A computational cognitive theory of temporal reasoning

Sangeet Khemlani
sunny.khemlani@nrl.navy.mil
Navy Center for Applied Research in Artificial Intelligence
US Naval Research Laboratory, Washington, DC 20375 USA

Abstract

I describe a novel model-based theory of how individuals reason deductively about temporal relations. It posits that temporal assertions refer to mental models -- iconic representations of possibilities -- of events. In line with recent accounts of spatial reasoning, the theory posits that individuals tend to build a single preferred model of a temporal description. The more models necessary to yield a correct answer, the harder that problem is. The theory is implemented in a computer program, mReasoner, which draws temporal deductions by building models. It varies three parameters governing separate factors in the process: the size of a model, the typicality of its contents, and the propensity to search for alternative models. Two experiments corroborate the predictions of the theory and its computational implementation. I conclude by discussing temporal and relational inference more broadly.

Keywords: temporal reasoning; events; mental models; reasoning; simulation

Introduction

People make temporal inferences when they schedule future events, reconcile past experiences, and attempt to understand ongoing scenarios. For instance, consider this description:

1. The car hit a pothole during the road trip.
   The car broke down after the road trip.
   Does it follow that the car hit a pothole before it broke down?

The words *during* and *after* are temporal relations – they describe how events and outcomes relate to one another – and reasoners have no difficulty inferring the correct answer ("yes") from the premises in (1). Indeed, English and other natural languages encode tense and aspect into every utterance, and so they provide abundant cues for drawing temporal conclusions. But some inferences are systematically easier than others. Contrast the example above with this problem (adapted from Schaeken et al., 1996):

2. The car hit a pothole before the road trip.
   The car's radio broke before the road trip.
   The car's windshield cracked when it hit a pothole.
   The car's headlights fused while on the road trip.
   Does it follow that the car's windshield cracked before its radio broke?

The correct answer – "no" – seems more difficult to infer compared to (1). Why? Many factors distinguish (2) from (1): it has more premises, and describes more events; it uses more temporal relations – *before*, *when*, and *while*; and its correct answer is negative instead of affirmative. Yet these factors don't provide an adequate explanation of the mental representations and processes humans use to reason about time. And no computational cognitive theory exists that's robust enough to simulate why (1) is an easy inference to make; why (2) is more difficult (though cf. the computer model described in Schaeken et al., 1996); and how people generate rational responses to even difficult temporal reasoning problems.

In what follows, I briefly summarize previous computational treatments of temporal deduction, and show how they are psychologically implausible. Next, I synthesize a theory of temporal reasoning based on how humans simulate the passage of time. I argue that to reason about time, humans can construct a mental timeline of events, i.e., an event model. These event models are easy to process when humans build and maintain just one in memory, but difficult to process when they need to maintain multiple event models. I describe a computational implementation of the theory and the predictions it makes, as well as a series of studies designed to test those predictions. And I show how the computational implementation fits data from those studies, and how it can model additional forms of temporal inference. I conclude by contrasting the theory with alternative proposals.

The logic of temporal reasoning

Systems of symbolic logic are designed to generate the correct answers to problems such as (1) and (2). Temporal logics, such as Prior's (1967) tense logic and Allen's (1983) interval calculus, treat each premise as a formula describing a temporal relation between events, and can be written as, e.g., `during(X,Y)` where `X` can stand in place for any proposition, such as `hitAPothole(car)`. Many systems of temporal logic in AI (e.g., Allen, 1991; Freksa, 1992; Øhrstrøm & Hasle, 1995; see also Fischer, Gabbay, & Vila, 2005; Goranko, Montanari, & Sciavicco, 2004 for reviews) posit primitive temporal relations that do not map into simple everyday English (Knauff, 1999) or other natural language expressions, and likewise, many temporal relations in natural language are flexible in ways AI systems cannot characterize. For instance, AI systems often neglect Reichenbach's (1947) distinction between different points of reference in natural language, and so they are insensitive to the distinction between, e.g., "I had done it" (past perfect tense) versus "I did it" (past tense). For many AI applications, these distinctions are irrelevant – but they ensure that such systems cannot interface with the full range of natural language capabilities (see, e.g., Khemlani & Johnson-Laird, 2019).

Event calculi (see, e.g., Kowalski & Sergot, 1986) may be more psychologically plausible, because they describe inference rules and axioms between two or more temporal relations, so they abide by the constraints of logic-based cognitive accounts of reasoning (e.g., Rips, 1994; see also Bringsjord & Govindarajulu, 2020). But, a limitation common to all temporal logics and event calculi are that they describe only valid deductions: they have no capacity of explaining what happens when reasoners err. And so they cannot explain why (1) is easy and why (2) is hard. For that, we turn to a psychological theory of temporal reasoning.

Mental models of events

The theory of temporal deduction I present is based on the tenets of mental model theory – the "model" theory for short (Johnson-Laird, 2006). The theory states that when people reason, they use language observation, and imagination to construct and mentally manipulate possibilities. The theory is based on several fundamental principles:

- **Mental models are iconic representations of possibilities.** That is, the structure of a mental model corresponds to the structure of what it represents as far as possible (Peirce, 1931-1958, Vol. 4). Models of temporal relations can use space to represent time by constructing mental timelines in which tokens represent events (Schaeken et al., 1996), or they can represent sequences of events as they unfold in time (Khemlani et al., 2013).
- **Models represent durations as discrete episodes.** Reasoners encode durations and intervals by representing episodes that mark the starts and ends of events (Khemlani et al., 2015a). By default, people do not maintain representations of metric time. To comprehend specific intervals, as in, *the meeting lasted 2 hours*, individuals tag events with ancillary information, and then reason arithmetically.
- **The principle of emergent consequences.** Logical relations are emergent consequences of iconic structure of models – and so no special logical rules, operations on formulas, or syntactic transformations are necessary for individuals to reason logically (Goodwin & Johnson-Laird, 2005).
- **Inferences are easier with one model; multiple models yield errors.** Human reasoning is based on two interacting sets of processes: one system produces rapid, intuitive inferences by building a single model. Hence, people are faster and make fewer errors when considering descriptions that yield only one model. When descriptions yield multiple models, i.e., when an initial model doesn't suffice, reasoners are more prone to errors (Khemlani & Johnson-Laird, 2017) and they take longer (Schaeken & Johnson-Laird, 2000).

The model theory posits that to simulate relations between events, people have two options: first, they can simulate a series of events in the same order as they would unfold. For example, to represent an individual's meals over the course of a day, you might simulate the individual eating breakfast, then lunch, then dinner, focusing only on each single meal at a time. By doing so, reasoners build kinematic mental models, i.e., they use time to represent time (Khemlani et al., 2013). Kinematic models may be particularly useful when following complex narratives, e.g., during discourse comprehension (Cain & Oakhill, 1999; Garnham, 2013; Graesser, Millis, & Zwaan, 1997; Zwaan & Rapp, 2006), though they can obscure the temporal relations between simultaneous events and events with durations. To reason directly about such relations, people can use space to represent time (Schaeken et al., 1996, 2000), e.g., they can construct a mental model for (1) in a way that can be depicted in the following diagram:

```
[   road-trip   ]   broke-down
   hit-pothole
```

The diagram represents the events iconically, i.e., with words that stand in place of mental simulations of the event itself. Its spatial layout presents the events in chronological order, from earliest to latest (see Kelly & Khemlani, 2020, under review). And it uses markers to designate the initiation (`[`) and conclusion (`]`) of a durative event, namely to depict that hitting a pothole occurred in the time between when the road trip started and ended (Kelly, Khemlani, & Johnson-Laird, 2020). The logical consequences emerge from the model's structure – by scanning it, reasoners can draw many different valid conclusions, e.g.,

the road trip happened before the car broke down;
the car hit a pothole before the breakdown;
the road trip ended after the car hit a pot-hole;
the road trip started before the breakdown;

and so on. Hence, the model serves as a compact, efficient representation to facilitate reasoning.

Models predict difficulty, because inferences that require multiple models place a higher demand on working memory resources. So, what makes (2) difficult is not just that it has more premises, or more relations. Rather, it's difficult because the description yields multiple models. This model satisfies the premises:

```
hit-pothole    radio-broke    [    road-trip     ]
windshield                       fused-headlight
```

but so does this one:

```
radio-broke    hit-pothole    [    road-trip     ]
               windshield       fused-headlight
```

Hence, the conclusion in (2) doesn't follow necessarily. To get the correct answer, reasoners must either initially build the second model above, or else keep both models in mind and compare the two. The theory accordingly predicts that all other things being equal, inferences that demand more models should be more difficult – they should produce more errors. I turn next to describe a computational implementation of the theory.

Temporal reasoning in mReasoner

mReasoner is a computational cognitive reasoning engine that implements the core tenets of the model theory (Khemlani & Johnson-Laird, 2022). The system is equipped with a small grammar that parses and builds iconic mental models for assertions concerning quantity (e.g., “Most of the potholes are large”), causality (e.g., “Hitting the pothole caused the breakdown”), and sentential inference (e.g., “The windshield cracked or else the headlights fused”), and it mimics patterns of human reasoning in all these domains (Briggs & Khemlani, 2019; Khemlani et al., 2015b, 2018). Updates to its components permitted it to reason about temporal relations. Figure 1 depicts a schematic of the system and shows how it draws the correct conclusion for (1). I review each updated component and their functionality in turn.

Building integrated models

The first component parses premises from natural language into *intensions*, which serve as blueprints for building models. Intensions provide a modal semantics for the meaning of an assertion. The system parses a variety of different temporal assertions, e.g., those describing connectives such as *before, after, while,* and *during.* The intensions of each assertion specify how to construct an initial model, as well as serve as a guide to the space of possible revisions on the model (see Khemlani & Johnson-Laird, 2022). The semantics is as follows:

A happened before B.	→	$A < B$
A happened after B.	→	$A > B$
A happened while B.	→	$A = B$
A happened during B.	→	$A \subseteq B$

mReasoner builds temporal models by integrating multiple temporal intensions, e.g., it builds an initial model of the first assertion in (1), and then updates that model with information about the second assertion, yielding an integrated model of the two relations. One subtlety of this procedure is that it is sensitive to the order in which it processes premises in that, by default, the system treats events in premises as punctate – but when necessary, it converts a punctate event into a durative one. This example illustrates the phenomenon:

The meeting happened before the conference.
The sale happened during the conference.

As in all temporal assertions, the events (*the meeting, the conference*) can be treated as single points or multiple points on a timeline. The system starts by building a model of the first premise:

```
meeting    conference
```

(such that *the sale* is contained within an interval), mReasoner breaks the punctate event into two markers because the second premise treats *the conference* as durative by explicitly represent its start and end, e.g.,

```
meeting   [ conference ]
                sale
```

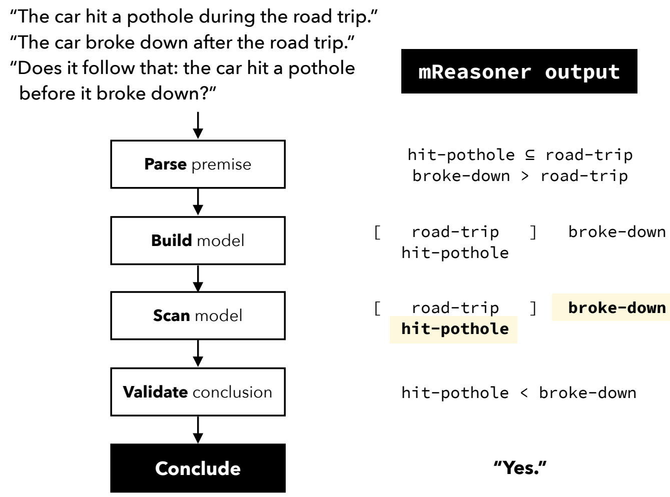

Figure 1. Four components of the mReasoner computational cognitive model that generate conclusions given temporal premises. The system *parses* premises into intensions; *builds* an initial model from those intensions; *scans* the model to locate events in a given premise; and *validates* a relation between those located events. If its deliberative system is engaged, mReasoner can engage a search for counterexamples to decide whether its initial conclusion necessarily follows, and it can modify the conclusion if necessary (see Khemlani & Johnson-Laird, 2022).

Another subtlety of the model-building component concerns how to construct indeterminate descriptions. Consider this set of premises:

The ceremony happened before the storm.
The newscast happened before the storm.

By default, mReasoner constructs and reasons with a single model at a time. But the description above is consistent with several different models, e.g., one in which the sale happens before the meeting, another in which the sale happens after the meeting, a third in which the meeting happens during the sale, and so forth. To build an initial model from indeterminate descriptions, mReasoner adopts heuristic strategies for constructing models initially developed for a theory of spatial reasoning (see Ragni & Knauff, 2013, p. 567). That is, by default mReasoner inserts new events at the first available location:

`ceremony  storm  →  newscast  ceremony  storm`

but it can also insert events so that they occur in a way that “spreads apart” existing events in the model, as in:

`ceremony  storm  →  ceremony  newscast  storm`

These two strategies are governed by a probabilistic ***typicality*** parameter that ranges from `0` to `1`, and controls the probability of engaging in the latter strategy (see Johnson-Laird et al., 2015; Khemlani & Johnson-Laird, 2022 for additional information on this parameter). In this way, the system mimics the variation in humans’ construction of temporal models.

Scanning models, drawing conclusions, and searching for counterexamples

To draw conclusions, mReasoner scans an integrated model with respect to a given temporal conclusion. For example, it scans an integrated model of (2) above for the two events specified in the conclusion (i.e., the windshield cracking and the reading breaking). If the events are represented in the model, the system generates an intension that describes their temporal relation, and converts that intension back into natural language – and if that relation happens to match the prompt (*the windshield cracked before the radio broke*) then the system responds affirmatively. In all other cases, including those in which it cannot locate an event in the model, the system responds negatively.

As previous investigations of temporal reasoning reveal, humans are able to reason about extraordinarily complex temporal descriptions. Hence, descriptions that concern multiple mental models may pose difficulties for reasoners, but many reasoners are skilled in their ability to overcome such difficulties. A viable theory of temporal reasoning must explain, not just why some problems are more difficult, but how certain individuals manage to provide correct responses despite such difficulties. The model theory proposes – and mReasoner implements – the idea that rational responses often depend on the interrogation of initial responses: people recognize, for instance, that descriptions are consistent with multiple models, and so they attempt to build those models. Their attempts may result in a model in which the premises are true but their initial conclusion is false – i.e., a "counterexample". The model theory further proposes that the search for counterexamples is not a sampling procedure (pace Phillips, Morris, & Cushman, 2019). Instead, the theory posits that reasoners make incremental changes to the events represented in their initial model. Evidence supporting such a procedure comes from the fact that difficult problems are easier when they require counterexamples that have a smaller "edit distance" to the initial model (Ragni, Khemlani, & Johnson-Laird, 2014).

For temporal reasoning, counterexample search depends on some combination of the following 5 strategies: i) shifting an event earlier in time; ii) shifting an event later in time; iii) converting a punctate event into a durative one; iv) converting a durative event into a punctate one; v) expanding a durative event, i.e., shifting a token representing its start to an earlier time and a token representing its end to a later time. The system attempts each strategy in turn in a recursive fashion, and stops when it discovers a counterexample. But, its ability to search for counterexamples in the first place is not turned on by default. It is governed by a ***search*** parameter (see Khemlani & Johnson-Laird, 2022) that controls the probability of engaging in a search for counterexamples.

These augmentations to the mReasoner computational model provide it with the means to mimick human temporal reasoning. The next section describes experiments designed to test the theory's prediction that one-model problems are easier than multiple-model problems; and the section that follows describes mReasoner's fit to the resulting data.

Experiments 1 and 2

We conducted two experiments to test the computational model described in the previous section. Each experiment presented participants with the same 8 reasoning problems, though the contents of the premises were randomized. Here is an example problem:

The suspect set up surveillance before he closed his bank account.
The suspect destroyed the laptop after he closed his bank account.
The suspect hired the lawyer while he set up surveillance.

The model theory predicts that the problem should be easy, since the premises in Experiment 1 are consistent with only one model, this one:

```
surveillance   closed-account   destroyed-laptop
hired-lawyer
```

Half of the problems were consistent with one model, and the other half were consistent with multiple models. In all other respects, namely, the specific contents, the number of premises, the events in the premises, and the number and type of temporal relation, the 4 one-model problems and 4 multiple-model problems were matched.

Participants in Experiment 1 were given three separate conclusions: a valid conclusion, a foil, and a null conclusion, i.e., "there's not enough information to conclude anything". Participants in Experiment 2 carried out the same problems, but instead generated their own responses to questions of the form:

What is the relationship between when the subject hired the lawyer and when he destroyed the laptop?

Participants' natural responses were coded for accuracy.

Method

Participants. A total of 61 participants (31 in Experiment 1 and 30 in Experiment 2) were recruited through Amazon Mechanical Turk. Participants who failed to answer attention checks, misunderstood the task, or performed the entire study under 2 minutes were dropped from analysis. This resulted in data from 56 participants (28 in Experiment 1 and another 28 in Experiment 2).

Design, procedure, and materials. Each participant was presented with 10 three-premise causal inference problems: 4 were predicted to be one-model problems and 4 that were multiple-model. The other 2 were practice problems that also served as attention checks, and were discarded from analysis. Each problem consisted of three premises describing temporal relations that were randomly selected from a pool of events that described the activities of a criminal suspect, e.g., "shredded the documents", "transferred the drug funds", "build the explosive", and so on. Each premise consisted of a pair of activities linked by 1 of 4 temporal connectives (*before, after, during,* and *while*). The activities were chosen such that they could be interpreted as durative or punctate (see Kelly et al., 2020), and yield a coherent narrative no matter how they were ordered (e.g., in the example above, the narrative would be coherent even if the suspect hired a lawyer before he transferred the drug funds). The order in which the participants carried out the 10 problems was randomized, as was the assignment of the contents of the premises.

Task. Experiment 1 provided participants with three response options: a valid conclusion, an invalid conclusion, and a null conclusion. In Experiment 2, participants typed out their responses to a question relating two events in the problem, i.e., "What is the relationship between __ and __?" I coded their responses for accuracy blind to the specific condition.

Open science. Data, materials, experimental code, mReasoner code, and synthetic data derived from computational simulations are available at https://osf.io/26ckg/.

Results

Both experiments showed that participants were more accurate for one-model problems than multiple-model problems (in Experiment 1, one- vs. multiple-model: 70% vs. 37%; Wilcoxon test, z = 5.08, p < .001, Cliff's δ = .33; in Experiment 2, one- vs. multiple-model: 78% vs. 44%; Wilcoxon test, z = 5.25, p < .001, Cliff's δ = .35). The results corroborate the model theory of temporal reasoning. In addition, Experiment 2 captured the response time between when participants read the three premises and when they began to type out a response. Analysis of Winsorized response times revealed that participants were faster to respond to one-model problems (51.77 s) than multiple-model problems (60.01 s; Wilcoxon test, z = 3.15, p = .002, Cliff's δ = .19). These results, too, corroborate the model theory's difficulty prediction. For brevity, I omit further analyses of the data in favor of describing the mReasoner's simulations of the two studies.

Simulation of Experiments 1 and 2

To simulate the 8 problems in Experiments 1 and 2, mReasoner generated datasets by systematically varying the settings of two of its parameters (Busemeyer & Diederich, 2010), i.e., the ***`atypicality`*** and ***`search`*** parameters described above, along with a ***`size`*** parameter that stochastically limited the size of each model. The parameter settings were quantized to span their ranges as follows:

```
       size: 2.0, 2.5, 3.0, 3.5, 4.0, 4.5*, 5.0**
atypicality: 0.0, 0.2, 0.4, 0.6, 0.8, 1.0
     search: 0.0, 0.2, 0.4, 0.6, 0.8, 1.0

                              * Exp 1   ** Exp 2
```

Hence, the system generated 7 × 6 × 6 = 252 separate simulated datasets. The system carried out the 8 problems 100 times for each of the 252 parameter settings. A grid search was used to locate the best fitting parameter settings for the data for Experiments 1 and 2. The grid search depended on minimizing the root mean squared error (RMSE) between the dataset and the proportions of responses in each simulated dataset across the 8 problems. Once the grid search located the best-fitting parameter settings (which were quite similar, and bolded above), the parameters were fixed and mReasoner carried out the 8 problems 1000 times each. Figure 2 plots the computational modeling simulations against the results from each dataset.

The computational model yielded a close fit to the data (r = .95, RMSE = .22 for Experiment 1; r = .93, RMSE = .20). And the optimizing parameter values located from by the grid search were sensible: the computational model fit the data when the size of the models was large (> 4), when the system considered atypical models 40% of the time, and when the system never engaged in a search for counterexamples. Searching for counterexamples is demanding, and most of the time, particularly for complex problems, reasoners appear to satisfice and base their inferences on the first model they construct.

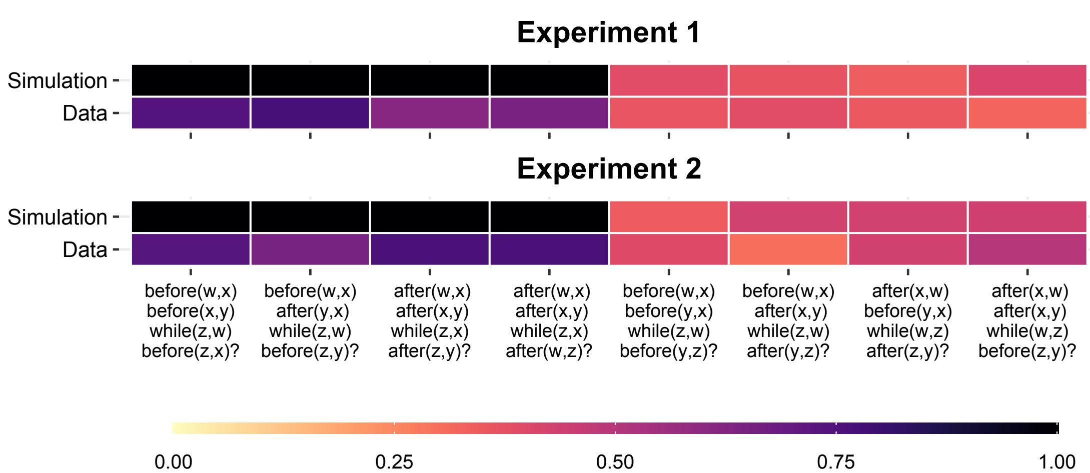

Figure 2. The proportions of correct responses to the 8 problems in Experiments 1 and 2, along with the proportions of correct responses generated by mReasoner's best-fitting simulations (rs = .95 and .93 for Experiments 1 and 2, respectively). The 8 problems in the two experiments are provided using schematic formulas in place of the natural language sentences participants received, e.g., participants saw premises akin to, "The suspect destroyed the laptop after he closed his bank account" instead of `after(y,x)`. Participants' evaluated a given response – denoted by the question mark – in Experiment 1, and specified the relation between two events in Experiment 2.

General discussion

Human reasoning about time is complex: events can be punctate or stretch across other events; they can be cyclical, as in the passage of seasons; and they can endure across fixed units that can be enumerated. Nevertheless, humans must make rapid inferences about relations to understand narratives and plan for future scenarios. I describe a theory of temporal cognition that relies on the construction, maintenance, and manipulation of event models. The theory accounts for how people represent durations and what makes reasoning about time difficult. The theory is embodied in mReasoner, a computational cognitive implementation of the model theory of thinking and reasoning (Khemlani & Johnson-Laird, 2022). In this paper, I described innovations to the system that can predict which temporal reasoning problems prompt reasoners to make errors; I described experiments designed to test the theory's central predictions; and I showed how the computational model fit the data from those studies.

The computational model explains only a small subset of temporal reasoning phenomena: it doesn't account for how people rapidly process and interpret tense and aspect, or how they cope with information about metric time. But, the theory does explain how people without any background in temporal logic can make valid deductions from temporal premises. Previous psychological accounts of reasoning have argued that people maintain axiom systems and build proofs to make temporal deductions (see, e.g., Rips, 1994). Meanwhile, probabilistic frameworks of reasoning either build off such logical frameworks, or else eschew any consideration of temporal inference whatsoever (see Knauff & Gazzo Castańeda, 2022). Neither approach can explain the systematic errors people make in Experiments 1 and 2. The system I describe therefore serves as a cognitive process model of temporal reasoning: it specifies both the structure of the mental simulations people build, as well as the algorithms they use to process those representations.

Acknowledgments

This research was supported by funding from the Naval Research Laboratory to study temporal reasoning. I'm grateful to Kalyan Gupta, Danielle Paterno, Kevin Zish, and Bryan Auslander at the Knexus Research Corporation for their help in conducting the experiments. For their helpful suggestions and critiques, I also thank Bill Adams, Branden Bio, Gordon Briggs, Ruth Byrne, Hillary Harner, Tony Harrison, Phil Johnson-Laird, Laura Kelly, Markus Knauff, Andrew Lovett, Marco Ragni, and Greg Trafton.

References

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

Allen, J. F. (1991). Time and time again: The many ways to represent time. *International Journal of Intelligent Systems*, 6, 341–355.

Briggs, G. & Khemlani, S. (2019). A cognitively plausible algorithm for casual inference. In T. Stewart (Ed.), *Proceedings of the 17th International Conference on Cognitive Modeling.*

Bringsjord,S., & Govindarajuli, N. S. (2020). Rectifying the mischaracterization of logic by mental model theorists. *Cognitive Science, 44,* e12898.

Busemeyer, J. R., & Diederich, A. (2010). *Cognitive modeling*. Sage.

Cain, K., & Oakhill, J. (1999). Inference making ability and its relation to comprehension failure in young children. *Reading and Writing,* 11.

Fisher, M. D., Gabbay, D. M., & Vila, L. (2005). *Handbook of temporal reasoning in artificial intelligence*. Amsterdam: Elsevier.

Freksa, C. (1992). Temporal reasoning based on semi-intervals. *Artificial Intelligence*, 54, 199–227.

Garnham, A. (2013). *Mental models and the interpretation of anaphora.* Psychology Press.

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

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.

Graesser, A. C., Millis, K. K., & Zwaan, R. A. (1997). Discourse comprehension. *Annual Review of Psychology*, 48, 163-189.

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

Johnson-Laird, P. N., Khemlani, S., & Goodwin, G.P. (2015). Logic, probability, and human reasoning. *Trends in Cognitive Sciences*, 19.

Kelly, L. & Khemlani, S. (2020). Directional biases in durative inference. In B. Armstrong, S. Denison, M. Mack, and Y. Xu (Eds.), *Proceedings of the 42nd Annual Conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.

Kelly, L., & Khemlani, S. (under review). Iconicity bias and duration. Manuscript under review.

Kelly, L., Khemlani, S., & Johnson-Laird, P.N. (2020). Reasoning about durations. *Journal of Cognitive Neuroscience,* 32, 2103-2116.

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

Khemlani, S., Harrison, A. M., & Trafton, J. G. (2015a). Episodes, events, and models. *Frontiers in Human Neuroscience*, 9, 590, 1-13.

Khemlani, S., & Johnson-Laird, P.N. (2013). The processes of inference. *Argument & Computation*, 4, 1-20.

Khemlani, S., & Johnson-Laird, P.N. (2017). Illusions in reasoning. *Minds & Machines*, 27, 11-35.

Khemlani, S., & Johnson-Laird, P.N. (2022). Reasoning about properties: A computational theory. Manuscript in press at *Psychological Review*.

Khemlani, S., & Johnson-Laird, P.N. (2019). Why machines don't (yet) reason like people. *KI – Künstliche Intelligenz*, 33, 219-228.

Khemlani, S., Lotstein, M., Trafton, J.G., & Johnson-Laird, P.N. (2015b). Immediate inferences from quantified assertions. *Quarterly Journal of Experimental Psychology*, 68.

Khemlani, 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.

Khemlani, S., Orenes, I., & Johnson-Laird, P.N. (2012). Negation: a theory of its meaning, representation, and use. *Journal of Cognitive Psychology*, 24.

Khemlani, S., Wasylyshyn, C., Briggs, G., & Bello, P. (2018). Mental models of omissive causation. *Memory & Cognition*, 46.

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

Kowalski, R., & Sergot, M. (1986). A logical-based calculus of events. *New Generation Computing*, 4, 67–95.

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.

Phillips, J., Morris, A., & Cushman, F. (2019). How we know what not to think. *Trends in Cognitive Sciences*, 23, 1026-1040.

Ragni, M., Khemlani, S., & Johnson-Laird, P.N. (2014). The evaluation of the consistency of quantified assertions. *Memory & Cognition*, 42, 53-66.

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

Reichenbach, H. (1947). *Elements of symbolic logic.* New York: Free Press.

Rips, L. J. (1994). *The psychology of proof: Deductive reasoning in human thinking.* Cambridge, MA: MIT Press.

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

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

Zwaan, R. A., & Rapp, D. N. (2006). Discourse comprehension. In *Handbook of psycholinguistics* (pp. 725-764). Academic Press.