AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Experimental Psychology: Learning, Memory, and Cognition

ISSN: 0278-7393

2023, Vol. 49, No. 11, 1715–1731
https://doi.org/10.1037/xlm0001269

Iconicity Bias and Duration

Laura Jane Kelly and Sangeet Khemlani
US Naval Research Laboratory, Navy Center for Applied Research in Artificial Intelligence, Washington, District of Columbia, United States

Descriptions of durational relations can be ambiguous, for example, the description "one meeting happened during another" could mean that one meeting started before the other ended, or it could mean that the meetings started and ended simultaneously. A recent theory posits that people mentally simulate descriptions of durational events by representing their starts and ends along a spatial axis, that is, an iconic representation of time. To draw conclusions from this iconic mental model, reasoners consciously scan it in the direction of earlier to later timepoints. The account predicts an iconicity bias: People should prefer descriptions that are congruent with an iconic scanning procedure—descriptions that mention the starts of events before the ends of events—over logically equivalent but incongruent descriptions. Six experiments corroborated the prediction; they show that iconicity biases in temporal reasoning manifest in cases when reasoners consciously evaluate the durations of events.

Keywords: events, temporal reasoning, durational relations, mental models, iconicity bias

People construct *iconic* mental simulations when they reason about space and time: Their simulations mimic the scenarios they intend to represent (see Peirce, 1931–1958, Vol. 4). Iconic simulations can leave distinct traces in behavior—for instance, studies by Ye et al. (2012) show that people are faster to process this sentence: "After the paper was submitted, the journal changed its policy …" than this one: "Before the paper was submitted, the journal changed its policy …" because the first sentence, but not the second, introduces the events in the same chronological order as they would have unfolded in real life. Indeed, much evidence suggests that humans construct mental timelines—a form of iconic mental simulation—when they comprehend how multiple events relate to one another (Bergen et al., 2012; Casasanto & Jasmin, 2012; Cooperrider & Núñez, 2009; Fuhrman & Boroditsky, 2010; Gevers et al., 2003, 2004; Ishihara et al., 2008; Leone et al., 2018; Maass & Russo, 2003; Santiago et al., 2007, 2010; Torralbo et al., 2006; Vallesi et al., 2014; Weger & Pratt, 2008; for reviews, see Bonato et al., 2012; Hoerl & McCormack, 2019). Children have difficulty processing sentences that describe events in reverse chronological order (Amidon & Carey, 1972; Clark, 1971; Pyykkönen et al., 2003; Pyykkönen & Järvikivi, 2012; cf., Amidon, 1976). Some researchers make the stronger claim that biases in temporal thinking come from the organization of neural structures, such that iconic mental timelines emerge from low-level attentional mechanisms (Chatterjee et al., 1999; Vicario et al., 2007) or more general cognitive mechanisms that apply across a variety of domains, such as space and number (see Winter et al., 2015 for a review).

Most investigations into how people reason about time have focused on punctate events—events that can be described as points in time, and whose durations are irrelevant to understanding their relations to other events. For instance, if you know that *the dinner occurred before the movie*, it may not matter if the dinner was hurriedly eaten in 15 min in the car on the way to the theater or if it was a longer meal at a restaurant. One reason for the focus on punctate events is that durations can be difficult to comprehend, particularly for young children. Children appear to use words that denote durations such as *day*, *week*, and *year* without understanding how long each referent lasts (Tillman & Barner, 2015; Tillman et al., 2018). Durations can be difficult for adults to encode, too: Descriptions of durations can be ambiguous, for example, the sentence *one meeting happened during another* could mean that one meeting started before the other ended or it could mean that the meetings both started and ended at the same time. They can be misperceived, too: The same slice of time can seem shorter or longer depending on how many events took place within it (Wang & Gennari, 2019; see also Isham et al., 2011; Xu & Kwok, 2019).

Before they ever learn talk about durations, infants can encode them nonsymbolically (Brannon et al., 2007; Provasi et al., 2011 ; vanMarle & Wynn, 2006). For example, Provasi et al. (2011) ;analyzed 4-month olds' looking times to discover that they can discriminate short-lasting sounds (500 ms) from long-lasting ones (1,500 ms). Indeed, much work on duration processing concerns such timing tasks in animals, children, and adults (see Gibbon, 1977; Meck & Church, 1983; Odic, 2018); many theorists posit that people encode in memory only the temporal order of events (Anderson, 1982) and not

This article was published Online First September 7, 2023.
Sangeet Khemlani https://orcid.org/0000-0002-1158-553X
Portions of this manuscript were presented at CogSci 2020, at Psychonomics 2020, and at the University of Granada in 2022, and we thank the audiences for their feedback and critiques.

This research was performed while the first author held an National Academy of Sciences Research Associateship award at the U.S. Naval Research Laboratory. We are grateful to Kalyan Gupta and Danielle Paterno at the Knexus Research Corporation for their help in conducting the experiments. And we thank Bill Adams, Branden Bio, Gordon Briggs, Monica Bucciarelli, Hillary Harner, Tony Harrison, Laura Hiatt, Phil Johnson-Laird, Joanna Korman, Markus Knauff, Andrew Lovett, Greg Trafton, and Ben Wright for their advice and comments.

Raw data, statistical analyses, and code for the experiments are available at https://osf.io/3waqe/.
Correspondence concerning this article should be addressed to Sangeet Khemlani, US Naval Research Laboratory, Navy Center for Applied Research in Artificial Intelligence, 4555 Overlook Avenue South West, Washington, DC 20375, United States. Email: sunny.khemlani@nrl.navy.mil

information about their durations, which they can infer from those temporal orders (see also Ornstein, 1969). In contrast, Droit-Volet and Rattat (1999) argue that between the ages of 3 and –5, children learn to map actions with typical durations. Earlier studies show that durational thinking continues to develop into late childhood (Levin, 1977). Few studies investigate children's explicit, conscious reasoning about durations and intervals, perhaps because they may not comprehend the formal definitions of time words (e.g., *minute*) until they become a part of their formal instruction (Tillman & Barner, 2015).

By adulthood, reasoners can consciously reason about durations without any special training in logic, though very few studies investigate the mechanisms underlying conscious durational thinking. For instance, suppose you know the following:

1. The 2022 harvest lasts from August to October.
   The subsequent winter lasts from December to March.

The following temporal inferences follow from (1):

2. a. The harvest happened before the winter.
   b. The harvest *started* before winter *started*.
   c. The harvest *ended* before winter *started*.
   d. The harvest *started* before winter *ended*.
   e. The harvest *ended* before winter *ended*.

The inference in (2a) concerns a relation, *before*, between two different events, the *harvest* and the *winter*. This inference does not reflect the durational nature of the events, that is, the events could be treated as punctate. The remaining inferences (2b–e) concern relations that refer to when some events starts and when others end. People often treat events as having parts and subevents, that is, they can be organized into partonomic hierarchies (Miller & Johnson-Laird, 1976; Tversky, 1989; Tversky & Hemenway, 1984; Zacks & Tversky, 2001). To refer to a part of an event is to imply there are other parts—that the event is extended in time. Hence, the inferences in (2b–e) are inherently durational.

To represent an event's duration, a reasoning system must, at a minimum, represent its start and its end. However, theories have yet to propose what people represent when they reason about durations expressed using natural language relations, for example, words such as "during," and few studies have examined systematic patterns of durational reasoning. People may represent mental timelines for durational relations, just as they appear to do for punctate ones. Mental timelines are a precursor to the technologies used for metric timekeeping, for example, calendar and scheduling systems. Mental timelines may be different from the representations that underlie recent artificial intelligence systems developed to describe temporal behavior (such as linear temporal logic; see, e.g., Giacomo & Vardi, 2013; Rozier, 2011) because they are likely to be constrained in ways that make them easy for people to encode, recall, and reason about.

In what follows, we first review treatments of durational reasoning in artificial intelligence and formal logic and explain their limitations as the basis for cognitive theories. We then present a computational cognitive theory of how people build mental models from descriptions of durations (Khemlani, 2022; Khemlani et al., 2015). The theory extends and clarifies the proposal that people construct a mental timeline to represent events. We show why a central prediction of the theory—that inferences emerge from the way people construct and scan models—yields a bias in the way people consciously process descriptions of durations. We describe six experiments that test and validate the bias, and marshal the evidence in light of theories of temporal cognition.

The Logic and Psychology of Durational Reasoning

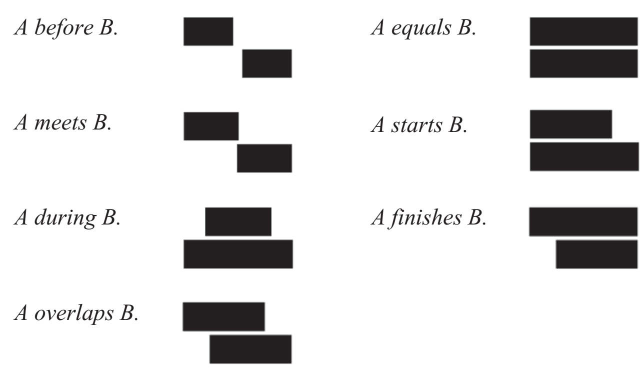
Durational inferences can vary in complexity. Consider this deduction:

3. The 2022 harvest happened during the fall.
   The annual vacation happened during the winter.
   The winter happened after the fall.
   Therefore, the vacation happened after the harvest.

The final conclusion is *valid*, that is, it is true in every possible case in which the premises hold (Jeffrey, 1981, p. 1). Now consider the conclusion in (4):

4. The 2022 harvest happened during the fall.
   The annual vacation happened during the fall.
   The winter happened after the fall.
   Therefore, the vacation happened after the harvest.

The conclusion is invalid—it is possible, but not necessarily the case: The vacation could have occurred before or during the harvest within the fall season. Certain systems of logic can provide a formal way to deduce valid conclusions from temporal descriptions (e.g., Allen, 1983; Fischer et al., 2004; Freksa, 1992; Goranko et al., 2004; Kowalski & Sergot, 1989; Pnueli, 1977). Temporal logics provide a basis for formalizing interval relations, that is, the way in which one interval of time exists relative to another. Allen's interval algebra (1983), for instance, is a system that specifies all possible relations between the endpoints for two different events, *A* and *B*, as depicted in the following diagrams:



The horizontal lines represent the way an event endures across multiple points in time. Allen's algebra specifies 14 relations, which include the seven relations above along with their inverses (e.g., the inverse of *A before B* is equivalent to *B after A*), but we omit them for brevity. Some of the Allen relations have intuitive mappings onto temporal prepositions and connectives in natural language, for example, *Event A occurred before event B*. Other such relations can be expressed in natural language by combining relations and referring to parts of events, for example, the *meets* relation can be expressed in the following natural language description: *Event A ends at the same time as event B starts*. The description is composed of the temporal verbs *end* and *start*, as well as the temporal preposition *at the same time as*. In contrast, Allen's calculus treats the relation as primitive. As Knauff and colleagues have argued, the disparity between logic and language precludes systems such as the interval algebra from serving as the basis of plausible accounts of spatiotemporal reasoning (Knauff, 1999; Knauff et al., 2004; Rauh et al., 2005). Other sorts of logical system, such as linear temporal logic (Manna & Pnueli, 1992, 1995; Pnueli, 1977), introduce alternative temporal primitives, such as the operation *until* (as in, *Event A happens until X is true*) and *next* (as in, *Event B occurs in the next timepoint*). Such operators seem relevant to many applied domains, such as scheduling shipments and reservations, but bear little resemblance to the ways people describe and reason about duration.

Khemlani et al. (2015) sought to explain inferences such as (3) and (4) without recourse to any system of logic. They argued that people reason about durations by constructing mental simulations of events—mental models (Johnson-Laird, 2006). The "model" theory applies to relational reasoning across several different domains (Goodwin & Johnson-Laird, 2005), including reasoning about space (Jahn et al., 2007; Knauff, 2013), causality (Goldvarg & Johnson-Laird, 2001; Khemlani et al., 2015, 2021), counterfactuals (Byrne, 2005), and even punctate events that have no explicit duration (Schaeken et al., 1996; Schaeken & Johnson-Laird, 2000; von Hecker et al., 2019). The theory rests on three fundamental constraints:

First, models are *iconic*, that is, their structure maps onto the structure of the things they represent (Peirce, 1931–1958, Vol. 4). An iconic treatment of events permits one of two sorts of representation: Reasoners can represent events as tokens arranged along a spatial axis that represents time (Schaeken et al., 1996) or else as simulations that unfold in the same sequence as the events do in the real world (Khemlani et al., 2013). Evidence supports the construction of iconic simulations: For instance, people encode temporal narratives in memory by grouping events closer in conceptual time (Diessel, 2008; Nieuwland, 2015; Xu & Kwok, 2019; Zwaan, 1996; Zwaan et al., 2001) and by organizing those events in a particular direction (Matlock et al., 2005; von Hecker et al., 2019). They likewise take longer to process narrative flashbacks, for example, events that occurred a longer time ago in an ongoing narrative (Claus & Kelter, 2006). They may incorporate other iconic elements into temporal representations, for example, they may represent longer durations for slower and more complex motion (Matlock, 2004; von Sobbe et al., 2021; see also von Sobbe, 2022). To represent iconic simulations of durations, Khemlani et al. (2015) propose that they use discrete tokens to encode the starts and ends of events for further processing. Accordingly, if they engage in abstract reasoning about temporal intervals, reasoners may disregard information about motion or metric time and reason on the basis of intervals alone. Recent neurocognitive evidence suggests a neural substrate for such representations (Ezzyat & Davachi, 2021; Zheng et al., 2022). The following diagram depicts a mental model of (3):

```
    [ harvest ]    [ vacation ]
[     fall     ] [    winter    ]
```

It shows four events whose durations are denoted by tokens representing when events start and end, that is, the opening and closing brackets. We separate the events on different lines to make the diagram legible, but the mental representation underlying it needs only a single ordinal dimension to represent time. A computational model that implements the theory (Khemlani, 2022) represents the mental model as a list of discrete timepoints:

$\text{fall}_{\text{START}}$ $\text{harvest}_{\text{START}}$ $\text{harvest}_{\text{END}}$ fall_{END} $\text{winter}_{\text{START}}$ $\text{vacation}_{\text{START}}$ $\text{vacation}_{\text{END}}$ $\text{winter}_{\text{END}}$

The words in these tokens serve as a shorthand to label the events: they specify the event's content but not its structure. An advantage of the theory is that the representation is agnostic to the metric length of the duration, so the theory's predictions are the same for inferences about minutes or decades. Reasoners appear to have no difficulty reasoning about longer events, which suggests that metric time may play no part in many temporal inferences. Second, reasoners construct models incrementally. Figure 1 shows how reasoners can build a temporal mental model from a set of premises describing punctate and durative relations. Reasoners draw conclusions by building and scanning models in a particular direction, for example, they can scan the models in Figure 1A from the harvest to the end of the winter (see, for example, Bonato et al., 2016 for neurological evidence corroborating such a representation; and Khemlani & Johnson-Laird, 2022 for discussion of an analogous scanning procedure in quantitative reasoning). By default, people build and scan models iconically, from earlier to later events (see Figure 1A and 1C). In certain cases, they can scan models in the reverse chronological order, but doing so takes additional effort and can make them more difficult to process (Figure 1B; see also Münte et al., 1998; Ye et al., 2012). Nevertheless, evidence shows that they can spontaneously form strategies for reasoning about spatiotemporal relations (Ragni & Knauff, 2013; Schaeken & Johnson-Laird, 2000).

Models help explain how people understand durational relations in natural language, such as *during* and *while*. In some cases, durational relations emerge, not from relations such as *during*, but from explicit information about the beginning and ends of events (Figure 1C). Consider again the introductory example:

1. The 2022 harvest lasts from August to October.
   The subsequent winter lasts from December to March.

The model theory posits that durational relations concern events that are defined by their onsets and offsets: The harvest in (1), for instance, concerns its initiation in August and is completion in October. The specific months suffice to anchor the starts and the ends of the events, and so a resulting model of (1) may be:

$[_{\text{AUG}}$ harvest $]_{\text{OCT}}$ $[_{\text{DEC}}$ winter $]_{\text{MAR}}$

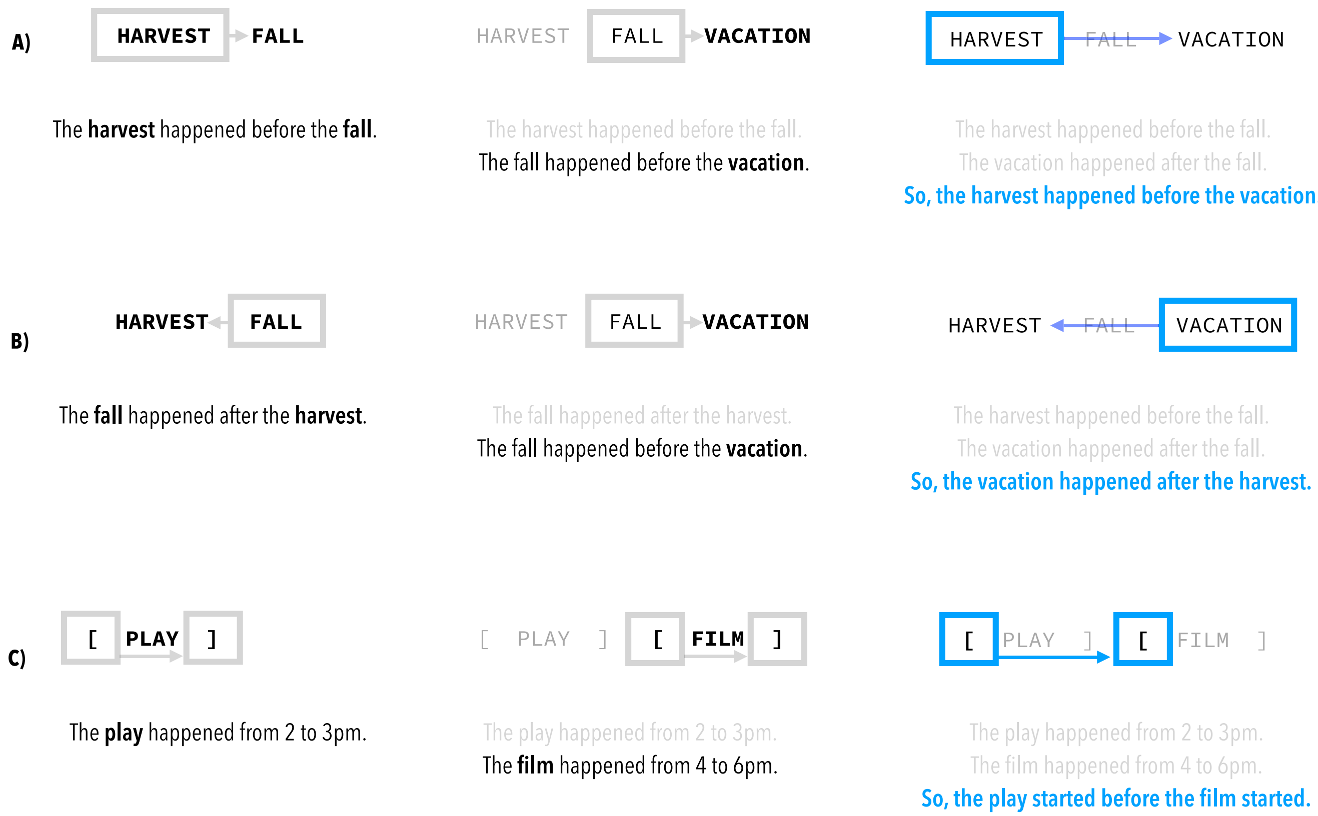
Unless it's important to keep track of the specific months, many reasoners may disregard them and reason on the basis of the intervals alone:

```
[   harvest  ]      [    winter    ]
```

By scanning the model, the relation between the two events becomes clear: The harvest happened before the winter, or, analogously, the winter happened after the harvest. The theory accordingly explains how reasoners can make rapid inferences about the organization of the starts and ends of events. For instance, given the premises in (1), is it true that the harvest ended after the winter started? No: Scanning the model shows that the harvest ended before the winter started (see Figure 1).

Figure 1

The Incremental Construction of Mental Models of Time



Note. Panels (A and B) show the way people interpret premises to update temporal models. Iconic biases are a consequence of model construction. The grey boxes highlight the structural elements of models that serve as reference points (arrows) from which new pieces of information (in bold) are incrementally added. The inference, indicted by shown in Panel (A) is easy to deduce: Reasoners need to scan the model in the same direction in which it was built, that is, left to right. The events in Panel (B) are similar to those in Panel (A), except that the premises introduce them in a different chronological order; they should be harder to process, and the inference should be more difficult to make. The same principle explains how people process durations (Panel C); the theory predicts that people should prefer conclusions that reflect the iconic pattern of construction, such as "The play started before the film started" (shown) over alternatives such as "The play ended before the film ended" (not shown). See the online article for the color version of the figure.

The Allen calculus, introduced above, enumerates the ways in which two durative events can relate to one another. To date, no studies have focused on how reasoners comprehend such configurations of events or how they process them. The model theory of durative reasoning posits that reasoners build and scan temporal models incrementally, and previous studies reveal systematic errors that reasoners make in line with the theory's predictions (Kelly et al., 2020). Here, we explore systematic *iconicity* biases that the theory predicts: that is, directional biases about how individuals process the starts and ends of events (Figure 1C). For instance, consider which of the following two statements is a better description for (1):

5. a. The harvest *started* before the winter *started*.
   b. The winter *ended* after the harvest *ended*.

The two descriptions are both accurate but incomplete characterizations of the scenario described in (1). Neither statement serves as an objectively better description than the other. No study has investigated whether people should express any preference between the two options. Conventional logical calculi treat both statements as valid inferences, so they predict no bias whatsoever. Theorists such as Casasanto (2016) point out cross-linguistic evidence that the ways reasoners speak about time may not align with the ways they think about time—so if a bias exists, it may have nothing to do with mental operations concerning temporal thinking. Perhaps people prefer the word *start* because it occurs more frequently, and so they should prefer (5a) to (5b) on the basis of word frequency alone. In contrast, if people reason by way of constructing and scanning iconic mental simulations, doing so may yield a scanning bias, one that prefers chronologically oriented descriptions of earlier timepoints to later ones. Hence, people should consider (5a) to be a better description than (5b) across a wide variety of tasks and scenarios.

We report six experiments that test and validate the presence of an iconicity bias, that is, a preference for descriptions of when events start to when they end. The experiments surveyed all seven Allen relations, that is, all the possible ways that two durations can exist relative to one another. Each of the studies asked participants to select the most appropriate description from a pair of incomplete but accurate alternatives, one of which was congruent with an iconic scanning bias. Study designs ruled out alternative explanations and addressed confounding variables. Across these methodologies, the data revealed robust biases in line with the model theory's prediction.

Experiment 1

To investigate whether people exhibit an iconic scanning bias, that is, whether they prefer descriptions relating earlier timepoints over later timepoints, Experiment 1 presented participants with a description of the durations of two events, for example,

> The encryption started at 1 pm and ended at 11 pm on Monday.
> The download started at 9 am and ended at 11 pm on Monday.

These premises reflect the *finishes* Allen relation (see the introductory part), because the two events end at the same time. Participants read descriptions of events that reflected *finishes* and three other Allen relations (*during, equals,* and *starts*).

For the example above, participants selected which of three different statements better summarized how the events relate to one another:

> The download started before the encryption started. (iconic)
> The encryption ended when the download ended. (noniconic)

The experiment provided one additional option, that is, "neither description is better than the other," to allow them to explicitly exhibit no preference.

Two separate outcomes of the experiment can validate the model theory: The first is a strong validation, where participants prefer descriptions of start points over the other two response options. The second is a pattern in which participants select the "neither" option most often, but secondarily prefer descriptions of start points to end points. The theory is falsified in any scenario in which participants select descriptions of end points more often than start points.

Method

Participants

In total, 167 participants completed the experiment for monetary compensation ($2.50) through Amazon Mechanical Turk (AMT), commensurate with minimum-wage standards. All of the participants were native English speakers, and 50% had taken one or fewer courses in introductory logic. Fourteen participants were excluded from the analysis, six for having a total time including reading the instructions of less than 2 min, six for having three or more trials with a response latency greater than 1.5 min, and two for nonsense input in a postexperimental questionnaire. The analyses reported below are based on the remaining 153 participants (74 female, $M_{\text{age}} = 36.9$).

Preregistration and Open Science

The experimental design was preregistered through the Open Science Framework platform (https://osf.io/3waqe/). The same link makes the data from the study available.

Task and Design

Participants carried out 16 problems describing the durations of two events. These durations corresponded to time intervals in the pattern of four different Allen relations (diagrammed above): *during*, *equals*, *finishes*, and *starts*, and participants received four problems of each type. Hence, the experiment implemented a fully within-participant design. Participants were asked to indicate which of two descriptions was better: a statement describing a relation between the start points of the two events or a statement describing a relation between the end points. The response options used the temporal connectives "before" and "when" in a manner that appropriately reflected the Allen relation provided. Hence, the two response options were incomplete but accurate descriptions of the given scenario; neither description was more accurate than the other.

Materials

Each problem description required two event labels and four timepoints (the start and end points for each event) to yield statements of the form: *(Event) started at (Timepoint 1) and ended at (Timepoint 2)*. The two event labels were randomly selected from a pool of 24 event labels that correspond to events relevant to computer networking. The set of events were designed to plausibly co-occur (see Appendix). Across the 16 problems, each pair of labels was unique and each label was used at most two times. The start and end points of the two events were randomly generated to correspond to the Allen relation assigned to each problem. To prevent participants from interpreting the two events as occurring on different days, each statement describing an event was appended with a temporal preposition describing a particular day of the week, for example, *the encryption started at 1 pm and ended at 11 pm on Friday*. The day of the week was randomly assigned. As a consequence, no two participants received the same set of materials. The study randomized the order of the 16 problems.

Results and Discussion

Table 1 reports the response distributions for Experiment 1 as a function of the four different types of problems given to participants, that is, as a function of different types of Allen relation. Participants selected iconic descriptions (those that used the word "started") more often than descriptions that used the word "ended" (45% vs. 38% of responses). In theory, participants would be justified in selecting the "neither" option on 100% of responses because the other options were both accurate. In reality, they did so less than 18% of the time: They selected "neither" most often (28% of the time) for descriptions of the *equals* relation.

To evaluate how each Allen relation predicted the distribution of participants' responses, we subjected their data to a multinomial discrete-choice logit regression (see Agresti, 2002) using the *lme4* package in R (Bates et al., 2015). Multinomial modeling is a generalization of logistic regression that predicts the outcome probabilities between more than two categorical outcomes, provided that reference categories for each discrete value can be stipulated. A statistically significant effect of a predictor implies that it predicts a reliable amount of unique variance in a particular discrete outcome. The analysis regressed the conditions in the study (the Allen relations, using *equals* as a reference category) against the outcomes ("A started …" or "neither," using "A ended …" as a reference category). Hence, the results of the analysis reveal whether a particular Allen relation (e.g., *during*) predicted participants' tendency to select a response over and above the reference categories.

Because "A started…" responses are iconic, and "A ended…" responses are noniconic, an iconicity bias should predict "A started…" responses for each of the individual Allen relations.

Table 1

The Percentages of Participants' Selections of Response Options Corresponding to A Started (Before/When) B Started and A Ended (Before/When) B Ended in Experiment 1 as a Function of the Four Separate Types of Problems They Received, Each of Which Corresponded to a Different Relation in Allen's (1983) Interval Algebra

	Allen relation between events A and B			
	During	Equals	Finishes	Starts
Type of description				
Iconic description				
A started (before/when) B started.	**47**	**44**	**45**	43
Nonconic description				
B ended (before/when) A ended.	38	28	41	43
Neither	15	28	14	13

Note. Bold values denote the most selected option.

Table 2 summarizes the results: They show that *during, starts,* and *finishes* explained a significant amount of the variance for "neither" responses relative to "A ended…" responses ($ps < .001$), and they show that *finishes* and *starts* explained a significant amount of the variance for "A started…" responses relative to "A ended…" responses ($ps < .003$ for *during* and *finishes*). The *during* condition yielded only anecdotal evidence ($p = .052$) in favor of a bias toward "A started…" responses.

To flesh out these results, we carried out a series of nonparametric pairwise analyses.

Nonparametric Analyses

To examine participants' response patterns, we subjected the frequencies of responses as a function of the four Allen relations to a Fisher exact test, which showed that the distributions were reliably different from chance ($p < .001$); separate follow-up Fisher's tests for each individual relation revealed analogous differences ($ps < .001$) in line with the multinomial model reported above.

Next, we dummy-coded the responses and subjected them to pairwise nonparametric analyses. Participants exhibited an overall preference for "A started…" descriptions vs. "A ended…" descriptions (45% vs. 38%; Wilcoxon test, $z = 3.24$, $p = .001$, Cliff's $\delta = 0.22$). They exhibited a preference for "A started…" versus "A ended…" in the *during* (47% vs. 38%; Wilcoxon test, $z = 2.04$, $p = .04$, Cliff's $\delta = 0.17$) and *equals* condition (44% vs. 28%; Wilcoxon test, $z = 3.98$, $p < .001$, Cliff's $\delta = 0.29$), but not in the *finishes* condition (45% vs. 41%; Wilcoxon test, $z = .88$, $p = .38$, Cliff's $\delta = 0.07$) or the *starts* condition (43% vs. 43%; Wilcoxon test, $z = 0.11$, $p = .91$, Cliff's $\delta = 0.002$).

The results of Experiment 1 were mixed: On the one hand, participants exhibited an iconicity bias insofar as they selected "A started…" more often than the other options. On the other hand, they did so on a minority of the trials, and they selected the "neither" option on roughly a fifth of the trials. In retrospect, providing participants with a "neither" option may have served to alert them to the possibility that both options were accurate and curtailed their natural preferences. If so, then experiments without such an option should yield stronger biases toward iconic responses. Alternatively, it may be that participants exhibit no bias between the two options whatsoever, in which case eliminating the "neither" option should result in no significant differences in selection between iconic and noniconic options. Indeed, it is possible that the iconicity bias only manifests in the presence of the "neither" option. Accordingly, Experiment 2 and the studies that follow provided participants with only two options on each trial, an iconic option and a noniconic option.

Experiment 2

Experiment 2 was equivalent to Experiment 1 in all respects save one: Participants evaluated two options of the relation between two durational events, an iconic and a noniconic description, without a "neither" option. If an iconicity bias in reasoning about durations is robust, people should exhibit it when they are not alerted to the

Table 2

Results of a Multinomial Regression Model That Regressed the Four Allen Relations Against Participants' Responses in Experiment 1

Response	Allen relation between events A and B	Coeff. estimate	*SE*	*z* value	*p* value
"Neither"	*during*	−0.94	0.16	−5.76	<.001
	finishes	−1.07	0.17	−6.46	<.001
	starts	−1.21	0.17	−7.25	<.001
"A started…"	*during*	−0.26	0.13	−1.94	.052
	finishes	−0.38	0.13	−2.90	.003
	starts	−0.47	0.13	−3.63	<.001

Note. The model treated the noniconic response ("A ended…") as well as the *equals* Allen relation as reference categories.

possibility that neither description is preferable. The theory predicts a preference in favor of descriptions concerning when events “started” to those when events “ended.”

Method

Participants

Because participants evaluated only two response options in Experiment 2, a power analysis revealed that 50 participants were sufficient to test the model theory’s prediction. They completed the study on AMT for monetary compensation ($1.50). All of the participants were native English speakers, and all but five had taken one or fewer courses in introductory logic. One participant was excluded from the analysis for always selecting the first response option. The analyses reported below are based on the remaining 49 participants (21 female, $M_{\text{age}} = 34.2$).

Preregistration and Open Science

The overall pattern observed in Experiment 1 was preregistered as the hypothesized results for this experiment and subsequent studies through the Open Science Framework platform (https://osf.io/3waqe/). The same link makes the data from the study available.

Task and Design

The task and design of Experiment 2 were similar to Experiment 1, except that Experiment 2 did not provided participants with a “neither” response option (see Experiment 1).

Procedure

Each experimental problem began by displaying the premises describing the durations of the two events. After a 2-s delay, two response options appeared as buttons on the screen along with the question: “Which of the following is a better description?” The participant responded by clicking on one of two buttons, corresponding to the following two options:

A started (before/when) B started. (iconic)
B ended (before/when) A ended. (noniconic)

Only once they registered their response would the experiment move onto the next trial. The order of the response options was randomized for each problem.

Design, Materials, and Procedure

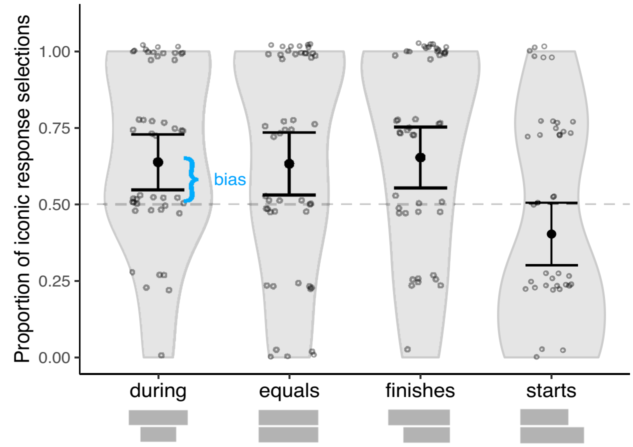
The instructions were altered to reflect the additional response option. Otherwise, the design, materials, and procedure were the same as in Experiment 1.

Results and Discussion

Figure 2 displays the proportion of iconic responses in Experiment 2. The data were subjected to a generalized linear mixed model (GLMM) with a maximal random effects structure regression analysis. The analysis utilized a maximal random-effects structure (following Barr et al., 2013) that controlled for variance contributed by participants, materials, and individual interpretations of Allen relations. It made use of a logit link function, and it treated

Figure 2

Violin Plots of the Proportions of Iconic Response Selections as a Function of the Four Separate Types of Problems in Experiment 2, Each of Which Corresponded One of Four Relations in Allen’s (1983) Interval Algebra



Note. Light gray circles denote individual participants’ mean proportions of iconic responses. Dark black circles denote mean proportions across all participants. The dashed gray line denotes chance performance. Error bars denote 95% confidence intervals. See the online article for the color version of the figure.

Allen relations as fixed effects. Overall, participants chose iconic descriptions over noniconic ones; they did so 58% of the time, a rate significantly above chance, as the intercept of the model revealed ($B = -0.45$, $SE = 0.20$, $p = .02$). Indeed, 33 of the 49 participants displayed this pattern (binomial test, $p = .02$). The result corroborates an iconicity bias in durational thinking. Three of the four individual conditions likewise yielded such a bias (see Figure 2): Participants selected iconic descriptions for *during* 64% of the time ($B = 1.10$, $SE = 0.22$, $p < .001$); for *equals*, 63% of the time ($B = 1.08$, $SE = 0.22$, $p < .001$); and for *finishes*, 65% of the time ($B = 1.18$, $SE = 0.23$, $p < .001$). They showed the reverse bias for *starts:* For this relation, they selected iconic descriptions 40% of the time. Nonparametric analyses (provided at https://osf.io/cunbk) corroborated the results of the regression.

Experiment 2 accordingly replicated and extended the results of Experiment 1. One limitation of the preceding experiments, however, is that they both used verbal premises of the form:

The encryption started at ___ and ended at ___…
The download started at ___ and ended at ___…

That is, the premise introducing each event described the start of that event before its end—so, the iconicity bias exhibited by the participants may have been an artifact of the premises and not a bias inherent to how reasoners process durations. Experiment 3 ruled out this deflationary explanation.

Experiment 3

Experiment 3 was similar to the preceding experiments: It provided participants with information about two events that related one to the other in four different ways that correspond to the *during,* *equals, finishes,* and *starts* relations in Allen's interval calculus. Instead of presenting verbal descriptions to participants, however, the events appeared in diagrams that depicted the two events as bars (akin to diagram conventions explored by Gantt, 1910; see Figure 3).

The experiment instructed participants on how to comprehend such charts, and it asked participants to select from two accurate but partial descriptions identical to those provided in the previous studies, that is, an iconic option and a noniconic option.

Method

Participants

Fifty participants (23 female; $M_{\text{age}} = 36.5$) completed the experiment for monetary compensation ($1.50) through AMT. All of the participants were native English speakers, and 44 participants had taken one or fewer courses in introductory logic.

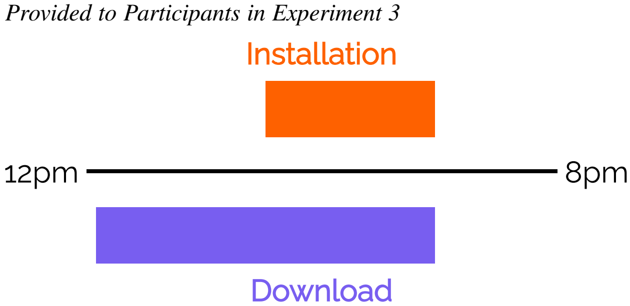
Design and Materials

The task and design of Experiment 3 was similar to Experiment 2 in all ways except the following: Experiment 3 used diagrams (see Figure 3) instead of verbal descriptions to inform participants about when events started and ended. Each diagram consisted of two bars of different colors. Colors were selected from a palette that could be easily distinguished by individuals with colorblindness (IBM Corp, 2021). The diagrams also consisted of a bar denoting an 8-hr span of time: The span was restricted to ensure concise and interpretable diagrams. Each diagram included labels for the event names, start and end times, and the timespan. Unlike in Experiment 2, no information about the day of the week was provided to participants because each diagram was designed to depict a single span of time. Participants selected a preferred description from two alternatives: one iconic and one noniconic.

Procedure

The experiment made use of different instructions to introduce the diagrams that depicted two events. Otherwise, the procedure was the same as in Experiment 2.

Figure 3
Diagram of Two Events, the Installation and the Download, Provided to Participants in Experiment 3



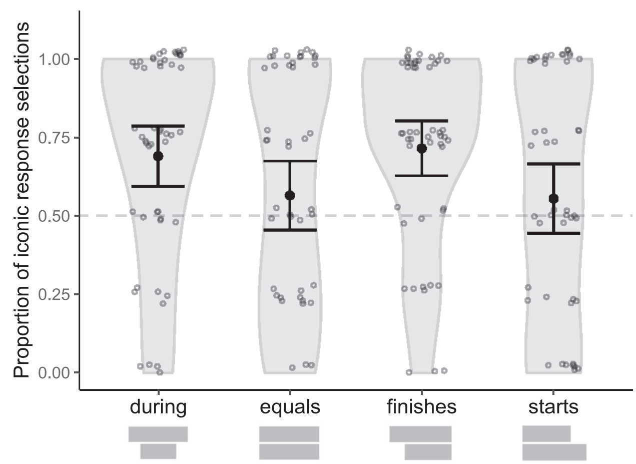
Note. Each event had a unique label and color, and the events were depicted relative to a timeline depicted in black, which spanned 8 hr. See the online article for the color version of the figure.

Results and Discussion

Figure 4 shows the proportion of iconic response selections in Experiment 3 as a function of the four different types of problems in the study. Data were subjected to a GLMM regression, but the maximal random-effects structure failed to converge, and so we reduced the structure until it yielded a model that did converge. Participants preferred iconic descriptions 63% of the time, that is, in the predicted direction, yet this overall preference was unreliable (intercept: $B = 0.30$, $SE = 0.31$, $p = .32$); 34 of the 50 participants displayed this pattern (binomial test: $p = .02$). Two of the four individual relations yielded a significant bias (see Figure 4): Participants selected iconic descriptions for *during* 69% of the time ($B = 0.73$, $SE = 0.23$, $p = .002$) and they did so 72% of the time for *finishes* ($B = 0.89$, $SE = 0.24$, $p < .001$). while neither *starts* nor *equals* reliably predicted iconic responses, they both yielded more iconic than noniconic responses, that is, the *starts* condition did not yield a bias in the opposite direction. Nonparametric analyses (provided at https://osf.io/4jndx) corroborated the results of the regression.

The results of Experiment 3 reveal iconicity biases in participants' preferences for descriptions that relate two visually depicted durations. Yet, Experiment 3 and the preceding studies all had response options with the same limitations: In each of the experiments, participants selected their preferences for descriptions of the form *A started before B started* or else of the form *A started when B started.* One explanation for the observed iconicity biases may be that they came from how people process the temporal connective "before," that is, "before" may make participants prefer response options that describe chronological orders to those that describe reverse chronological orders. Experiment 4 ruled out this explanation by counterbalancing descriptions that use "before" and "after" as temporal connectives.

Figure 4
Violin Plots of the Proportions of Iconic Response Selections as a Function of the Four Separate Types of Problems in Experiment 3, Each of Which Corresponded One of Four Relations in Allen's (1983) Interval Algebra



Note. Light gray circles denote individual participants' mean proportions of iconic responses. Dark black circles denote mean proportions across all participants. The dashed gray line denotes chance performance. Error bars denote 95% confidence intervals.

Experiment 4

Experiment 4 sought to test whether the iconicity bias observed in the previous studies came about because participants processed the temporal connective "before." Some research supports the idea that "before" can be easier to process than "after" (see, e.g., Hoeks et al., 2004; but cf., Politzer-Ahles et al., 2017). Other accounts suggest that people have difficulty processing events described in reverse chronological order (see, e.g., Münte et al., 1998; Ye et al., 2012). It may be that presenting sentences of the form *A started before B started* creates, rather than measures, a preference for descriptions in chronological order. To address the issue, Experiment 4 presented participants with verbal descriptions, for example,

> The encryption started at 12 pm and ended at 2 pm on Monday.
> The download started at 5 am and ended at 11 pm on Monday.

It manipulated the temporal connectives ("before" vs. "after") in the two options participants could select from. For instance, in one condition, participants chose between both iconic and noniconic descriptions using the connective "after":

> The encryption started after the download started. (iconic)
> The download ended after the encryption ended. (noniconic)

As in previous experiments, both of the options are accurate. The difference between the two options is that the first concerns the initiation of two durational events—initiations occur earlier in a durational model—and the second concerns their conclusion, which occurs later in the model. So, reasoners who prefer the first description to the second do so because of a bias toward iconicity, that is, scanning an iconic model from earliest to latest timepoints. Another condition presented the two options using "before," for example,

> The download started before the encryption started. (iconic)
> The encryption ended before the download ended. (noniconic)

And two other conditions presented alternative descriptions in which one used "before" and the other used "after," for example,

> The download started before the encryption started. (iconic)
> The download ended after the encryption ended. (noniconic)

Because of these different presentation conditions, and to keep the number of problems tractable for participants, the study presented participants with only two Allen relations: *before* (see the introductory part) and *during*. In sum, the experiment implemented a 2 (Allen relation: *before* vs. *during*) × 4 (temporal connectives: "before" vs. "after" in the iconic and noniconic descriptions) repeated-measures design. The results of the study are straightforward: In all eight conditions, participants preferred iconic descriptions.

Method

Participants

Fifty participants completed the experiment for monetary compensation ($2.00) through AMT, commensurate with minimum-wage standards. One participant was excluded from the analysis for always selecting the first response option. The analyses reported below are based on the remaining 49 participants (eight female, $M_{\text{age}} = 36.2$, 48 native English speakers). Thirty-eight participants had taken one or fewer courses in introductory logic.

Design, Materials, and Procedure

Participants carried out 24 problems. In each problem, they received a description of the durations of two events. The durations corresponded to two distinct intervals described in Allen's calculus: *during* and *before.* As in each of the previous studies, participants had to decide whether an iconic or a noniconic description of the events better characterized the scenario. In addition to the intervals, the experiment manipulated whether the iconic description used "before" or "after" as a temporal connective and likewise whether the noniconic description used "before" or "after." Participants carried out three problems in each of the eight conditions of the experiment. The experiment randomized the display order of the descriptions. In all other respects, the experiment was similar to Experiment 2.

Results and Discussion

Table 3 shows the percentages of iconic descriptions selected in Experiment 4 as a function of the two different relations and the four pairs of response options given to participants. Participants preferred iconic descriptions 64% of the time across the experiment as a whole. As in Experiment 2, data were subjected to a GLMM with a maximal random effects structure; the analysis yielded a significant intercept, which reveals an overall bias in favor of iconic descriptions across all study conditions ($B = 0.85$, $SE = 0.24$, $p < .001$). No other main effects or interactions were significant. Follow-up nonparametric analyses corroborate the results from the regression (provided at https://osf.io/tmw2q).

Experiment 4 allowed us to examine whether participants prefer the chronology simulated in a mental model—*iconic* chronology—over the chronology expressed in the response options. The results reveal that people exhibit a bias toward iconic chronology, as the theory predicts, and not a bias toward response chronology. The two routinely coincide, and previous studies have suggested that reasoners have difficulty processing reverse chronological orders (Münte et al., 1998; Politzer-Ahles et al., 2017; Ye et al., 2012), but the results of this experiment reveal a more subtle effect. That is, participants preferred descriptions of the form, *B started after A started*, over descriptions such as, *B ended after A ended* (see Table 3); both of descriptions reflect a reverse chronological order. The difference between the two is that the first concerns the initiation of two durational events—initiations occur earlier in a model of the events—while the second concerns their conclusion, which occurs later in the model.

Despite the converging results of the preceding studies, Experiments 1–4 reflect several notable constraints. In each of the studies, response options placed the adverbials "after," "before," and "when" in the middle of each pair of temporal expressions, for example,

> A started *before* B started.

which linguists refer to as *sentence-final* position because it comes at the end of the first temporal expression (Diessel, 2008; Politzer-Ahles et al., 2017). An alternative way of expressing the same relation is to place a temporal adverb before the first temporal expression (that is, *sentence-initial* position):

> *After* A started, B started.

To maintain the meaning of the relation between the two events, the adverbial "before" must be swapped with its opposite, that is,

Table 3
*The Percentages of Participants' Selections of the Iconic Response Options in Experiment 4 as a Function of Two Types of Allen Relations (*During *and* Before*), and as a Function of Four Separate Option Pairs Using the Temporal Connectives "Before" and "After"*

Temporal relation		Four separate pairs of descriptions for events A and B			
		A started before B started. B ended before A ended.	B started after A started. B ended before A ended.	A started before B started. A ended after B ended.	B started after A started. A ended after B ended.
during		57%	58%	60%	66%
		A started before B started. A ended before B ended.	B started after A started. A ended before B ended.	A started before B started. B ended after A ended.	B started after A started. B ended after A ended.
before		69%	60%	71%	67%

Note. For each pair of options, the top is iconic and the bottom is noniconic. The remainder of the percentages in each cell is the selection of noniconic responses.

"after." This constraint presents an opportunity to test an interaction predicted by the iconic processing (one that was highlighted to us by an anonymous reviewer). Experiment 5 tested the prediction.

Another limitation of the previous experiments is that they examined five of the seven possible relations between two events as described in Allen's (1983) interval calculus: *during*, *equals*, *finishes*, *starts*, and *before*. The other remaining relations, that is, the *meets* relation (which is when a preceding event ends at exactly the same time as a succeeding event starts) and the *overlaps* relation (when a preceding event ends after a succeeding event starts; see the introductory part) have gone unexplored. Experiment 5 addressed this discrepancy as well.

Experiment 5

If participants base descriptions of events on the way they scan an iconic mental model, then those scans should yield an interaction in the way they process sentence-initial and sentence-final descriptions:

> After A started, B started.
> A started before B started.

The two expressions above are both iconic, that is, they describe the starts of the events in chronological order. In contrast, these two expressions are noniconic:

> Before B started, A started.
> B started after A started.

In Experiment 5, participants selected from two different options, one iconic and one noniconic, just as in the previous studies. Unlike previous studies, half the trials in Experiment 5 presented participants with sentence-initial descriptions and half presented sentence-final descriptions (see Table 4). The study presented descriptions of events that characterized the following Allen (1983) intervals: *during, before, meets,* and *overlaps*.

Experiment 5's design helped to test a deflationary hypothesis for the biases revealed in the previous studies, namely that participants exhibit preferences, not for iconic descriptions, but for the word "started" over the word "ended." Such a preference may exist because people may use the verb *start* more frequently in speech—though a probe of the Corpus of Contemporary American English, COCA, revealed that the word "start" occurred 5.5 million times in the entire corpus and the word "end" occurred 6.2 million times, suggesting that people use the two words with roughly equal frequencies. Nevertheless, Experiment 5 presented a stronger test of the hypothesis: It presented option pairs that compared only the starts of events and not their ends, and so a preference for the word "started" cannot explain any remaining iconicity biases that participants exhibit.

The design was instrumental in testing a predicted interaction: Namely, that participants should prefer descriptions that contain the temporal adverbial "after" over those that contain "before" in sentence-initial form, for example, "After A started, B started," but this difference should reverse for sentence-final form, that is, they should prefer "A started before B started" over "B started after A started." Participants in Experiment 5 who exhibit iconicity biases in all conditions also exhibit this interaction as a consequence.

Table 4
The Two Conditions of Response Option Pairs in Experiment 5, Along With a Schematic of Each Option; Whether Each Option's Temporal Adverbial Is in Sentence-Initial or -Final Position; the Adverbial Itself; and Whether Each Option Is Iconic or Noniconic

Response option pairs	Temporal adverbial	Iconicity
Sentence-initial condition		
After A started, B started.	After	Iconic
Before B started, A started.	Before	Noniconic
Sentence-final condition		
B started after A started.	After	Noniconic
A started before B started.	Before	Iconic

Note. The theory predicts that participants should prefer iconic descriptions (bolded).

Method

Participants

Sixty participants completed the study on AMT for monetary compensation ($2.50). All of the participants were native English speakers, and all but 11 had taken one or fewer courses in introductory logic. Two participants were excluded from the analysis for producing nonsense responses in a postexperimental questionnaire. The analyses reported below are based on the remaining 58 participants (31 female, 25 male, one other, one prefer not to say; $M_{\text{age}} = 41.14$).

Task and Design

The task and design of Experiment 5 were similar to Experiment 2: Participants carried out 16 problems describing the durations of two events. These durations corresponded to time intervals in the pattern of four different Allen relations: *during*, *before*, *meets*, and *overlaps*, and participants received four problems of each type. Participants selected between two response options: one iconic and one noniconic. Half of the problems presented response options in which the temporal adverbial started the sentence while the balance of problems placed the adverbial in the middle of the sentence. For each pair of response options, one used the adverbial "before" and the other used the adverbial "after," though by design, the two conditions counterbalanced whether the "before" option corresponded to an iconic or noniconic response (see Table 4). Experiment 5 therefore reflected a fully repeated-measures design.

Design, Materials, and Procedure

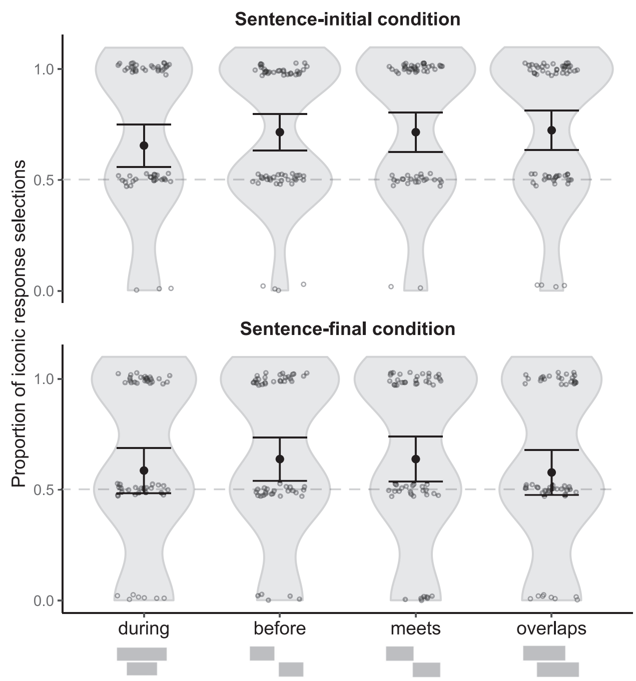
Same as Experiment 2.

Results and Discussion

Figure 5 shows the proportion of iconic response selections in Experiment 5 as a function of the four different types of problems in the study and whether response options were in sentence-initial or sentence-final form. Overall, participants produced iconic responses on 66% of responses. Data were subjected to a GLMM regression with a maximal random effects structure; the analysis revealed a significant intercept ($B = 0.57$, $SE = 0.18$, $p = .001$), as well as a significant predictor of sentence type ($B = 0.44$, $SE = 0.14$, $p = .002$). Forty-six of the 58 participants displayed this preference (binomial test: $p < .001$). Nonparametric analyses (provided at https://osf.io/369a8) corroborated the results of the regression. We use them to probe the significant difference between sentence-initial and sentence-final descriptions.

Figure 5

Violin Plots of the Proportions of Iconic Response Selections as a Function of the Four Separate Types of Problems in Experiment 5, and as a Function of Whether Participants Evaluated Response Options in Sentence-Initial (e.g., "After A Started, B Started") Construction or Sentence-Final Construction



Note. Light gray circles denote individual participants' mean proportions of iconic responses. Dark black circles denote mean proportions across all participants. The dashed gray lines denote chance performance. Error bars denote 95% confidence intervals.

Nonparametric Analyses

Across the four conditions, participants preferred iconic descriptions on 62%, 68%, 65%, and 68% of trials in *during*, *before*, *meets*, and *overlaps* conditions (Wilcoxon tests against chance, i.e., 50%; $zs > 3.82$, $ps < .001$, Cliff's δs > 0.36). They preferred iconic descriptions more often in sentence-initial construction than sentence-final construction, but this difference was unreliable (70% vs. 61%; Wilcoxon test, $z = 1.82$, $p = .07$, Cliff's $\delta = 0.19$). On each trial, participants received one response option that contained the temporal adverbial "before" and one option that contained "after." A consequence of a robust iconicity bias is that participants should prefer descriptions containing "after" a majority of the time in the sentence-initial condition and a minority of the time in the sentence-final condition; they exhibited such a pattern (70% vs. 40%; Wilcoxon test, $z = 5.8$, $p < .001$, Cliff's $\delta = 0.63$).

Experiment 5 generalized the iconicity bias in a number of ways: It showed that people exhibit the bias for two additional Allen relations (*meets* and *overlaps*), and it showed that they exhibit the bias for constructions of the form "After A started, B started," that is, sentence-initial descriptions. Indeed, the bias was stronger for these descriptions, and the study suggested that it overrides preferences for temporal adverbials (see Politzer-Ahles et al., 2017, who report a similar finding for nondurative events). The study also helped to rule out the possibility that people prefer the verb "start" over the verb "end" instead of considering the iconicity of the description; the two response options provided to participants in Experiment 5 only contained the verb "start," and yet the bias persisted across all conditions.

Nevertheless, Experiment 5 and the preceding studies are limited in at least three different ways. First, all the previous studies concerned only one domain, that is, the domain of computer network events, and no evidence generalizes the iconicity bias beyond those materials. Second, no evidence generalizes the bias beyond the words "start" and "end" to describe durational events. In everyday language, people use many words to describe durations, for example, you can describe the hours of operation of a particular store based on when the store "opens" and "closes." Preferences for iconic descriptions should generalize to such verbs. The third limitation is perhaps the most subtle: Each of the previous studies assigned numerical timepoints randomly to create appropriate time intervals. For instance, in Experiment 5, one trial may have described the *before* relation between two events using the following timepoints:

The download started at 10 am and ended at 2 pm.
(download: 4 hr long)
The encryption started at 6 pm and ended at 7 pm.
(encryption: 1 hr long)

While another trial may have assigned the timepoints as follows:

The download started at 10 am and ended at 2 pm.
(download: 4 hr long)
The encryption started at 4 pm and ended at 8 pm.
(encryption: 4 hr long)

The parentheticals provide the durations of the corresponding events. They show that on some trials, the two events had equal durations, and on others, the durations differed. Some participants may have carried out simple arithmetic to calculate the durations of the events. In Experiments 1–5, participants may have preferred shorter durations to longer ones, and these preferences may have translated into preferences for start times over end times. Indeed, anecdotal evidence suggests that reasoners can take duration into account when considering descriptions of events: They find events with equal durations easier to process than events with durations that differ (see Kelly et al., 2020). Experiment 6 accordingly (a) tested the effect of whether the durations between start times and end times are equal or not, (b) generalized the iconicity bias to descriptions containing the words "opened" and "closed," and (c) generalized the bias to a novel domain.

Experiment 6

Experiment 6 sought to generalize beyond previous studies, to eliminate confounds, and to test whether reasoners take into account the relative durations of events when they consider alternative descriptions, and so it implemented many changes. It presented participants with materials that concerned a music festival involving multiple stages with performances that could co-occur, and they were provided a graphic of two stages labeled by color names (see https://osf.io/utkq2). Participants read the following information:

> Music festivals often have multiple stages as can be seen in this example festival map. … Different stages will open and close throughout the day to keep the crowds moving through the festival and give the staff time to clean and set up for a future set of performances.

The study proper provided participants with descriptions of when two separate stages (labeled by color names) opened and closed for a particular set of performances, for example,

The red stage opened at 1 pm and closed at 4 pm on Friday.
The blue stage opened at 4 pm and closed at 7 pm on Friday.

The experiment presented descriptions whose intervals corresponded to the *before*, *meets*, and *overlaps* relations in Allen's (1983) calculus. It also varied whether the two intervals had the same duration, as in the example above, or whether they had different durations.

Method

Participants

Fifty-one participants completed the experiment for monetary compensation ($2.50) through AMT, commensurate with minimum-wage standards. One participant was excluded from the analysis for always selecting the first response option on all trials. Five additional participants were excluded as suspected bots. The analyses reported below are based on the remaining 45 participants (23 female, $M_{age} = 37.2$). All participants were native English speakers, and 29 participants had taken one or fewer courses in introductory logic.

Design, Materials, and Procedure

Participants selected whether they preferred an iconic or a noniconic description to best characterize a pair of premises that stipulated when two different stages opened and closed. The stages were named based on color (e.g., "the orange stage"). The experiment varied the Allen relation to which the timepoints in the premises corresponded, that is, the premises could describe a *before*, *meets,* or *overlaps* relation (see the introductory part). It also varied whether the two events had the same duration or whether they had different durations, and so the experiment implemented a 3 × 2 within-participant design. Participants carried out three problems in each of the six conditions, that is, 18 problems in total. As in previous studies, the order of the 18 trials, and the order in which the two response options appeared on the screen, were randomized.

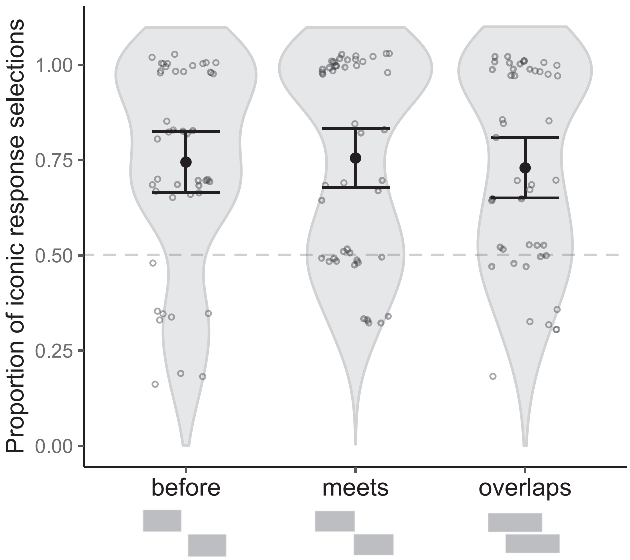
Results and Discussion

Participants selected iconic responses 74% of the time in Experiment 6. When the data were subjected to a GLMM regression with maximal random-effects structure, it revealed a significant intercept ($B = 1.54$, $SE = 0.33$, $p < .001$), that is, a significant bias in favor of iconic responses. No other predictors were significant, including the manipulation of whether the two events in the premises had same or different durations ($B = 1.14$, $SE = 18$, $p = .45$). Thirty-seven of the 45 participants exhibited a preference for iconic descriptions (binomial test: $p < .001$), and nonparametric analyses (see https://osf.io/akf4j) corroborated the reliability of the effect. Figure 6 shows the proportion of iconic response selections in Experiment 6.

In sum, Experiments 1 through 6 revealed a systematic bias toward iconic descriptions of durations, that is, descriptions that interrelate events that occur earlier in a temporal model of events. The experiments surveyed all seven possible relations between two durations. In rare cases, it found preferences for noniconic descriptions (e.g., in the *starts* condition of Experiment 2). In some cases, participants exhibited no bias or else a slight but statistically unreliable bias (cf., the *starts* condition in Experiment 1; the *equals* and *starts* conditions in Experiment 3). Experiments 4, 5, and 6 generalized the bias across a variety of Allen relations, response conditions, and materials, and they ruled out the hypothesis that people prefer descriptions using the word "started": Experiment 4 used only descriptions of when events start, and Experiment 6 made use of the words "opened" and "closed" to describe when events started and ended. The latter study also controlled for whether

Figure 6

Violin Plots of the Proportions of Iconic Response Selections as a Function of the Three Separate Types of Problems in Experiment 6, Each of Which Corresponded to a Relation in Allen's (1983) Interval Algebra



Note. Light gray circles denote individual participants' mean proportions of iconic responses. Dark black circles denote mean proportions across all participants. The dashed gray lines denote chance performance. Error bars denote 95% confidence intervals.

the duration of events diminished the bias. The result validates the model theory of durational reasoning (Kelly et al., 2020; Khemlani et al., 2015) and its central tenet that reasoners construct and scan iconic mental simulations of durational events.

General Discussion

This paper sought to test whether people comprehend durational language by building iconic mental simulations, that is, models. In many cases, mental simulations can bias the way in which concepts interrelate. For example, Zwaan and Yaxley (2003) show that individuals are slower to react to pairs of words such as *attic* and *basement* when they are spatially arranged as follows:

basement
attic

than when they are arranged in an order congruent with an appropriate spatial mental model of the scenario:

attic
basement.

Many theorists argue that reasoners can represent time along a spatial axis (Bonato et al., 2012; Hoerl & McCormack, 2019) and likewise that people are faster to process descriptions of events in chronological order (Münte et al., 1998; Ye et al., 2012). Our studies are the first to link these lines of exploration with the study of how individuals consciously understand and reason about durations.

Six experiments showed that people have explicit preferences for some durational descriptions over others, a preference that is best explained as a bias derived from how people build and scan mental models of durations. Participants in each experiment reliably selected descriptions that reflect iconic scanning of durational models (see Figure 1), namely descriptions that relate the start points of two events over descriptions about their end points.

Experiment 1 set up a competition to contrast the two accurate but incomplete descriptions of events, but it allowed participants to optionally indicate that that neither description was better than the other. Nevertheless, more often than not, participants chose the option that described start points more often than the one that described end points. Experiment 2 was designed to create maximal competition between two ambiguous relations between durational events because participants were forced to choose between descriptions such as:

The configuration started before the encryption started.
The encryption ended when the configuration ended.

The results do not appear to be an artifact of processing sentences because participants exhibited similar results when they considered diagrams (Experiment 3), and because the usage of terms such as "before" and "after" did not attenuate the bias (Experiments 4 and 5). The bias generalizes to multiple domains and multiple ways of describing intervals between events, such as the usage of words such as "opened" and "closed" (Experiment 6).

The results cannot be explained by logical calculi that deal with temporal information because such systems do not distinguish between two valid temporal deductions. Participants' patterns of response are consistent with the general hypothesis that people build a mental timeline when reasoning about the temporal relations of events (see Bonato et al., 2012), but studies corroborating a mental timeline seldom concern how people engage in conscious reasoning tasks. The present results are predicted directly by the more specific hypothesis that reasoners construct and scan mental models of events arranged in such a timeline when they reason about durational relations (Kelly et al., 2020; Khemlani et al., 2015).

The results likewise cannot be explained by the frequency of response options, for at least three reasons. First, words such as "started" and "ended," that is, those used in most of the response options provided to participants, occur with equal frequency in natural language. Second, people exhibit iconicity biases even when provided with response options that describe only the starts of events. Third, people exhibit such biases for words such as "opened" and "closed."

Across all six experiments, people exhibited iconicity biases for every Allen relation tested except for the *starts* relation (see, e.g., Experiments 1 and 2; but cf., Experiment 3). Here is an example of a description that yields a *starts* relation:

The backup started at 3 pm and ended at 7 pm.
The cyberattack started at 3 pm and ended at 11 pm.

For such scenarios, participants preferred descriptions relating the ends of the two events, such as *the backup ended before the cyberattack ended*. One explanation for this reversal may be because people prefer timepoint asynchrony, that is, they preferred any description that references two different timepoints (e.g., when the *backup* and the *cyberattack* ended) over any description that references only one timepoint (e.g., when the two events started). However, such a preference cannot explain why people exhibited iconicity biases in the case of *equals* relation—in which both timepoints are synchronous—and in the case of *during* relations—in which both timepoints are asynchronous. Another explanation for the reversal may be to amend the proposal that reasoners scan models in order from earlier to later events: When such a scanning procedure discovers that two events started at the same time as one another, it may reverse the direction of the scan. Such an amendment may yield additional testable empirical predictions of reasoners' preferences for temporal conclusions, but it too has difficulty explaining why people exhibited an iconicity bias for *equals* relations. In any case, the behavior hints that reasoners are driven by more than iconicity to select competing descriptions. Perhaps the best evidence for noniconic processing is the fact that iconic biases did not yield uniform preferences: Roughly 40% of the time across all the studies, participants preferred noniconic descriptions. Indeed, many reasoners may process temporal mental models in flexible and idiosyncratic ways. Previous research shows that people are capable, but slower, at processing events in reverse chronological order (see, e.g., Claus & Kelter, 2006; Politzer-Ahles et al., 2017) and that their neural processing indexes such difficulty (Münte et al., 1998). The current studies shows that reverse chronologies do not just affect nonverbal responses; it also affects conscious processing and preferences for linguistic descriptions of events. Future research may investigate the spontaneous strategies reasoners use to make inferences about time, and how reasoners flexibly think about reverse chronological orders (see Schaeken & Johnson-Laird, 2000).

Descriptions of durational relations can be ambiguous. The phrase *the rain happened during the concert* may imply that the rain started after the concert began. Additional verbs such as *start* and *end* can help clarify the relations between the two events, for example, *the rain started after the concert began and ended before the concert wrapped up*. Partial descriptions can be consistent with several different relations between events. The present studies showed that people prefer some ambiguous relations over others and that their preferences are systematic. The results support the proposal that humans simulate events with durations along an iconic mental representation and that they consciously scan that simulation to reason about time.

References

Agresti, A. (2002). Logistic regression. In A. Agresti (Ed.), *Categorical data analysis* (2nd ed.). Wiley Online Library. https://doi.org/10.1002/0471249688.ch5

Allen, J. F. (1983). Maintaining knowledge about temporal intervals. *Communications of the ACM*, *26*(11), 832–843. https://doi.org/10.1145/182.358434

Amidon, A. (1976). Children's understanding of sentences with contingent relations: Why are temporal and conditional connectives so difficult?. *Journal of Experimental Child Psychology*, *22*(3), 423–437.

Amidon, A., & Carey, P. (1972). Why five-year-olds cannot understand before and after. *Journal of Verbal Learning and Verbal Behavior*, *11*(4), 417–423. https://doi.org/10.1016/S0022-5371(72)80022-7

Anderson, A. H. (1982). *Text comprehension: The influence of temporal information on processing and reading rate*. University of Glasgow.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bergen, B., Lau, C., & Ting, T. (2012). Writing direction affects how people map space onto time. *Frontiers in Psychology*, *3*, Article 109. https://doi.org/10.3389/fpsyg.2012.00109

Bonato, M., Saj, A., & Vuilleumier, P. (2016). Hemispatial neglect shows that "before" is "left". *Neural Plasticity*, *2016*, Article 2716036. https://doi.org/10.1155/2016/2716036

Bonato, M., Zorzi, M., & Umiltà, C. (2012). When time is space: Evidence for a mental time line. *Neuroscience and Biobehavioral Reviews*, *36*(10), 2257–2273. https://doi.org/10.1016/j.neubiorev.2012.08.007

Brannon, E. M., Suanda, S., & Libertus, K. (2007). Temporal discrimination increases in precision over development and parallels the development of numerosity discrimination. *Developmental Science*, *10*(6), 770–777. https://doi.org/10.1111/j.1467-7687.2007.00635.x

Byrne, R. (2005). *The rational imagination. How people create alternatives to rality*. MIT Press.

Casasanto, D. (2016). A shared mechanism of linguistic, cultural, and bodily relativity. *Language Learning*, *66*(3), 714–730. https://doi.org/10.1111/lang.12192

Casasanto, D., & Jasmin, K. (2012). The hands of time: Temporal gestures in English speakers. *Cognitive Linguistics*, *23*(4), 643–674. https://doi.org/10.1515/cog-2012-0020

Chatterjee, A., Southwood, M. H., & Basilico, D. (1999). Verbs, events and spatial representations. *Neuropsychologia*, *37*(4), 395–402. https://doi.org/10.1016/S0028-3932(98)00108-0

Clark, E. V. (1971). On the acquisition of the meaning of before and after. *Journal of Verbal Learning and Verbal Behavior*, *10*(3), 266–275. https://doi.org/10.1016/S0022-5371(71)80054-3

Claus, B., & Kelter, S. (2006). Comprehending narratives containing flashbacks: Evidence for temporally organized representations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(5), 1031–1044. https://doi.org/10.1037/0278-7393.32.5.1031

Cooperrider, K., & Núñez, R. (2009). Across time, across the body: Transversal temporal gestures. *Gesture*, *9*(2), 181–206. https://doi.org/10.1075/gest.9.2.02coo

Diessel, H. (2008). Iconicity of sequence: A corpus-based analysis of the positioning of temporal adverbial clauses in English. *Cognitive Linguistics*, *19*(3), 465–490. https://doi.org/10.1515/COGL.2008.018

Droit-Volet, S., & Rattat, A. C. (1999). Are time and action dissociated in young children's time estimation?. *Cognitive Development*, *14*(4), 573–595. https://doi.org/10.1016/S0885-2014(99)00020-9

Ezzyat, Y., & Davachi, L. (2021). Neural evidence for representational persistence within events. *The Journal of Neuroscience*, *41*(37), 7909–7920. https://doi.org/10.1523/JNEUROSCI.0073-21.2021

Fischer, M., Gabbay, D., & Vila, L. (2004). *Handbook of temporal reasoning in artificial intelligence*. Elsevier.

Freksa, C. (1992). Temporal reasoning based on semi-intervals. *Artificial Intelligence*, *54*(1–2), 199–227. https://doi.org/10.1016/0004-3702(92)90090-K

Fuhrman, O., & Boroditsky, L. (2010). Cross-cultural differences in mental representations of time: Evidence from an implicit nonlinguistic task. *Cognitive Science*, *34*(8), 1430–1451. https://doi.org/10.1111/j.1551-6709.2010.01105.x

Gantt, H. L. (1910). *Work, wages and profit*. Hive Publishing.

Gevers, W., Reynvoet, B., & Fias, W. (2003). The mental representation of ordinal sequences is spatially organized. *Cognition*, *87*(3), B87–B95. https://doi.org/10.1016/S0010-0277(02)00234-2

Gevers, W., Reynvoet, B., & Fias, W. (2004). The mental representation of ordinal sequences is spatially organised: Evidence from days of the week. *Cortex*, *40*(1), 171–172. https://doi.org/10.1016/S0010-9452(08)70938-9

Giacomo, G., & Vardi, M. Y. (2013). *Linear temporal logic and linear dynamic logic on finite traces* [Conference session]. Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence, Beijing, China. https://dl.acm.org/doi/proceedings/10.5555/2540128

Gibbon, J. (1977). Scalar expectancy theory and Weber's law in animal timing. *Psychological Review*, *84*(3), 279–325. https://doi.org/10.1037/0033-295X.84.3.279

Goldvarg, E., & Johnson-Laird, P. N. (2001). Naive causality: A mental model theory of causal meaning and reasoning. *Cognitive Science*, *25*(4), 565–610. https://doi.org/10.1207/s15516709cog2504_3

Goodwin, G. P., & Johnson-Laird, P. N. (2005). Reasoning about relations. *Psychological Review*, *112*(2), 468–493. https://doi.org/10.1037/0033-295X.112.2.468

Goranko, V., Montanari, A., & Sciavicco, G. (2004). A road map of interval temporal logics and duration calculi. *Journal of Applied Non-Classical Logics*, *14*(1–2), 9–54. https://doi.org/10.3166/jancl.14.9-54

Hoeks, J. C., Stowe, L. A., & Wunderink, C. (2004). Time is of the essence: Processing temporal connectives during reading. In K. D. Forbus, D. Gentner, & T. Regier (Eds.), *Proceedings of the Twenty-Sixth Annual Conference of the Cognitive Science Society* (pp. 577–582). Erlbaum.

Hoerl, C., & McCormack, T. (2019). Thinking in and about time: A dual systems perspective on temporal cognition. *Behavioral and Brain Sciences*, *42*, Article e244.

IBM Corp. (2021). *IBM design language: Color*. https://www.ibm.com/design/language/color

Isham, E. A., Banks, W. P., Ekstrom, A. D., & Stern, J. A. (2011). Deceived and distorted: Game outcome retrospectively determines the reported time of action. *Journal of Experimental Psychology: Human Perception and Performance*, *37*(5), 1458–1469. https://doi.org/10.1037/a0023111

Ishihara, M., Keller, P. E., Rossetti, Y., & Prinz, W. (2008). Horizontal spatial representations of time: Evidence for the STEARC effect. *Cortex*, *44*(4), 454–461. https://doi.org/10.1016/j.cortex.2007.08.010

Jahn, G., Knauff, M., & Johnson-Laird, P. N. (2007). Preferred mental models in reasoning about spatial relations. *Memory and Cognition*, *35*, 2075–2087. https://doi.org/10.3758/BF03192939

Jeffrey, R. (1981). *Formal logic: Its scope and limits* (2nd ed.). McGraw-Hill.

Johnson-Laird, P. N. (2006). *How we reason*. Oxford University Press.

Kelly, L. J., Khemlani, S., & Johnson-Laird, P. N. (2020). Reasoning about durations. *Journal of Cognitive Neuroscience*, *32*(11), 2103–2116. https://doi.org/10.1162/jocn_a_01621

Khemlani, S., & Johnson-Laird, P. N. (2022). Reasoning about properties: A computational theory. *Psychological Review*, *129*(2), 289–312. https://doi.org/10.1037/rev0000240

Khemlani, S. S. (2022). *A computational theory of temporal reasoning* [Conference session]. International conference on cognitive modeling 2022.

Khemlani, S. S., Bello, P., Briggs, G., Harner, H., & Wasylyshyn, C. (2021). Much ado about nothing: The mental representation of omissive relations. *Frontiers in Psychology*, *11*, Article 3854. https://doi.org/10.3389/fpsyg.2020.609658

Khemlani, S. S., Harrison, A. M., & Trafton, J. G. (2015). Episodes, events, and models. *Frontiers in Human Neuroscience*, *9*, Article 590. https://doi.org/10.3389/fnhum.2015.00590

Khemlani, S. S., Mackiewicz, R., Bucciarelli, M., & Johnson-Laird, P. N. (2013). Kinematic mental simulations in abduction and deduction. *Proceedings of the National Academy of Sciences*, *110*(42), 16766–16771. https://doi.org/10.1073/pnas.1316275110

Knauff, M. (1999). The cognitive adequacy of Allen's interval calculus for qualitative spatial relation and reasoning. *Spatial Cognition and Computation*, *1*(3), 261–290. https://doi.org/10.1023/A:1010097601575

Knauff, M. (2013). *Space to reason: A spatial theory of human thought*. MIT Press.

Knauff, M., Strube, G., Jola, C., Rauh, R., & Schlieder, C. (2004). The psychological validity of qualitative spatial reasoning in one dimension. *Spatial Cognition and Computation*, *4*(2), 167–188. https://doi.org/10.1207/s15427633scc0402_3

Kowalski, R., & Sergot, M. (1989). A logic-based calculus of events. In *Foundations of knowledge base management* (pp. 23–55). Springer.

Leone, M. J., Salles, A., Pulver, A., Golombek, D. A., & Sigman, M. (2018). Time drawings: Spatial representation of temporal concepts. *Consciousness and Cognition*, *59*, 10–25. https://doi.org/10.1016/j.concog.2018.01.005

Levin, I. (1977). The development of time concepts in young children: Reasoning about duration. *Child Development*, *48*(2), 435–444. https://doi.org/10.2307/1128636.

Maass, A., & Russo, A. (2003). Directional bias in the mental representation of spatial events: Nature or culture? *Psychological Science*, *14*(4), 296–301. https://doi.org/10.1111/1467-9280.14421

Manna, Z., & Pnueli, A. (1992). *The temporal logic of reactive and concurrent systems*. Springer.

Manna, Z., & Pnueli, A. (1995). *The temporal logic of reactive and concurrent systems: Safety*. Springer-Verlag.

Matlock, T. (2004). Fictive motion as cognitive simulation. *Memory and Cognition*, *32*(8), 1389–1400. https://doi.org/10.3758/BF03206329

Matlock, T., Ramscar, M., & Boroditsky, L. (2005). On the experiential link between spatial and temporal language. *Cognitive Science*, *29*(4), 655–664. https://doi.org/10.1207/s15516709cog0000_17

Meck, W. H., & Church, R. M. (1983). A mode control model of counting and timing processes. *Journal of Experimental Psychology: Animal Behavior Processes*, *9*(3), 320–334. https://doi.org/10.1037/0097-7403.9.3.320

Miller, G. A., & Johnson-Laird, P. N. (1976). *Language and perception*. Belknap Press.

Münte, T., Schiltz, K., & Kutas, M. (1998). When temporal terms belie conceptual order. *Nature*, *395*(6697), 71–73. https://doi.org/10.1038/25731

Nieuwland, M. S. (2015). The truth before and after: Brain potentials reveal automatic activation of event knowledge during sentence comprehension. *Journal of Cognitive Neuroscience*, *27*(11), 2215–2228. https://doi.org/10.1162/jocn_a_00856

Odic, D. (2018). Children's intuitive sense of number develops independently of their perception of area, density, length, and time. *Developmental Science*, *21*(2), e12533. https://doi.org/10.1111/desc.12533

Ornstein, R. E. (1969). *On the experience of time*. Penguin.

Peirce, C. S. (1931–1958). Collected papers of Charles Sanders Peirce. In C. Hartshorne, P. Weiss, & A. Burks (Eds.), Harvard University Press.

Pnueli, A. (1977). *The temporal logic of programs* [Conference session]. Proceedings of the 18th Annual Symposium on Foundations of Computer Science.

Politzer-Ahles, S., Xiang, M., Almeida, D., & Filik, R. (2017). "Before" and "after": Investigating the relationship between temporal connectives and chronological ordering using event-related potentials. *PLoS One*, *12*(4), e0175199. https://doi.org/10.1371/journal.pone.0175199

Provasi, J., Rattat, A. C., & Droit-Volet, S. (2011). Temporal bisection in 4-month-old infants. *Journal of Experimental Psychology: Animal Behavior Processes*, *37*(1), 108–113. https://doi.org/10.1037/a0019976

Pyykkönen, P., & Järvikivi, J. (2012). Children and situation models of multiple events. *Developmental Psychology*, *48*(2), 521–529. https://doi.org/10.1037/a0025526

Pyykkönen, P., Niemi, J., & Järvikivi, J. (2003). Sentence structure, temporal order and linearity: Slow emergence of adult-like syntactic performance in Finnish. *SKY Journal of Linguistic*, *16*, 113–138.

Ragni, M., & Knauff, M. (2013). A theory and a computational model of spatial reasoning with preferred mental models. *Psychological Review*, *120*(3), 561–588. https://doi.org/10.1037/a0032460

Rauh, R., Hagen, C., Knauff, M., Kuss, T., Schlieder, C., & Strube, G. (2005). Preferred and alternative mental models in spatial reasoning. *Spatial Cognition and Computation*, *5*(2-3), 239–269. https://doi.org/10.1207/s15427633scc052&3_7

Rozier, K. Y. (2011). Linear temporal logic symbolic model checking. *Computer Science Review*, *5*(2), 163–203. https://doi.org/10.1016/j.cosrev.2010.06.002

Santiago, J., Lupáñez, J., Pérez, E., & Funes, M. J. (2007). Time (also) flies from left to right. *Psychonomic Bulletin and Review*, *14*(3), 512–516. https://doi.org/10.3758/BF03194099

Santiago, J., Román, A., Ouellet, M., Rodríguez, N., & Pérez-Azor, P. (2010). In hindsight, life flows from left to right. *Psychological Research Psychologische Forschung*, *74*(1), 59–70. https://doi.org/10.1007/s00426-008-0220-0

Schaeken, W., & Johnson-Laird, P. N. (2000). Strategies in temporal reasoning. *Thinking and Reasoning*, *6*(3), 193–219. https://doi.org/10.1080/13546780050114500

Schaeken, W., Johnson-Laird, P. N., & d'Ydewalle, G. (1996). Mental models and temporal reasoning. *Cognition*, *60*(3), 205–234. https://doi.org/10.1016/0010-0277(96)00708-1

Tillman, K. A., & Barner, D. (2015). Learning the language of time: Children's acquisition of duration words. *Cognitive Psychology*, *78*, 57–77. https://doi.org/10.1016/j.cogpsych.2015.03.001

Tillman, K. A., Tulagan, N., Fukuda, E., & Barner, D. (2018). The mental timeline is gradually constructed in childhood. *Developmental Science*, *21*(6), Article e12679. https://doi.org/10.1111/desc.12679

Torralbo, A., Santiago, J., & Lupiáñez, J. (2006). Flexible conceptual projection of time onto spatial frames of reference. *Cognitive Science*, *30*(4), 745–757. https://doi.org/10.1207/s15516709cog0000_67

Tversky, B. (1989). Parts, partonomies, and taxonomies. *Developmental Psychology*, *25*(6), 983–995. https://doi.org/10.1037/0012-1649.25.6.983

Tversky, B., & Hemenway, K. (1984). Objects, parts, and categories. *Journal of Experimental Psychology: General*, *113*(2), 169–193. https://doi.org/10.1037/0096-3445.113.2.169

Vallesi, A., Weisblatt, Y., Semenza, C., & Shaki, S. (2014). Cultural modulations of space–time compatibility effects. *Psychonomic Bulletin and Review*, *21*(3), 666–669. https://doi.org/10.3758/s13423-013-0540-y

vanMarle, K., & Wynn, K. (2006). Six-month-old infants use analog magnitudes to represent duration. *Developmental Science*, *9*(5), F41–F49. https://doi.org/10.1111/j.1467-7687.2006.00508.x

Vicario, C. M., Caltagirone, C., & Oliveri, M. (2007). Optokinetic stimulation affects temporal estimation in healthy humans. *Brain and Cognition*, *64*(1), 68–73. https://doi.org/10.1016/j.bandc.2006.12.002

von Hecker, U., Klauer, K. C., & Abfalg, A. (2019). A robust anchoring effect in linear ordering. *Quarterly Journal of Experimental Psychology*, *72*(11), 2680–2689. https://doi.org/10.1177/1747021819855234

von Sobbe, L. (2022). *Instantiations of the cognitive grounding of linguistically expressed temporal information* [Doctoral dissertation]. University of Tübingen.

von Sobbe, L., Maienborn, C., Reiber, F., Scheifele, E., & Ulrich, R. (2021). Speed or duration? Effects of implicit stimulus attributes on perceived duration. *Journal of Cognitive Psychology*, *33*(8), 877–898. https://doi.org/10.1080/20445911.2021.1950736

Wang, Y., & Gennari, S. P. (2019). How language and event recall can shape memory for time. *Cognitive Psychology*, *108*, 1–21. https://doi.org/10.1016/j.cogpsych.2018.10.003

Weger, U. W., & Pratt, J. (2008). Time flies like an arrow: Space–time compatibility effects suggest the use of a mental timeline. *Psychonomic Bulletin and Review*, *15*(2), 426–430. https://doi.org/10.3758/PBR.15.2.426

Winter, B., Marghetis, T., & Matlock, T. (2015). Of magnitudes and metaphors: Explaining cognitive interactions between space, time, and number. *Cortex*, *64*, 209–224. https://doi.org/10.1016/j.cortex.2014.10.015

Xu, X., & Kwok, S. C. (2019). Temporal-order iconicity bias in narrative event understanding and memory. *Memory*, *27*(8), 1079–1090. https://doi.org/10.1080/09658211.2019.1622734

Ye, Z., Kutas, M., St. George, M., Sereno, M., Ling, F., & Münte, T. (2012). Rearranging the world: Neural network supporting the processing of temporal connectives. *NeuroImage*, *59*(4), 3662–3667. https://doi.org/10.1016/j.neuroimage.2011.11.039

Zacks, J., & Tversky, B. (2001). Event structure in perception and conception. *Psychological Bulletin*, *127*(1), 3–21. https://doi.org/10.1037/0033-2909.127.1.3

Zheng, J., Schjetnan, A., Yebra, M., Gomes, B., Clayton, M., Mosher, C. P., Kalia, S. K., Valiante, T. A., Mamelak, A. N., Kreiman, G., & Rutishauser, U. (2022). Neurons detect cognitive boundaries to structure episodic memories in humans. *Nature Neuroscience*, *25*(3), 358–368. https://doi.org/10.1038/s41593-022-01020-w

Zwaan, R. A. (1996). Processing narrative time shifts. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(5), 1196–1207. https://doi.org/10.1037/0278-7393.22.5.1196

Zwaan, R. A., Madden, C. J., & Stanfield, R. A. (2001). Time in narrative comprehension. In D. H. Schram & G. J. Steen (Eds.), *Psychology and sociology of literature* (pp. 71–86). John Benjamins.

Zwaan, R. A., & Yaxley, R. H. (2003). Spatial iconicity affects semantic relatedness judgments. *Psychonomic Bulletin and Review*, *10*(4), 954–958. https://doi.org/10.3758/BF03196557

(*Appendix follows*)

Appendix

Table A1 presents the names of events used in the descriptions of problems in Experiments 1–4. The events were selected such that any pair of them can concern independent and unrelated network processes.

Table A1
Names of Events

Backup	Rendering
Buffering	Security check
Compilation	Server update
Compression	Signal
Cyber attack	Software update
Defragmentation	Streaming
Download	Sync
Encryption	Testing
Installation	Transmission
Maintenance	Transfer
Recalibration	Uninstallation
Reconfiguration	Upload

Received December 11, 2021
Revision received January 9, 2023
Accepted May 12, 2023 ■