system within which this hypothesis can be put to a test. Obviously, the linguistic reference points for relative time-telling differ in Dutch and English. Conveniently, the eyes reveal much about how time expressions are prepared; indeed, the way the eyes move in service of language planning might be regarded as an unexpected variety of linguistic determinism. If the differences between Dutch and English time expressions motivate different ways of parsing clock displays, in accordance with linguistic relativity, it should be readily apparent in distinct patterns of eye movements.

In fact, there were few differences to speak of, and none that could be allocated a role in supporting production of the contrasting relative time frameworks in English and Dutch. The enormous disparities in eye movements and linguistic formulation associated with

different displays and different expressions were consistent and consistently large for both languages, with specific language factors adding little to an account of the transition from seeing to saying. This is all the more remarkable in light of how opposed the expression preferences in the two languages are (Fig. 4) and how peculiar the average native English speaker finds the Dutch hourly framework to be. Intuitively, it strikes Americans as a very odd way of telling the time.

Of course, mundane time frameworks may play their part during perceptual or cognitive computations that are inaccessible to our measures. A prime candidate is apprehension, during the initial perceptual uptake of information. Surely, frameworks must also play a role in learning how to tell the time. However, we suspect on the basis of what failed to materialize in our data that frameworks do not play a part during the linguistic formulation process, in active support of lexical retrieval or grammatical encoding. Another way of putting it is that if the Dutch "see" or "think about" the clock differently than English speakers do, the differences do not affect the dance of looking-and-saying.

We did observe one large, albeit localized, effect of the linkage between the language of time-telling and the process of time-telling. When using relative expressions to tell time from analog clocks, some American English speakers exhibited a formulation process that was rather less fluid than a dance, despite the fluency of the eventual utterances. The next section considers the reasons behind the disruptions.

Seeing for saying

In Experiments 2 and 3, producing relative expressions to analog clocks seemed to have untoward consequences for some English speakers. Compared to Dutch performance, error rates were high and the onset of speech was delayed. Perhaps most strikingly, for two speakers (one in Experiment 2 and another in Experiment 3) the coordination between the order of fixating the term-relevant regions and the order of the terms themselves in speech did not follow the "look at minute, say minute, look at hour, say hour" routine.

This is a small piece of evidence that speakers can indeed differ in how they approach the task of interfacing their perceptions with their language. It indicates that the counterpoint between the order of fixations and the order of words is not a prerequisite of fluent speech, and that this kind of incrementality has exceptions (see also Ferreira & Swets, 2002; van der Meulen, Meyer, & Levelt, 2001). More speculatively, it suggests that language may have an impact on the uptake of visual information in the manner implied by weaker versions of linguistic determinism (e.g., Hunt & Agnoli, 1991). We assume that the deviant speakers shared the American predilection for absolute expressions. Maintaining this preference in a world with analog clocks may demand a lot of practice looking first at the little hand and then at the big hand, and such experience need not succumb immediately to the demands of a different manner of speaking.

Without making too much of single instances, it is worth considering whether these exceptions prove any linguistic determinism rule, either weak or strong. The existence of a consistent relationship between the order in which information is expressed and the order in which it is tapped makes it hard to deny that there is an effect of language on an aspect of perception. Equally, the pervasive relationship across speakers, languages, and types of expressions makes it hard to deny that how we speak has a powerful effect on how we move our eyes over the world, in a literal sense. But the very literalness of the implied connection, the arguably superficial role that language plays in enforcing it, and the absence of any discernible effect on eye movements of a genuine difference between two languages in their preferred frameworks for mundane time all militate against the supposition that the conceptual structures for mundane time differed for speakers who used different kinds of time expressions. It could still be so, but we have not seen any evidence of it. What we observed, instead, was that most (but not all) speakers tended to sample information from a perceptual image in a manner that was efficient for what they were preparing to say. What is efficient for one kind of expression is less so for another, and most speakers adapt accordingly. The clear implication is that the perceptual information *remains the same* in the face of different linguistic requirements; what changes are the processes that effect the transformation between perception and language.

This is an explicit variant of what Slobin (1996) dubbed *thinking for speaking*. We call it *seeing for saying*. Slobin argued that speakers of languages in which, for instance, singular and plural number is part of the grammar must deal with number-relevant information when preparing every utterance; speakers of languages that lack grammatical number specifications need not do so. The latter speakers may have the same perceptions and conceptions of numerosity, but lack experience with bringing the information to bear during the preparation of utterances. When a language, or an expression within a language, makes demands on a particular piece of information, the interface must bring that information on line and ideally, bring it on line when the language can make best use of it. If we take eye movements as tracing the structure of an interface between visual perception and language, we can conclude that speakers with diverse time-telling preferences differ not so much in what they see as in how they use what they see in preparing to talk.

Parsing the clock

Paradoxically, the weakness of specific linguistic differences in clock viewing occurred in company with

very strong language-directed guidance of eye movements. Little in what the eyes did stemmed from the properties of the visual displays, except in mediating the uptake of the information needed for speaking. Even digital clocks, whose left-to-right numbering format ought to coax the eyes into their strong left-to-right reading habit, were fixated first on the right and then on the left when expressions called for minute information first. One consequence is that we can say virtually nothing about the perception of the clock displays. We do know that 100 ms of viewing was sufficient to provide much of the visual information that was needed to perform the task. With a 30-fold increase in viewing time, the increase in the number of fluent, correct responses was just 6%. The targeting of initial fixations suggested that enough detail was available to identify the appropriate starting points for the linguistic formulation of upcoming expressions. The obvious inference from these facts is that most of the perceptual action took place prior to programming the first saccade.

What, then, are the eyes doing during the next three seconds or so? We have argued that after apprehension, they are guided by a linguistic plan, and not the other way around. We assume that the eye movements serve to ground the needs of this plan, perhaps supporting the retrieval of information from memory (Ballard, Hayhoe, Pook, & Rao, 1997), particularly lexical information (Meyer et al., 1998; Zelinsky & Murphy, 2000), and verifying the constituents of the plan itself.

Despite its general plausibility, there is something not entirely satisfying about this hypothesis in the present context. The difficulty of repeatedly retrieving twenty-one or even fewer number names in the course of a single experimental session seems unlikely to require elaborate, sustained support from visual input, however dutiful and otherwise unoccupied the visual system may be. An alternative hypothesis is that the eyes help to enforce the linearization and separation of information placed in serial order. In the terms of Dell, Burger, and Svec (1997), the creation of serial order requires suppressing the past, activating the present, and preparing the future. A current object of fixation supports the activation of associated information in memory, but may also serve the function of de-activating previous things (Irwin, 1998; Irwin & Brockmole, 2000; Irwin & Carlson-Radvansky, 1996). Vision may also help to prepare the future by bringing upcoming objects into parafoveal preview and readying saccades in their direction. Accordingly, by recruiting the eyes in their service, syntactic plans may be implemented more efficiently and, perhaps, more fluently.

Timely idioms

Expressions for mundane time are a kind of nonfigurative idiom, construction, or formula (Fillmore, Kay, & O'Connor, 1988; Kuiper, 1996). They constitute relatively invariant ways of speaking that, with few exceptions (see Van der Henst et al., 2002), are shaped more by convention and history than by the changing communicative demands of specific conversational settings. In this respect they function more like names (Malt, Sloman, Gennari, Shi, & Wang, 1999) than the phrases that they are.

Their linguistic formulation nonetheless proceeds compositionally, built up from separable elements, regardless of whether the expression is applied to a typical reference domain (e.g., absolute expressions to digital clocks) or to an atypical one (e.g., absolute expressions to analog clocks). In both cases, we found that separate fixations were allocated to minute and hour information prior to the onsets of the corresponding time terms. Moreover, the timing of these fixations for most speakers, most of the time, formed a counterpoint with the beginnings of the relevant words, punctuating their individuality within the whole.

These findings converge with others that show that even prototypical idioms are not produced as indivisible chunks, but as structured sequences of words (Cutting & Bock, 1997; Peterson, Burgess, Dell, & Eberhard, 2001). Sprenger, Levelt, and Kempen (2000) proposed that this compositional process is guided by a unitary lexical entry that they called a *superlemma*. An idiom's superlemma forms an overarching structure linking together the individual words, which must be encoded one by one. Such structures could also serve in the production of time expressions. In that role, superlemmas tell a different and simpler story about the compatibility effects observed in Experiment 3. If the presence of misfitting perceptual information delays the retrieval of a superlemma, the observed increases for first terms in expression onset, in gaze durations, and eye-voice spans follow directly.

Conclusion

The experiments traced some of the relationships among alternative means of displaying time, alternative means of telling time, and the consequences for visual perception and language production. The results point to the existence of strong and interpretable relationships between how clock displays are viewed and how time is told. Evidently, the eyes serve as a sensitive and powerful meter of the psycholinguistic processes behind the production of time expressions.

The significance of these findings lies less in what they reveal about how people tell time than in what they reveal about how people talk. In this context, time-telling represents a small, tractable approach to a large, intractable problem. The problem concerns the normal ability to convey thought in language and the normal influence of language on how we perceive and conceive the world.

Acknowledgments

References

Aaronson, D., & Scarborough, H. S. (1976). Performance theories for sentence coding: Some quantitative evidence. *Journal of Experimental Psychology: Human Perception and Performance, 2*, 56–70.

Ballard, D. H., Hayhoe, M. M., Pook, P. K., & Rao, R. P. N. (1997). Deictic codes for the embodiment of cognition. *Behavioral and Brain Sciences, 20*, 723–767.

Berman, R. A., & Slobin, D. I. (1994). *Relating events in narrative: A crosslinguistic developmental study*. Hillsdale, NJ: Erlbaum.

Biederman, I. (1981). On the semantics of a glance at a scene. In M. Kubovy, & J. R. Pomerantz (Eds.), *Perceptual organization* (pp. 213–224). Hillsdale, NJ: Erlbaum.

Bock, J. K. (1996). Language production: Methods and methodologies. *Psychonomic Bulletin & Review, 3*, 395–421.

Bock, J. K., Eberhard, K. M., Cutting, J. C., Meyer, A. S., & Schriefers, H. (2001). Some attractions of verb agreement. *Cognitive Psychology, 43*, 83–128.

Bock, J. K., & Griffin, Z. M. (2000). Producing words: How mind meets mouth. In L. R. Wheeldon (Ed.), *Aspects of language production* (pp. 7–47). London: Psychology Press.

Boroditsky, L. (2001). Does language shape thought? Mandarin and English speakers' conceptions of time. *Cognitive Psychology, 43*, 1–22.

Bowerman, M., & Levinson, S. C. (Eds.). (2001). *Language acquisition and conceptual development*. Cambridge, UK: Cambridge University Press.

Cahn, L. A., & Diesfeldt, H. E. (1977). Bij de tijd zijn: Een studie over desoriëntatie in tijd en het aflezen van klokken bij geestelijk gestoorde oudere patiënten [To know the score: A study of disorientation in time and clockreading by mentally disturbed older patients]. *Nederlands Tijdschrift voor Gerontologie, 8*, 213–220.

Cutting, J. C., & Bock, J. K. (1997). That's the way the cookie bounces: Syntactic and semantic components of experimentally elicited idiom blends. *Memory & Cognition, 25*, 57–71.

Dell, G. S., Burger, L. K., & Svec, W. R. (1997). Language production and serial order: A functional analysis and a model. *Psychological Review, 104*, 123–147.

Ferreira, F., & Swets, B. (2002). How incremental is language production? Evidence from the production of utterances requiring the computation of arithmetic sums. *Journal of Memory and Language, 46*, 57–84.

Fillmore, C. J., Kay, P., & O'Connor, M. C. (1988). Regularity and idiomaticity in grammatical constructions. *Language, 64*, 501–538.

Gell, A. (1992). *The anthropology of time: Cultural constructions of temporal maps*. Oxford, UK: Berg.

Gernsbacher, M. A., & Hargreaves, D. (1988). Accessing sentence participants: The advantage of first mention. *Journal of Memory and Language, 27*, 699–717.

Goolkasian, P., & Park, D. C. (1980). Processing of visually presented clock times. *Journal of Experimental Psychology: Human Perception and Performance, 6*, 707–717.

Gordon, R., 1999. The time-course of attention during scene perception. Unpublished doctoral dissertation, University of Illinois at Urbana-Champaign.

Griffin, Z. M., & Bock, J. K. (2000). What the eyes say about speaking. *Psychological Science, 11*, 274–279.

Grondin, S. (2001). From physical time to the first and second moments of psychological time. *Psychological Bulletin, 127*, 22–44.

Hawking, S. (1998). *A brief history of time: The updated and expanded tenth anniversary edition*. New York: Bantam.

Hayhoe, M. (2000). Vision using routines: A functional account of vision. *Visual Cognition, 7*, 43–64.

Hollich, G., Hirsh-Pasek, K., & Golinkoff, R. M. (2000). *Breaking the language barrier: An emergentist coalition model for the origins of word learning*. Chicago: University of Chicago Press.

Huang, Z, Alleva, F., Hon, H., Hwang, M., Rosenfeld, R., 1992. The SPHINX-II speech recognition system. Technical Report CMU-CS-92-112. Pittsburgh, PA: Carnegie-Mellon University, School of Computer Science.

Hunt, E., & Agnoli, F. (1991). The Whorfian hypothesis: A cognitive psychology perspective. *Psychological Review, 98*, 377–389.

Irwin, D. E. (1992). Memory for position and identity across eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*, 307–317.

Irwin, D. E. (1998). Lexical processing during saccadic eye movements. *Cognitive Psychology, 36*, 1–27.

Irwin, D. E., & Brockmole, J. (2000). Mental rotation is suppressed during saccadic eye movements. *Psychonomic Bulletin & Review, 7*, 654–661.

Irwin, D. E., & Carlson-Radvansky, L. A. (1996). Cognitive suppression during saccadic eye movements. *Psychological Science, 7*, 83–88.

Jurafsky, D., & Martin, J. H. (2000). *Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition*. Upper Saddle River, NJ: Prentice-Hall.

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review, 87*, 329–354.

Kartsounis, L. D., & Crewes, H. (1994). Horologagnosia: An impairment of the ability to tell the time. *Journal of Neurology, Neurosurgery and Psychiatry, 57*, 384–385.

Kuiper, K. (1996). *Smooth talkers: The linguistic performance of auctioneers and sportscasters*. Mahwah, NJ: Erlbaum.

Le Poidevin, R. (1993). *The philosophy of time*. Oxford, UK: Oxford University Press.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.

Levinson, S. C., Kita, S., Haun, D. B. M., & Rasch, B. H. (2002). Returning the tables: Language affects spatial reasoning. *Cognition, 84*, 155–188.

Li, P., & Gleitman, L. (2002). Turning the tables: Language and spatial reasoning. *Cognition, 83*, 265–294.

Loftus, G. R., & Mackworth, N. H. (1978). Cognitive determinants of fixation location during picture viewing. *Journal of Experimental Psychology: Human Perception and Performance, 4*, 565–572.

Logan, G. D. (1995). Linguistic and conceptual control of visual spatial attention. *Cognitive Psychology, 28*, 103–174.

Lucy, J. A. (1992a). *Grammatical categories and cognition: A case study of the linguistic relativity hypothesis*. Cambridge, UK: Cambridge University Press.

Lucy, J. A. (1992b). *Language diversity and thought: A reformulation of the linguistic relativity hypothesis*. Cambridge, UK: Cambridge University Press.

MacWhinney, B. (1977). Starting points. *Language, 53*, 152–168.

Malt, B. C., Sloman, S. A., Gennari, S., Shi, M., & Wang, Y. (1999). Knowing versus naming: Similarity and the linguistic categorization of artifacts. *Journal of Memory and Language, 40*, 230–262.

Meyer, A. S., Sleiderink, A., & Levelt, W. J. M. (1998). Viewing and naming objects: Eye movements during noun phrase production. *Cognition, 66*, B25–B33.

Meyer, A. S., & van der Meulen, F. F. (2000). Phonological priming effects on speech onset latencies and viewing times in object naming. *Psychonomic Bulletin & Review, 7*, 314–319.

Miller, R. J., & Penningroth, S. (1997). The effects of response format and other variables on comparisons of digital and dial displays. *Human Factors, 39*, 417–424.

Mitchell, D. C., & Green, D. W. (1978). The effects of context and content on immediate processing in reading. *Quarterly Journal of Experimental Psychology, 30*, 609–636.

Moore, C., Angelopoulos, M., & Bennett, P. (1999). Word learning in the context of referential and salience cues. *Developmental Psychology, 35*, 60–68.

Osgood, C. E. (1971). Where do sentences come from?. In D. D. Steinberg, & L. A. Jakobovits (Eds.), *Semantics: An interdisciplinary reader in philosophy, linguistics, and psychology* (pp. 497–529). Cambridge, UK: Cambridge University Press.

Osgood, C. E. (1980). *Lectures on language performance*. New York: Springer.

Osgood, C. E., & Bock, J. K. (1977). Salience and sentencing: Some production principles. In S. Rosenberg (Ed.), *Sentence production: Developments in research and theory*. Hillsdale, NJ: Erlbaum.

Paivio, A. (1978). Comparisons of mental clocks. *Journal of Experimental Psychology: Human Perception and Performance, 4*, 61–71.

Peterson, R. R., Burgess, C., Dell, G. S., & Eberhard, K. M. (2001). Dissociation between syntactic and semantic processing during idiom comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*, 1223–1237.

Potter, M. C. (1975). Meaning in visual search. *Science, 187*, 965–966.

Ravishankar, M.K., 1996. Efficient algorithms for speech recognition. Technical Report CMU-CS-96-143. Pittsburgh, PA: Carnegie-Mellon University, School of Computer Science.

Roberson, D., Davies, I., & Davidoff, J. (2000). Color categories are not universal: Replications and new evidence from a stone-age culture. *Journal of Experimental Psychology: General, 129*, 369–398.

Shanon, B. (1985). How to tell the time in Hebrew. *Journal of Psycholinguistic Research, 14*, 317–329.

Siegler, R. S., & McGilly, K. (1989). Strategy choices in children's time-telling. In I. Levin, & D. Zakay (Eds.), *Time and human cognition: A life-span perspective* (pp. 185–218). Amsterdam: Elsevier.

Slobin, D. I. (1996). From "thought and language" to "thinking for speaking". In J. Gumperz, & S. C. Levinson (Eds.), *Rethinking linguistic relativity* (pp. 70–96). Cambridge, UK: Cambridge University Press.

Snodgrass, J. G., & Yuditsky, T. (1996). Naming times for the Snodgrass and Vanderwart pictures. *Behavior Research Methods, Instruments, & Computers, 28*, 516–536.

Spelke, E. S., & Tsivkin, S. (2001). Language and number: A bilingual training study. *Cognition, 78*, 2045–2088.

Sprenger, S.A., Levelt, W.J.M., Kempen, G., 2000, March. The production of fixed expressions: Idiom representation and access. Poster presented at the CUNY Sentence Processing Conference, La Jolla, CA.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science, 268*, 1632–1634.

ter Meulen, A. G. B. (1995). *Representing time in natural language: The dynamic interpretation of tense and aspect*. Cambridge, MA: MIT Press.

Van der Henst, J.-B., Carles, L., & Sperber, D. (2002). Truthfulness and relevance in telling the time. *Mind and Language, 17*, 457–466.

van der Meulen, F., Meyer, A. S., & Levelt, W. J. M. (2001). Eye movements during the production of nouns and pronouns. *Memory & Cognition, 29*, 512–521.

Whitrow, G. J. (1988). *Time in history*. Oxford, UK: Oxford University Press.

Whorf, B. L. (1956). *Language, thought, and reality*. Cambridge, MA: MIT Press.

Zelinsky, G. J., & Murphy, G. L. (2000). Synchronizing visual and language processing: An effect of object name length on eye movements. *Psychological Science, 11*, 125–131.