Directional biases in durative inference

Laura Kelly and Sangeet Khemlani
{laura.kelly.ctr, sangeet.khemlani}@nrl.navy.mil
Navy Center for Applied Research in Artificial Intelligence
US Naval Research Laboratory, Washington, DC 20375 USA

Abstract

Descriptions of durational relations can be ambiguous, e.g., the description 'two different meetings happened at the same time' could mean that one meeting started before the other ended, or it could mean that the meetings both started and ended simultaneously. A recent theory posits that people mentally simulate events with durations by representing the starts and ends of events along a chronological axis (Khemlani et al., 2015). To draw conclusions from this durational mental model, reasoners consciously scan it in the direction of earlier time points to later time points. The account predicts that people should prefer descriptions that are congruent with a chronological scanning procedure, e.g., descriptions that mention the starts of events before the ends of events. Two experiments corroborate the prediction, and show that chronological biases in temporal reasoning manifest in cases when reasoners consciously evaluate the durations of events.

Keywords: events, temporal reasoning, durational relations, mental models, mental timeline

Introduction

People exhibit directional biases when thinking about time. Many studies have examined how people reason about events organized by relations such as *before* and *after*. In general, they find that people appear to construct a "mental timeline" that is organized on a horizontal, left-to-right axis (for reviews, see Hoerl & McCormack, 2019 and Bonato, Zorzi, & Umiltà, 2012). Evidence in support of such a timeline comes from the way people spatially arrange items representing temporal events and concepts (e.g., Bergen, Lau & Ting, 2012; Fuhrman & Boroditsky, 2010; Leone, et al., 2018; Maass & Russo, 2003), as well as their implicit differences in reaction times on temporal judgment tasks (Gevers, Reynvoet, & Fias, 2003, 2004; Ishihara et al., 2008; Santiago, et al., 2007, 2010; Torralbo et al., 2006; Vallesi et al., 2014; Weger & Pratt, 2008). Likewise, people's spontaneous temporal gestures can reflect a horizontal axis that represents time (Casasanto & Jasmin, 2012; Cooperrider & Nunez, 2009). Some researchers even make the stronger claim that chronological biases in temporal thinking come from the organization of neural structures, such that a mental timeline is an emergent result of low-level attentional biases (Chatterjee, Southwood, & Basilico, 1999; Vicaro et al., 2007) or cross-domain representational mechanisms (see Winter, Marghetis, & Matlock, 2015, for a review).

Most research on temporal reasoning has focused on punctate events – events whose durations are irrelevant to understanding their temporal relations to other events. For instance, if you know that *the dinner occurred before the movie*, it doesn't matter if the dinner was hurriedly eaten within 15 minutes 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 until later in development (Tillman & Barner, 2015; Tillman, et al., 2018). Even for adults, descriptions of durations can be ambiguous. For instance, if you're told that *two different meetings happened at the same time*, it 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.

Adult reasoners can make simple durational inferences without any special training in logic. Events are often naturally treated as having parts and subevents, i.e., they can be organized into partonomic hierarchies (Miller & Johnson-Laird, 1976; Tversky, 1989; Tversky & Hemenway, 1984; Zacks & Tversky, 2001). For instance, suppose you know the following:

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

It is easy to make the following temporal inferences:

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

The first inference concerns a relation, *before*, between two different events, the *harvest* and the *winter*. This inference does not reflect the durational nature of the events: in (2a), the events could be treated as punctate. The remaining inferences (2b-e) are durational because they concern relations that reference parts of events, namely, when an event starts and when an event ends. To reference a part of an event is to imply there are other parts – that the event is extended in time.

To represent an event's duration, a reasoning system must, at a minimum, represent its start and its end. But few theories have proposed what people represent when they reason about durational relations such as *during*, and few studies have examined systematic patterns of durational reasoning. The claims of research into mental timelines with punctate events appear to imply that people should also represent mental timelines for durational relations.

In the following, 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 more recent computational cognitive theory that proposes that people build mental models from descriptions of durations (Khemlani, Harrison, & Trafton, 2015). The theory extends the proposal that people construct a mental timeline to represent events. And we show why a central prediction of the theory – i.e., that inferences emerge from the way people scan models – yields a conscious chronological bias in preferring some inferences to others. We describe two 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 be simple and they can be complex. Consider this deduction:

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

The final conclusion in (3) is *valid*, i.e., it is true in any situation in which the premises are true as well (Jeffery, 1981, p. 1). Now consider this one:

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

The final conclusion in (4) is invalid – it is possibly true, but not necessarily the case. Systems of logic were developed to provide a formal way to derive logically valid temporal deductions (e.g., Allen, 1983; Fischer, Gabbay, & Vila, 2005; Freksa, 1992; Goranko, Montanari, & Sciavicco, 2004; Kowalski & Sergot, 1989). Temporal logics provide a basis for formalizing interval relations, i.e., the way in which one interval exists relative to another. As a result, temporal logics make use of relations that have no connection to concepts in natural language. 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:

```
A —               A before B.     A ———        A equals B.
B        —                        B ———

A ——              A meets B.      A —           A starts B.
B     ——                          B —————

A     —           A during B.     A          — A finishes B.
B —————                           B —————

A ———          A overlaps B.
B     ———
```

The horizontal lines represent the way an event endures across multiple points in time. Allen's algebra specifies 13 relations, i.e., the 7 relations above along with their inverses (the inverse of *before* is *after*), but we omit them for brevity. Some of the relations have intuitive mappings onto temporal prepositions and connectives in natural language, e.g., *Event A occurred before event B*. Other relations can only be expressed using combinations of durational relations, e.g., 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 *ended* and *started*, 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).

A recent theory sought to explain inferences like (3) and (4). Khemlani, Harrison, and Trafton (2015) 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, Knauff, & Johnson-Laird, 2007; Ragni & Knauff, 2013), causality (Goldvarg & Johnson-Laird, 2001; Khemlani, Barbey, & Johnson-Laird, 2015; Khemlani, Byrne, & Johnson-Laird, 2018), and punctate events (Schaeken, Johnson-Laird, & d'Ydewalle, 1996; Schaeken & Johnson-Laird, 2000). The theory rests on three fundamental constraints: first, models are *iconic*, i.e., their structure maps onto the structure of the things they represent (see Peirce, 1931-1958, Vol. 4). An iconic treatment of events, for instance, suggests that reasoners represent them as tokens arranged along a mental time line (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). To represent durations iconically, Khemlani et al. (2015) proposed that people use discrete tokens to represent the starts and ends of events, but that they do not maintain or reason with intermediate time points. Hence, a mental model of (3) can be represented in the following diagram:

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

The diagram depicts four durational events whose durations are denoted by the starts and the ends of events, i.e., the opening and closing brackets. The words in the tokens are merely used to label the events, i.e., they specify the event's content but not its structure. An advantage of the theory is that the representation is agnostic to the length of the duration, and so it can be applied uniformly to events lasting minutes and events lasting years, something that adult reasoners do with equal ease.

Second, reasoners draw inferences by scanning models in a particular direction, e.g., they can scan the model above from the start of the fall to the end of the winter. By default, they scan models in chronological order (from earliest to latest), but they can also scan models in the opposite order when a particular inferential task demands it. Indeed, they can spontaneously form strategies for reasoning about time (Schaeken & Johnson-Laird, 2000), and it may be that they can scan models in a flexible fashion.

Finally, the model theory uniquely explains why some durational reasoning problems are easier than others: problems that demand more models should be more difficult (Schaeken & Johnson-Laird, 1996). For instance, it should be harder to reason about the premises in (4) than in (3), because (4) requires reasoners to keep track of multiple models whereas reasoners can represent (3) using only one model. Recent tests of this hypothesis confirmed the difference in difficulty between one- and multiple-model problems (Kelly & Khemlani, 2019; Kelly, Khemlani, & Johnson-Laird, under review).

The model theory makes a surprising prediction about how people evaluate durations. Consider again the introductory example:

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

Ignoring the specific points in time that are used to construct the models, i.e., the months, a model of (1) would be:

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

We invite the reader to consider which of the following two statements is a better description for (1):

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

In general, (5a) and (5b) are ambiguous in the same way: they're incomplete and compatible with multiple sorts of models, and so neither statement serves as a "better" description than the other – logical frameworks such as the Allen algebra or the event calculus would treat both statements as ambiguous. However, if reasoners exhibit a chronological bias in scanning models – as the model theory predicts – then they should consider (5a) to be better than (5b). We report two experiments that test and validate this prediction.

Experiment 1

To investigate whether participants exhibit a chronological scanning bias, i.e., whether they prefer descriptions relating start times or descriptions relating end times, Experiment 1 presented participants with a description of the durations of two events, e.g.,

The encryption started at 12pm and ended at 11pm…
The download started at 9am and ended at 11pm…

Participants then selected which of two different statements better summarized how the events related to one another:

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

The model theory predicts that reasoners should build a durational model of the events and then scan it from earlier times to later times – hence, they should exhibit a preference for descriptions of start times over descriptions of end times.

Method

Participants. 50 participants completed the experiment for monetary compensation ($1.50) through Amazon's Mechanical Turk, commensurate with minimum-wage standards. All of the participants were native English speakers, and all but 5 had taken one or fewer courses in introductory logic. 1 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, mean age = 34.2).

Preregistration and data-availability. The experimental design was pre-registered through the Open Science Framework platform (https://osf.io/f7ezg/). 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 4 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 times of the two events or a statement describing a relation between the end times. The two response options were incomplete descriptions of the given scenario.

Materials. Each problem description required 2 event labels and 4 timepoints (start and end times 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 computer/network events. The set of events were designed to plausibly co-occur. Across the 16 problems, each pair of labels was unique. The start and end times of the two events were randomly generated to correspond to the Allen relation assigned to each problem. Finally, to prevent participants from interpreting the two events as occurring on different days, each statement describing an event was appended with a preposition describing a particular day of the week, e.g., *the encryption started at 12pm and ended at 11pm on Friday*. The day of the week was randomly assigned. In total, each participant saw a unique set of problems.

Procedure. Each experimental problem began by displaying the event information. After a 2 second delay, the question and response options appeared. The participant selected a description by clicking on the button labeled with that description, which would trigger the display to load the next trial. The order of the response options was randomized for each problem.

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 Experiments 1 and 2 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. Bold values denote the most selected option.

Experiment 1	"started"	"ended"	*Experiment 2*	"started"	"ended"	neither
during	**64**	36	during	**51**	36	14
equals	**63**	37	equals	**38**	26	37
finishes	**65**	35	finishes	**51**	37	13
starts	40	**60**	starts	39	**44**	18
Total	**58**	42	Total	**44**	36	20

Results and discussion

The left side of Table 1 reports the response distributions for Experiment 1 as a function of the four different types of problems given to participants, i.e., as a function of different types of Allen relation. Overall, participants chose the statements describing start times over those describing end times 58% of the time, a rate significantly above chance (Wilcoxon test, $z = 2.68$, $p = .007$, Cliff's $\delta = 0.35$) and 33 of the 49 participants displayed this pattern (binomial test, $p = .02$).

The data were subjected to a Friedman analysis of variance, which showed that their responses differed as a function of the type of problem they received (Friedman test, $\chi^2 = 11.71$, $p = .008$). Post-hoc analyses showed that the *starts* relation differed from each of the three other relations (pairwise Wilcoxon tests, *p*s < .0001, Cliff's δs > .35). At least 27 of the 49 participants showed the pattern for each of the significant pairwise comparisons (binomial tests, *p*s < .05).

One limitation of Experiment 1 is the forced-choice nature of the task. The bias for preferring descriptions of the start of a set of events over descriptions of the end of a set of events may manifest only because participants had no option to respond that the two descriptions are roughly equivalent to one another. Experiment 2 therefore provided participants with a third response option.

Experiment 2

Experiment 2 was equivalent to Experiment 1 in all respects save one: participants in Experiment 2 were provided with the option to respond that neither of the two descriptions was better than the other. If a chronological bias in reasoning about durations is robust, people should exhibit it even when given an alternative option.

Method

Participants. 50 participants completed the experiment for monetary compensation ($1.50) through Amazon Mechanical Turk. All of the participants were native English speakers, and all but 12 had taken one or fewer courses in introductory logic. 2 participants were excluded from the analysis, 1 for always selecting the second response option and 1 for nonsense input in a post-experimental questionnaire. The analyses reported below are based on the remaining 48 participants (18 female, mean age = 36.0).

Open science. The pattern observed in Experiment 1 was pre-registered as the hypothesized results for the present experiment through the Open Science Framework platform (https://osf.io/f7ezg/). The same link makes the data from the study available.

Task and design. The task and design of Experiment 2 are similar to Experiment 1, except that Experiment 2 provided participants with the option to respond that *Neither description is better than the other.*

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

The left side of Table 1 displays the response distributions for Experiment 2; it shows that adding a "neither" response option did not qualitatively change the results. The pattern of the most frequently chosen responses by condition echoes the pattern of Experiment 1. The difference between a participant's proportion of start response selections and their proportion of end response selections (setting aside their "neither" responses) provides an index of the degree to which they preferred start to end times. Participants' overall difference scores were biased in favor of start times statements over end time statements (Wilcoxon test, $z = 2.20$, $p = .03$, Cliff's $\delta = .33$). Each individual Allen relation appeared to show the same bias as in Experiment 1. However, because participants selected the "neither" option 20% of the time, there was insufficient power to detect the chronological bias at the condition level (Wilcoxon tests, *z*s < 2.31, *p*s > .02, where .0125 serves as the corrected α using the Holm–Šidák correction for multiple comparisons).

The frequencies of responses to the three response options as a function of the four types of problem were subjected to a Fisher exact test, which showed that the distribution of participants' responses was significantly different from chance (Fisher's exact test, $p < .001$). Separate follow-up Fisher's tests were conducted for each individual relation: only the responses to the *equals* relation was at chance (Fisher's exact test, $p = 0.24$) while responses to the other 3 were significantly different from chance (Fisher's exact tests, *p*s < .002).

We ran a series of comparisons on participants' frequencies of responses for all pairwise comparisons of different Allen relations. In pairwise comparisons, the *equals* response distribution was different from the other 3 relation conditions (Fisher's exact tests, equals comparisons *p*s < 0.0001; all others, *p*s > .0125). The central reason for the difference appears to be because people were much more likely to provide a "neither" response for the *equals* relation.

In sum, Experiments 1 and 2 both revealed a robust chronological scanning bias. The result is predicted by the model theory of durational reasoning (Kelly & Khemlani, 2019; Khemlani et al., 2015), and it cannot be explained by formal systems of temporal reasoning, such as the Allen algebra.

General discussion

Two experiments showed that people have explicit preferences for some durational inferences over others, a preference that is best explained as a chronological scanning bias. Participants in each experiment reliably selected descriptions relating the start times of two durational events over descriptions about the end times. Experiment 1 was designed to create maximal competition between two ambiguous relations between durational events, since participants were forced to choose between descriptions such as:

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

Experiment 2 relaxed the competition by allowing for a correct response – that neither of the two events is better than the other. Nevertheless, more often than not, participants chose the option that described start times more often than the one that described end times. 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 & Khemlani, 2019; Kelly et al., under review; Khemlani et al., 2015).

Across both experiments, people exhibited a chronological scanning bias for every Allen relation tested except for the *starts* relation. Here is an example of a description that yields a *starts* relation:

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

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 due to a preference for timepoint asynchrony: participants preferred the description that referenced two different timepoints (e.g., when the *backup* and the *cyberattack* ended) over the description that referenced only one timepoint (e.g., when the two events started). But such a preference cannot explain people's chronological biases in the case of *equals* relations – in which both timepoints were synchronous – or *during* relations – in which both timepoints were asynchronous. Another explanation for the reversal may be to amend the proposal that reasoners scan models in chronological order: 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. And such an amendment may yield additional testable empirical predictions of reasoners' preferences for temporal conclusions, but it too has difficulty explaining why people exhibited a chronological bias for *equals* relations. In any case, the behavior suggests that something more than the chronological scanning bias affects people's inferential preferences.

Descriptions of durational relations can be ambiguous. You may say, for instance, that *the rain happened during concert*, to mean that the rain started after the concert began. To be more precise about the scenario, you might clarify it using durational verbs such as *started* and *ended*, e.g., *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 show that people prefer some ambiguous relations over others. The results suggest that humans simulate events with durations along a mental timeline, and that they consciously scan the simulation in order to reason about it.

Acknowledgments

This research was performed while the first author held an NRC Research Associateship award at the U.S. Naval Research Laboratory. It was also supported by a grant from the Office of Naval Research to the second author. 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, 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.

References

Allen, J. F. (1983). Maintaining knowledge about temporal intervals. *Communications of the ACM*, *26*, 832–843.

Bergen, B., Lau, C., & Ting, T. (2012). Writing direction affects how people map space onto time. *Frontiers in Psychology*, *3*, 109.

Bonato, M., Zorzi, M., & Umiltà, C. (2012). When time is space: Evidence for a mental time line. *Neuroscience and Biobehavioral Reviews*, *36,* 2257-2273.

Casasanto, D., & Jasmin, K. (2012). The hands of time: Temporal gestures in English speakers. *Cognitive Linguistics*, *23*, 643– 674.

Chatterjee, A., Southwood, M. H., & Basilico, D. (1999). Verbs, events and spatial representations. *Neuropsychologia*, *37*, 395-402.

Cooperrider, K., & Núñez, R. (2009). Across time, across the body: Transversal temporal gestures. *Gesture*, *9*, 181-206.

Fuhrman, O., & Boroditsky, L. (2010). Cross-cultural differences in mental representations of time: Evidence from an implicit nonlinguistic task. *Cognitive Science*, *34*, 1430-1451.

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*, 199-227.

Gevers, W., Reynvoet, B., & Fias, W. (2003). The mental representation of ordinal sequences is spatially organized. *Cognition, 87*, B87-B95.

Gevers, W., Reynvoet, B., & Fias, W. (2004). The mental representation of ordinal sequences is spatially organised: evidence from days of the week. *Cortex*, *40*, 171-172.

Goldvarg, E., & Johnson-Laird, P. N. (2001). Naive causality: A mental model theory of causal meaning and reasoning. *Cognitive Science*, *25*, 565-610.

Goodwin, G.P., & Johnson-Laird, P.N. (2005). Reasoning about relations. *Psychological Review*, *112*, 468–493.

Goranko, V., Montanari, A., & Sciavicco, G. (2004). A road map of interval temporal logics and duration calculi. *Journal of Applied Non-Classical Logics*, *14*, 9–54.

Hoerl, C., & McCormack, T. (2019). Thinking in and about time: A dual systems perspective on temporal cognition. *Behavioral and Brain Sciences*, *25*, 1–78.

Ishihara, M., Keller, P. E., Rossetti, Y., & Prinz, W. (2008). Horizontal spatial representations of time: Evidence for the STEARC effect. *Cortex*, *44*, 454-461.

Jahn, G., Knauff, M., & Johnson-Laird, P. N. (2007). Preferred mental models in reasoning about spatial relations. *Memory & Cognition*, *35*, 2075-2087.

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

Kelly, L., & Khemlani, S. (2019). The consistency of durative relations. In A. Goel, C. Seifert, & C. Freksa (Eds.), *Proceedings of the 41st Annual Conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.

Kelly, L., Khemlani, S., & Johnson-Laird, P.N. (under review). Reasoning about durations. Manuscript under review.

Khemlani, S., Byrne, R.M.J., & Johnson-Laird, P.N. (2018). Facts and possibilities: A model-based theory of sentential reasoning. *Cognitive Science, 42,* 1887–1924.

Khemlani, S., Barbey, A., & Johnson-Laird, P. N. (2014). Causal reasoning with mental models. *Frontiers in Human Neuroscience*, *8*, 849.

Khemlani, S., Harrison, A. M., & Trafton, J. G. (2015). Episodes, events, and models. *Frontiers in Human Neuroscience*, *9*, 590.

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*, 16766–16771.

Knauff, M. (1999). The cognitive adequacy of Allen's interval calculus for qualitative spatial relation and reasoning. *Spatial Cognition & Computation, 1*, 261-290.

Knauff, M., Strube, G., Jola, C., Rauh, R., & Schlieder, C. (2004). The psychological validity of qualitative spatial reasoning in one dimension. *Spatial Cognition & Computation*, *4*, 167-188.

Kowalski, R., & Sergot, M. (1989). A logic-based calculus of events. In *Foundations of Knowledge Base Management*. Berlin: 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.

Maass, A., & Russo, A. (2003). Directional bias in the mental representation of spatial events: Nature or culture? *Psychological science*, *14*, 296-301.

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

Peirce, C.S. (1931-1958). *Collected papers of Charles Sanders Peirce. 8 vols.* C. Hartshorne, P. Weiss, and A. Burks, (Eds.). Cambridge, MA: Harvard University Press.

Ragni, M., & Knauff, M. (2013). A theory and a computational model of spatial reasoning with preferred mental models. *Psychological Review*, *120*, 561-588.

Rauh, R., Hagen, C., Knauff, M., Kuss, T., Schlieder, C., & Strube, G. (2005). Preferred and alternative mental models in spatial reasoning. *Spatial Cognition & Computation*, *5*, 239-269.

Santiago, J., Lupáñez, J., Pérez, E., & Funes, M. J. (2007). Time (also) flies from left to right. *Psychonomic Bulletin & Review*, *14*, 512-516.

Santiago, J., Román, A., Ouellet, M., Rodríguez, N., & Pérez-Azor, P. (2010). In hindsight, life flows from left to right. *Psychological Research PRPF*, *74*, 59-70.

Schaeken, W., Johnson-Laird, P. N., & d'Ydewalle, G. (1996). Mental models and temporal reasoning. *Cognition*, *60*, 205-234.

Schaeken, W., & Johnson-Laird, P. N. (2000). Strategies in temporal reasoning. *Thinking & Reasoning, 6*, 193-219.

Tillman, K. A., & Barner, D. (2015). Learning the language of time: Children's acquisition of duration words. *Cognitive Psychology*, *78*, 57-77.

Tillman, K. A., Tulagan, N., Fukuda, E., & Barner, D. (2018). The mental timeline is gradually constructed in childhood. *Developmental Science*, *21*, e12679.

Torralbo, A., Santiago, J., & Lupiáñez, J. (2006). Flexible conceptual projection of time onto spatial frames of reference. *Cognitive Science*, *30*, 745-757.

Tversky, B. (1989). Parts, partonomies, and taxonomies. *Developmental Psychology*, *25*, 983-995.

Tversky, B., & Hemenway, K. (1984). Objects, parts, and categories. *Journal of experimental psychology: General*, *113*, 169-193.

Vallesi, A., Weisblatt, Y., Semenza, C., & Shaki, S. (2014). Cultural modulations of space–time compatibility effects. *Psychonomic Bulletin & Review*, *21*, 666-669.

Vicaro, C. M., Caltagirone, C., & Oliveri, M. (2007). Optokinetic stimulation affects temporal estimation in healthy humans. *Brain and Cognition*, *64*, 68–73.

Weger, U. W., & Pratt, J. (2008). Time flies like an arrow: Space-time compatibility effects suggest the use of a mental timeline. *Psychonomic Bulletin & Review, 15*, 426-430.

Winter, B., Marghetis, T., & Matlock, T. (2015). Of magnitudes and metaphors: Explaining cognitive interactions between space, time, and number. *Cortex*, *64*, 209-224.

Zacks, J., & Tversky, B. (2001). Event structure in perception and conception. *Psychological Bulletin,*127, 3-21.