Reasoning with online and offline knowledge

Branden J. Bio* and Sangeet Khemlani
{branden.bio.ctr, sangeet.khemlani}@nrl.navy.mil
*NRC Postdoctoral Fellow
Navy Center for Applied Research in Artificial Intelligence
U.S. Naval Research Laboratory, Washington, DC 20375 USA

Abstract

Knowledge affects how humans think and reason: people use background knowledge to interpret natural language, and reason over those interpretations. We show one way in which offline knowledge, which is stored in semantic memory, interacts with online knowledge, that is, knowledge acquired through the use of *factive* mental state verbs such as *know* and *discover*. The interaction tests a theory of human thinking that assumes people construct simulations of possibilities – mental models – when they reason. It predicts that offline knowledge can affect reasoning through a process known as modulation, which blocks the mental construction of possibilities; and that online knowledge can cause reasoners to make presuppositions about facts. It also describes the mechanisms by which the mind updates mental models and separates fact from belief. An experiment tested the theory and corroborated its predicted interaction effect. We discuss the results in light of recent proposals of reasoning with knowledge.

Keywords: knowledge; belief; epistemic reasoning; mental models; cognitive verbs

Introduction

Humans base their everyday reasoning on both *online knowledge*, that is, the facts and presuppositions they pick up as through expression, and *offline knowledge*, which is stored in memory and recalled as needed. Suppose, for instance, that you are told the following:

1. Howard knows that pterodactyls are not dinosaurs.

The statement directly expresses knowledge that Howard possesses, and it indirectly asserts a fact about the world (paleontologists classify pterodactyls as pterosaurs, not dinosaurs). Non-experts can use statements such as (1) to establish facts and correct misconceptions online, because the example uses the verb *know* to establish what is true. Cognitive scientists and linguists refer to verbs such as *know, think, believe,* and *discover* as "propositional attitude verbs", "epistemic verbs", "cognitive verbs", "mental state verbs", and "verbs of knowledge", among other terms. Children acquire them early in development, and by age 5, people are capable of making sophisticated inferences from expressions of knowledge (e.g., Adrián, Clemente, & Villanueva, 2007; Abbeduto & Rosenberg, 1985; Booth & Hall, 1995; Forrester, 2017; Lewis, Hacquard, & Lidz, 2017; MacWhinney, 2000; Moore, 2013; Perner & Roessler, 2012; Schwanenflugel, Henderson, & Fabricius, 1998). Early in development, children understand that some verbs, such as *know* and *discover*, are factive: their complements presuppose some true condition about the world (e.g., Cohen, 1992, p. 91; Dudley, Rowe, Hacquard, & Lidz, 2017; Fetzer & Johansson, 2010; Kiparsky & Kiparsky, 1970; Stalnaker, 1999, p. 55; but cf. Hazlett, 2012).

As cognitive scientists discovered in the middle 20th C., background knowledge – *offline* knowledge – can systematically change how people reason (see Khemlani, 2018, for a review). Consider your background knowledge when making this inference:

2. Aiden is in Port-au-Prince or he is in Haiti.
   Aiden is in Port-au-Prince.
   Does it follow that he is in Haiti?

The logical structure of (2) is:

3. P or Q or both.
   P.
   Does it follow that Q?

where *P* and *Q* can stand in place for propositions. Under any orthodox system of logic, *Q* doesn't necessary follow given the premises in (3) – it is not a valid conclusion, because the premises can be true even when *Q* is false (see Jeffrey, 1981). The primary difference between (2) and (3) is the knowledge people may possess, namely that Port-au-Prince is Haiti's capital. This knowledge changes the validity of the inference – the correct answer to (2) is "yes".

Offline knowledge concerns the commonsense facts and concepts that reasoners possess in the form of declarative, episodic, and semantic memory (Kumar, 2021; Renoult, Irish, Moscovitch, & Rugg, 2019; Squire, 2004). Consider again that knowledge can be expressed (see example 1 above). The effects of offline knowledge on human reasoning are robust and wide-ranging: for instance, people are better at reasoning about familiar contents than abstract ones (Wason & Shapiro, 1971; Wilkins, 1929; see also Johnson-Laird & Wason, 1972 for a review), and contemporary linguistic AI systems that learn from large corpora of natural language appear to reflect such patterns (Dasgupta et al., under review). Likewise, reasoners prefer conclusions that are believable – that is, conclusions that accord with background knowledge – over unbelievable conclusions, even when those believable conclusions are logically invalid (Morgan & Morton, 1944; for reviews on the "belief bias" effect, see Dube, Rotello, & Heit, 2010; Evans et al., 1993; Johnson-Laird, 1999).

In everyday reasoning, people use both the knowledge they glean from expression (online knowledge) as well as knowledge they already possess (offline knowledge) to make inferences. While many theories of reasoning accept that people base inferences on the knowledge they have, few specify mechanisms to explain how online and offline knowledge interact. In what follows, we explore such a theory of reasoning. The theory argues that people use offline knowledge to simulate sets of possibilities; that they convert discourse and online knowledge into simulated sets of possibilities; and that they reconcile online and offline possibilities by combining these two sets. We present a novel phenomenon uniquely predicted by the theory, i.e., that certain patterns of reasoning depend on the consideration of both online and offline knowledge. We also describe a preregistered experimental test of the phenomenon. The experiment corroborates the theory's prediction and rules out alternative hypotheses. We conclude by considering limitations of the results and drawing lessons for advancing theories of epistemic reasoning.

Mental models of online and offline knowledge

Many linguists consider epistemic verbs, such as *know*, *believe*, *think*, *discover*, and *infer*, to be "modal" in nature: they refer to possibilities (Portner, 2009). On their analysis, to say that somebody "knows" or "believes" something is to refer to a potentially infinite number of "possible worlds" of those mental states, and theorists propose semantic accounts based on computing such infinite sets (see, e.g., McKinsey, 199; Richard, 1983). Partee (1979) notes that possible worlds are too big for individual minds to compute, and are accordingly implausible as psychological theories. Nevertheless, recent psychological theories appeal to possibilities to explain higher cognitive operations (Carey, Leahy, Redshaw, & Suddendorf, 2020; Johnson-Laird & Ragni, 2019; Phillips, Morris, & Cushman, 2019; though cf. Johnson-Laird, Khemlani, & Goodwin, 2015, for a review of theories of reasoning that eschew possibilities). For instance, researchers argue that people construct possibilities when they comprehend cause and effect (Henne et al., 2019, 2021; Morris et al., 2019), make moral judgements (Phillips & Cushman, 2017; Schtulman & Tong, 2013), reason about spatiotemporal arrangements (Knauff, 2013; Radvansky & Zacks, 2014), and think about what could have been (Byrne, 2005, 2016; Trickett & Trafton, 2007).

Such accounts can be traced to Craik (1943), who was the first psychologist to argue that humans make decisions by constructing "small-scale models" of the world and by mentally "trying out various alternatives" on those models. Johnson-Laird (1983, 2006) theorized that mental models underlie, not just decision-making, but perception, communication, reasoning, and imagination as well. Mental models – i.e., mental simulations of possible situations – account for reasoning about cause and effect (Goldvarg & Johnson-Laird, 2001; Khemlani et al., 2018, 2021), quantifiers such as all, some, and most (Khemlani & Johnson-Laird, 2022), space and time (Kelly, Khemlani, & Johnson-Laird, 2020; Knauff, 2013; Ragni & Knauff, 2013), and sentential operators such as *and*, *if*, and *or* (Khemlani et al., 2018). The "model" theory rests on three assumptions:

1. People construct models to reason. Assertions in natural language refer to alternative possibilities, and people reason by mentally constructing finite models – small sets of such possibilities – consistent with relevant information (Johnson-Laird, 2006; Khemlani, Byrne, & Johnson-Laird, 2018). Like pictorial representations, mental models are iconic, i.e., they reflect the structure of the scenarios they represent (Peirce, 1931-1958, Vol. 4); and they are coherent, i.e., they are internally consistent, because they cannot directly represent impossible scenarios, such as an object that is both a circle and a square. But, unlike pictorial representations, models can include abstract symbols, such as the symbol for negation (Khemlani, Orenes, & Johnson-Laird, 2012). And they can be dynamic, i.e., they can represent elements that change over time (Khemlani, Mackiewicz, Bucciarelli, & Johnson-Laird, 2013).

2. Default models. People default to a core "modal semantics" in the absence of knowledge to the contrary. The meanings of causal, temporal, conditional, and other sorts of assertions pick out those situations that are possible and impossible. To economize working memory resources, the models people initially construct tend to concern only one of the possibilities consistent with an assertion, and they do not represent what is false or impossible (Khemlani & Johnson-Laird, 2017). For instance, to reason about conditionals such as *if it rains then it's windy*, reasoners often represent only the possibility in which it rains and it's windy (Johnson-Laird & Byrne, 2002). Further deliberation helps them consider possibilities consistent with the conditional, such as cases in which it doesn't rain.

3. Offline knowledge "modulates" possibilities. Reasoners can use their general knowledge concerning the clauses in assertions (e.g., *it rains* and *it's windy*) to add information to the models they construct. When that information introduces causal, spatiotemporal, or other relational dependencies, it can prevent individuals from considering possibilities that core meanings would otherwise allow (Johnson-Laird & Byrne, 2002), thereby narrowing the space of alternative models to consider.

We illustrate the theory and its mechanisms for building and combining models by example. Consider this conditional statement:

4. If it's an animal, then it's hidden.

Reasoners can understand it and make systematic inferences from it. For instance, they readily infer that:

5. It is possible that it's an animal and it's hidden.

That is, they can draw modal conclusions about possibility from non-modal premises (Hinterecker, Knauff, & Johnson-Laird, 2016; Johnson-Laird & Ragni, 2019; Khemlani et al., 2018). According to the model theory, people interpret (4) by building a mental model of a possibility, which is embodied in the following diagram:

```
animal      hidden
         ...
```

where the ellipsis denotes that individuals implicitly track that other scenarios are possible. Sentential logic does not concern possibilities, though other sorts of logic do (Hughes & Cresswell, 1996). Unlike these systems, the model theory treats possibilities as rudimentary: models *are* possibilities, and all reasoning – even reasoning about truth and falsity – is a form of modal reasoning (Johnson-Laird & Ragni, 2019; Khemlani et al., 2018). Hence, no special token denoting possibilities is necessary.

As Espino, Byrne, and Johnson-Laird (2020) show, reasoners understand that the force of a conditional is to describe some situation (i.e., that it's hidden) conditioned on some other situation (i.e., that it's an animal). At first, they explicitly build one model to interpret the conditional in (4), though many implicitly realize that it may not be an animal. By deliberating, reasoners can flesh out these implicit possibilities to a fully explicit representation of a conditional:

animal	hidden
¬ animal	hidden
¬ animal	¬ hidden

where each row denotes a different possibility, and the '¬' symbol denotes negation. The second row depicts the situation in which it's not an animal and hidden; the third row depicts the situation in which it's not an animal and not hidden. The three possibilities capture all and only those arrangements of the two events that render the conditional true (Byrne & Handley, 1997). The theory is therefore dual process (see Khemlani & Johnson-Laird, 2013, 2022; and De Neys, 2022 for discussion): a fast system 1 reduces load on working memory by representing only one possibility in memory; a slower system 2 represents fully explicit models.

The theory posits that reasoners make inferences by incrementally building and combining models of premises. Consider this *modus ponens* inference:

6. If it's an animal, then it's hidden.
   It's an animal.
   What, if anything, follows?

Reasoners combine the fully explicit models of an abstract conditional, *if A then B*, with a categorical, *A,* to illustrate the inference in (6):

"If it's animal then	*it's hidden."*	*"It's an animal."*		*The combined model*	
animal	hidden	animal	→	animal	hidden
¬ animal	hidden	animal	→	--	
¬ animal	¬ hidden	animal	→	--	

Null models, denoted by the dashes, are disregarded, and so the resulting model of both premises is:

animal	hidden

The example illustrates how reasoners without any specialized training in formal logic can make *modus ponens* inferences, which are easy for both adults and children (see Braine, Reiser, & Rumain, 1983; Johnson-Laird, Byrne, & Schaeken, 1992; Osherson, 1974-1976). The same principles for building and combining models and generating conclusions account for various phenomena, such as how people draw necessary and possible conclusions from conditionals, conjunctions, and disjunctions (Johnson-Laird & Byrne, 2002; Khemlani et al., 2018) as well as various sorts of quantifier, such as *all* and *most* (Johnson-Laird & Khemlani, 2022; Khemlani & Johnson-Laird, 2022).

Offline knowledge and modulation. The theory posits that individuals can construct models from offline knowledge. Offline knowledge serves to introduce relational dependencies, which can restrict the kinds of models reasoners consider depending on the meaning of clauses in assertions. In the conditional in (4) above, the *if*-clause and the *then*-clause are unrelated to one another. But suppose that the conditional is instead this one:

7. If it's an animal, then it's a wolf.

Reasoners have commonsense knowledge that wolves are a type of animal, and so the information between the *if*-clause and the *then*-clause in (7) *modulates* the default meaning of a conditional statement: it rules out the possibility that it's a wolf and not an animal. Hence, reasoners who consider alternative possibilities may only build the following explicit models:

animal	wolf
¬ animal	¬ wolf

Another consequence of modulation is that some inferences that would otherwise be invalid become valid. Consider the two inferences below:

8a. If it's an animal, then it's hidden.
   It's not an animal.
   Therefore, it's not hidden. [invalid]
 b. If it's an animal, then it's a wolf.
   It's not an animal.
   Therefore, it's not a wolf. [valid]

The first (8a) is invalid because something can be hidden without being an animal, e.g., buried treasure. The second (8b) is valid: wolves are animals, i.e., there's no possibility in which something is a wolf without also being an animal. This modulation effect was first discovered by Johnson-Laird and Byrne (2002), who argued that background knowledge can block the construction of specific possibilities by introducing causal, spatiotemporal, set-theoretic, and other relational dependencies between the clauses of a conditional.

The effects of modulation generalize far beyond conditional reasoning – indeed, it is a widespread phenomenon (Juhos, Quelhas, & Johnson-Laird, 2012; Orenes & Johnson-Laird, 2012; Quelhas & Johnson-Laird, 2017; Quelhas, Rasga, & Johnson-Laird, 2019). And it serves as an acid test of the model theory: no other accounts of reasoning predict or explain it. We return to this point in the General Discussion.

Interactions between online and offline knowledge. How do reasoners take into account online knowledge, i.e., knowledge expressed from descriptions of mental states, when they reason? Consider this inference:

9. If the dog is somebody's pet then it's neutered.
   Riva believes that the dog is somebody's pet.
   What follows?

The second premise in (9) expresses, not a fact about the real world, but rather a belief held by Riva. Reasoners understand that she could potentially hold a false belief, and so it does not necessarily follow that the dog is neutered. The epistemic verb *believe* does not presuppose any facts about the world. Other epistemic verbs do, e.g., *discover, remember*, and *know* – that is, they are factives. Hence, in this variation:

9'. If the dog is somebody's pet then it's neutered.
Riva discovers that the dog is somebody's pet.
What follows?

it follows that the dog is neutered.

We generalize the model theory to work with online knowledge, i.e., presupposed facts that come from expressions of language used to describe mental states, hypotheticals, and other bodies of knowledge that can potentially conflict with the real world, as in (9) and (9') above: reasoners integrate presupposed facts into models of discourse by combining them using the same algorithms used to combine other sorts of model (Johnson-Laird & Khemlani, 2022). For non-factive verbs, such as *think*, *assume, guess*, and *believe,* reasoners keep track of a separate model of an individual's beliefs (see Harner & Khemlani, 2022). Working memory constrains both the number and the detail of the beliefs people keep track of, whereas reasoners can combine presupposed facts with a model of discourse at the time of their presupposition.

This account of how people process expressions of knowledge predicts a novel pattern of reasoning. That is, while the modulation effect described above is widespread, it should not occur in scenarios expressing contingencies instead of presupposed facts. Consider this inference:

10. Loma knows that if it's an animal, then it's a wolf.
It's not an animal.
Does it follow that it's not a wolf?

Reasoners should make this denial of the antecedent (DA) inference and infer that it's not a wolf, because the conditional in (10) expresses a state of knowledge held by Loma. The presupposition is the conditional itself, which is not a fact of the world but rather a set of possibilities. Offline knowledge modulates the conditional by ruling out two possibilities: one in which it's an animal but not a wolf, and another in which it's a wolf but not an animal. This pattern should not hold for unmodulated conditionals, such as *Loma knows that if it's an animal, then it's hidden.*

In contrast, the effect of modulation should disappear for non-factive expressions of belief, as in:

10'. Loma believes that if it's an animal, then it's a wolf.
It's not an animal.
Does it follow that it's not a wolf?

The epistemic verb *believe* is non-factive and makes no presupposition, it should not yield any effect of modulation, and neither should unmodulated conditionals, such as *Loma believes that if it's an animal, then it's hidden*. The following experiment tested and corroborated this interaction.

Experiment

An experiment tested the interaction between online and offline knowledge: for expressions that concern the factive epistemic verb *know,* reasoners should exhibit a modulation effect, i.e., they should be more likely to make inferences such as affirmation of the consequent (AC) and denial of the antecedent (DA) for modulated than unmodulated conditionals. For expressions that concern the epistemic verb *believe,* the theory predicts no difference between modulated and unmodulated conditions.

Participants in the study saw problems such as this one:

Devon knows that if it's cloudy, then it's a warthog.
Devon knows that it's a warthog.
Is it cloudy?

The problem matches the structure of an AC problem, i.e.,

If P, then Q.
Q.
Does it follow that P?

though it embeds the premises in statements that ascribe knowledge to a particular individual. The model theory predicts that people should reject this inference for unmodulated conditionals such as:

...if it's cloudy, then it's a warthog.

but that they should accept the inference for modulated conditionals such as:

...if it's an animal, then it's a warthog.

And it predicts that this effect of modulation should hold for factive verbs (e.g., *knows*) but not for non-factive verbs (e.g., *believes*).

Method

Participants The experiment recruited 60 healthy members of the general North American public through the Cloud Research online platform (29 females, 31 males, 0 other/prefer not to say; mean age = 36.77, age range = 22-62) and compensated them $1.50 for a study that lasted less than 6 minutes; 6 participants were excluded from statistical analysis for failing to meet attention check criteria. 40 out of the remaining 54 participants had received no prior instruction in symbolic logic.

Design and materials Participants completed 12 problems in total. Each problem consisted of three premises: the first premise introduced an agent and an observation (e.g., "[Agent] notices something in the distance"); a second premise stipulated that the agent possessed conditional knowledge linking a state of affairs to an animal (e.g., "[Agent] knows that [if P then Q]"); and a third premise stipulated that the agent possessed categorical knowledge of one of the clauses of the conditional (e.g., "[Agent] knows that [Q]."). Participants then assessed whether a particular conclusion followed from the given premises (e.g., equivalent to, "Is it the case that [P]?") by registering their response on buttons marked "Yes", "No", and "I'm not sure".

The experiment's primary manipulation concerned semantic modulation, that is, whether the conditional premise described an *if*-clause and a *then*-clause that prohibited certain possibilities. The experiment constructed modulated conditionals by using the following *if*-clause: "it is an animal". For example: "...if it is an animal, then it's an ostrich." Unmodulated conditionals concerned a weather condition, e.g., "...if it's cloudy, then it's an ostrich." The materials in each unmodulated problem came randomly drawn from a pool of weather conditions (e.g., "it's cloudy") and the materials in each modulated problem were drawn from a pool of animals (e.g., "ostrich"). As the model theory predicts, the difference between the two is that reasoners' knowledge of various animals suppresses the consideration of any possibility in which it's an ostrich but not an animal (e.g., *not-P and Q*), whereas no such suppression occurs for scenarios in which it's an ostrich but not cloudy.

A secondary manipulation concerned the epistemic verb used to stipulate online knowledge. That is, half the problems concerned an agent's knowledge (e.g., "Devon knows that if...") and the other half concerned an agent's belief (e.g., "Devon believes that if..."). To vary both the presentation of the materials as well as the structures of the problems, the study also manipulated whether the problem structure reflected an affirmation of the consequent (AC) inference or a denial of the antecedent (DA) inference. The two inferences are logically invalid, but compelling (see, e.g., Barrouillet, Gauffroy, & Lecas, 2008; Evans, 1993; Oberauer, 2006; Singmann et al., 2014). The model theory predicts that epistemic verbs should affect the endorsement of inferences, but not problem structure. The experiment randomized the names used for the agents, the materials assigned to the conditions, the order of the problems, and the positions of the response buttons. Two attention check trials were similar in all respects to the 8 other problems in the experiment except that a separate button appeared on the screen for participants to press to indicate that they were paying attention. We excluded participants who missed both attention check trials from subsequent analyses. In addition, two "interpretation" trials were included to verify understanding of each epistemic verb. These trials consisted of an agent in a location and a weather event taking place in a different location. For the 'believes' problem, the prompt asked about the agent's belief about the weather event and for the 'knows' problem it asked about the agent's knowledge of it. These were included to get an understanding of how participants had interpreted each epistemic verb.

Open science The experimental code, materials, data, and statistical analyses are available through the Open Science Framework (https://osf.io/36b9u/), as are preregistrations for all analyses.

Results and discussion

Participants in the study endorsed inferences (e.g., accepted AC or DA inferences) reliably more often for modulated problems than for unmodulated problems (60% vs. 36%, Wilcoxon test, $z = 3.29$, $p < .001$, Cliff's $\delta = .48$), a pattern that corroborates the model theory's central prediction. Participants also did so more often when the epistemic verb was factive rather than non-factive (62% vs. 35%, Wilcoxon test, $z = 4.33$, $p < .001$, Cliff's $\delta = .43$). The structure of the problem, i.e., AC or DA, didn't affect their tendency to endorse inferences (49% vs. 47%, Wilcoxon test, $z = 0.52$, $p = .60$, Cliff's $\delta = .05$).

The results yielded a reliable two-way interaction between modulation and factivity as predicted by the model theory (see Figure 1; Wilcoxon test, $z = 2.66$, $p = .007$, Cliff's $\delta = .27$): participants accepted modulated factives 80% of the time, and they accepted all other problems less than 44% of the time. The results likewise yielded a two-way interaction between the type of verb and the type of problem, i.e., AC vs. DA (Wilcoxon test, $z = 2.42$, $p = .02$, Cliff's $\delta = .19$). It yielded no other significant interactions. Data were subjected to a generalized mixed-model regression (GLMM) to control for participant- and material-wise random effects; the GLMM corroborated nonparametric analyses, i.e., it yielded a main effect of modulation ($B = .82$, $SE = .36$, $p = .02$), a main effect of factivity ($B = 2.37$, $SE = .42$, $p < .001$), and an interaction between modulation and factivity ($B = 1.4$, $SE = .55$, $p = .009$).

In sum, the experiment revealed effects of modulation (an effect of offline background knowledge), of factivity (an

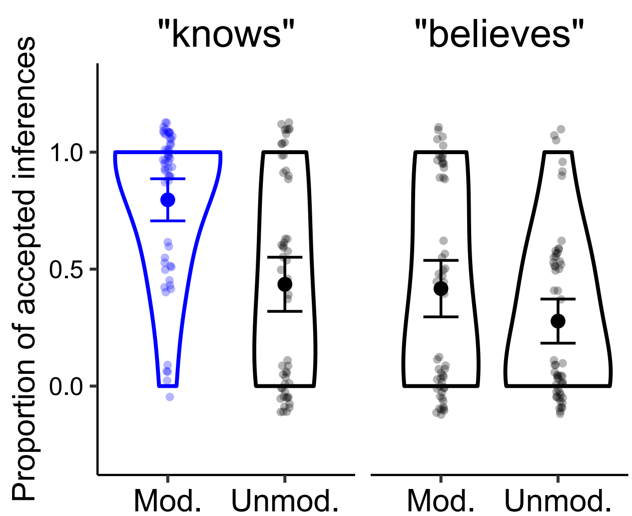

Figure 1. Participants in the experiment carried out problems of the form: *X [knows / believes] that if P then Q; Q is true; Does it follow that P?* The figure shows violin plots of the jittered proportions of accepted AC or DA inferences in the experiment as a function of whether the conditional (*if P then Q*) was modulated, and as a function of whether the epistemic verb was factive ("knows") or not ("believes") in each problem. Participants accepted inferences significantly more than chance only for modulated problems whose epistemic verb was factive (shown in **blue**) and not in any of the other conditions (shown in **black**).

effect of online knowledge ascription), and of the interaction between the two. The effects of modulation corroborate a central prediction of the model theory (Johnson-Laird & Byrne, 2002; Quelhas et al., 2019): they show that people do not reason based on the logical structure of premises, but rather on their meanings as embodied in models of possibilities. Meanings can rule out certain possibilities and introduce relations between entities (Juhos et al., 2012). Epistemic relations, i.e., relations between an agent's mental state and the facts of the matter, can interact with the effects of modulation.

General discussion

We show that reasoners integrate offline and online knowledge when they reason. Offline knowledge refers to conceptual structures in declarative, semantic, and episodic memory (see, e.g., Moscovitch, & Rugg, 2019), and many studies reveal its effects on the inferences people make (Dube, Rotello, & Heit, 2010; Evans et al., 1993; Johnson-Laird, 1999). Studies also reveal that people take into account online knowledge, i.e., the knowledge attributed to individuals in descriptions of mental states. To say that "Howard knows that pterodactyls are not dinosaurs" presupposes that Howard's knowledge of the world is correct, and from an early age, humans learn to distinguish factive verbs such as *know* from non-factive verbs such as *believe*. But no research has demonstrated how these two forms of knowledge interact. Contemporary frameworks of reasoning concur that knowledge affects the inferences people make, though disagreements exist about the representations that give rise to effects of content and context (see Khemlani, 2018, for a review).

To explore the matter, we ran a study that pitted considerations of offline knowledge against those of online knowledge. We presented participants with problems of the following form:

Devon knows that if it's an animal, then it's a warthog.
Devon knows that it's a warthog.
Is it an animal?

The conclusion is sensible: people should respond "yes". They do so by integrating offline knowledge (the fact that warthogs are a type of animal) and online knowledge (the fact that Devon's factive mental state implies that it is, indeed, a warthog). The model theory predicts that reasoners cannot construct any possibility in which it's a warthog but not an animal, i.e., every model they construct should reflect that it's both a warthog and an animal. Hence, reasoners should accept the conclusion above. The study we ran corroborates this effect, and it also shows that unless both of these conditions are in place, reasoners are likely to reject AC and DA conclusions more often than not.

Do other accounts of reasoning explain the effects of modulation and factivity observed here and elsewhere (e.g., Quelhas, Rasga, & Johnson-Laird, 2019)? No – as far as we know. Many theories of reasoning explore how people update prior beliefs in light of new knowledge (Oaksford & Chater, 2013), and some even appeal to the idea that reasoners construct mental simulations of the world to reason (Evans & Over, 2013), and so the effects of modulation would seem a reasonable test of such theories. To explain modulation effects, however, alternative accounts of reasoning need to show how people keep track of multiple possibilities and eliminate them in light of semantic constraints. No such theory provides such explanations – perhaps because probabilistic theories are often developed at the "computational" level, i.e., they concern the mathematical functions people compute and not the mental processes by which they compute them (see Knauff & Gazzo-Castañeda, in press). New alternatives to the model theory can be developed, of course, but an alternative account that explains modulation in terms of how reasoners eliminate possibilities may be indistinguishable from the mechanisms we described above. Hence, at present, only the model theory predicts or explains the effects of modulation.

Acknowledgments

This work was supported by an NRC Research Associateship Award to BB and funding from the Naval Research Laboratory to SK. We also thank Kalyan Gupta, Danielle Paterno, and the Knexus Research Corporation for their help in data collection, and we are indebted to Bill Adams, Gordon Briggs, Magda Bugajska, Tony Harrison, Andrew Lovett, Maria Kon, and Greg Trafton for their critiques and comments.

References

Abbeduto, L., & Rosenberg, S. (1985). Children's knowledge of the presuppositions of know and other cognitive verbs. *Journal of Child Language*, 12(3), 621-641.

Adrián, J. E., Clemente, R. A., & Villanueva, L. (2007). Mothers' use of cognitive state verbs in picture-book reading and the development of children's understanding of mind: A longitudinal study. *Child Development*, 78(4), 1052-1067.

Barrouillet, P., Gauffroy, C., & Lecas, J.-F. (2008). Mental models and the suppositional account of conditionals. *Psychological Review,* 115, 760—772.

Booth, J. R., & Hall, W. S. (1995). Development of the understanding of the polysemous meanings of the mental-state verb know. *Cognitive Development*, 10(4), 529-549.

Braine, M. D., Reiser, B. J., & Rumain, B. (1984). Some empirical justification for a theory of natural propositional logic. In *Psychology of Learning and Motivation* (Vol. 18, pp. 313-371). Academic Press.

Byrne, R. M., & Handley, S. J. (1997). Reasoning strategies for suppositional deductions. *Cognition*, 62(1), 1-49.

Byrne, R. M. J. (2005). *The rational imagination.* Cambridge, MA: MIT Press.

Byrne, R. M. J. (2016). Counterfactual thought. *Annual Review of Psychology*, 67, 135–157.

Carey, S., Leahy, B., Redshaw, J., & Suddendorf, T. (2020). Could it be so? The cognitive science of possibility. *Trends in Cognitive Sciences*, 24(1), 3-4.

Cohen, L. J. (1992). *An essay on belief and acceptance*. Oxford University Press.

Craik, K. J. W. (1943). *The nature of explanation.* University Press, Macmillan.

Dasgupta, I., et al. (under review). Language models show human-like content effects on reasoning. Manuscript under review.
De Neys, W. (2022). Advancing theorizing about fast-and-slow thinking. *Behavioral and Brain Sciences*, 1-68.
Dube, C., Rotello, C., & Heit, H. (2010). Assessing the belief bias effect with ROCs: It's a response bias effect. *Psychological Review,* 117, 831-863.
Dudley, R., Rowe, M., Hacquard, V., & Lidz, J. (2017). Discovering the factivity of "know". In *Semantics and Linguistic Theory* (Vol. 27, pp. 600-619).
Espino, O., Byrne, R. M., & Johnson-Laird, P. N. (2020). Possibilities and the parallel meanings of factual and counterfactual conditionals. *Memory & Cognition*, 48(7).
Evans, J. St. B. T. (1993). The mental model theory of conditional reasoning: Critical appraisal and revision. *Cognition*, 48, 1-20.
Evans, J. St. B. T., Newstead, S., & Byrne, R.M.J. (1993). *Human reasoning: The psychology of deduction.* UK: Lawrence Earlbaum Associates.
Evans, J. S. B., & Over, D. E. (2013). *Rationality and reasoning.* Psychology Press.
Fetzer, A., & Johansson, M. (2010). Cognitive verbs in context: A contrastive analysis of English and French argumentative discourse. *International Journal of Corpus Linguistics,* 15.
Forrester, J. (2017). *Thinking in cases*. John Wiley & Sons.
Goldvarg, E., & Johnson-Laird, P. N. (2001). Naive causality: A mental model theory of causal meaning and reasoning. *Cognitive Science*, 25(4), 565-610.
Harner, H., & Khemlani, S. (2022). Reasoning about *want*. *Cognitive Science,* 46.
Hazlett, A. (2012). Higher-order epistemic attitudes and intellectual humility. *Episteme*, 9, 205-223.
Henne, P., Niemi, L., Pinillos, A., De Brigard, F., & Knobe, J. (2019). A counterfactual explanation for the action effect in causal judgment. *Cognition, 190,* 157–164.
Henne, P., O'Neill, K., Bello, P., Khemlani, S., & De Brigard, F. (2021). Norms affect prospective causal judgments. *Cognitive Science, 44,* e12931.
Hinterecker, T., Knauff, M., & Johnson-Laird, P. N. (2016). Modality, probability, and mental models. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 42.
Hughes, G. E., & Cresswell, M. J. (1996). *A new introduction to modal logic*. London: Routledge.
Johnson-Laird, P. N., & Byrne, R. M. (2002). Conditionals: a theory of meaning, pragmatics, and inference. *Psychological Review*, 109.
Johnson-Laird, P. N., & Khemlani, S. (2022). Mental models and the algorithms of deduction. In R. Sun (Ed.) *Cambridge Handbook of Computational Cognitive Sciences.*
Johnson-Laird, P. N., & Ragni, M. (2019). Possibilities as the foundation of reasoning. *Cognition*, 193, 103950.
Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness*. Harvard University Press.
Johnson-Laird, P. N. (2006). *How we reason.* Oxford University Press: Oxford, UK.
Johnson-Laird, P. N. (1999). Deductive reasoning. Annual review of psychology, 50, 109-135.
Johnson-Laird, P. N., Byrne, R. M., & Schaeken, W. (1992). Propositional reasoning by model. *Psychological Review*, 99.
Johnson-Laird, P. N., Khemlani, S. S., & Goodwin, G. P. (2015). Logic, probability, and human reasoning. *Trends in Cognitive Sciences*, 19, 201-214.
Juhos, C., Quelhas, A. C., & Johnson-Laird, P. N. (2012). Temporal and spatial relations in sentential reasoning. *Cognition*, 122.
Kelly, L. J., Khemlani, S., & Johnson-Laird, P. N. (2020). Reasoning about durations. *Journal of Cognitive Neuroscience*, 32, 2103-2116.
Khemlani, S., & Johnson-Laird, P. N. (2013). The processes of inference. *Argument & Computation*, 4, 4-20.
Khemlani, S. S., & Johnson-Laird, P. N. (2017). Illusions in reasoning. *Minds and Machines*, 27, 11-35.
Khemlani, S., & Johnson-Laird, P.N. (2022). Reasoning about properties: A computational theory. *Psychological Review,* 129.
Khemlani, S., Bello, P., Briggs, G., Harner, H., & Wasylyshyn, C. (2021). Much ado about nothing: The mental representation of omissive relations. *Frontiers in Psychology*, 11, 609658.
Khemlani, S. S. (2018). Reasoning. In S. Thompson-Schill (Ed.), *Stevens' Handbook of Experimental Psychology and Cognitive Neuroscience*. Wiley & Sons.
Khemlani, S. 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. 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.
Khemlani, S., Orenes, I., & Johnson-Laird, P. N. (2012). Negation: A theory of its meaning, representation, and use. *Journal of Cognitive Psychology*, 24, 541-559.
Kiparsky, P., & Kiparsky, C. (1971). Fact. In L. Jakobovits and D. Steinberg (Eds.), *Semantics: An Interdisciplinary Reader*. Cambridge: Cambridge University Press.
Knauff, M. (2013). *Space to reason: A spatial theory of human thought.* MIT Press.
Knauff, M., & Gazzo Castaneda, L. (in press). When nomenclature matters: Is the "new paradigm" really a new paradigm for the psychology of reasoning?. *Thinking & Reasoning,* in press.
Kumar, A. (2021). Semantic memory: A review of methods, models, and current challenges. *Psychonomic Bulletin & Review,* 28.
Lewis, S., Hacquard, V., & Lidz, J. (2017). "Think" pragmatically: Children's interpretation of belief reports. *Language Learning and Development*, 13(4), 395-417.
MacWhinney, B. (2000). The CHILDES project: The database (Vol. 2). Psychology Press.
McKinsey, M. (1999). The semantics of belief ascriptions. *Noûs, 33*.
Moore, R. (2013). Social realism and the problem of the problem of knowledge in the sociology of education. *British Journal of Sociology of Education*, 34, 333-353.
Morgan, J., & Morton, J. (1944). The distortion of syllogistic reasoning produced by personal convictions. *Journal of Social Psychology*, 20, 39-59.
Morris, A., Phillips, J., Gerstenberg, T., & Cushman, F. (2019). Quantitative causal selection patterns in token causation. *PLoS One*, 14, e0219704.
Renoult, L., Irish, M., Moscovitch, M., & Rugg, M. D. (2019). From knowing to remembering: the semantic–episodic distinction. *Trends in Cognitive Sciences*, 23(12), 1041-1057.
Oaksford, M., & Chater, N. (2013). Dynamic inference and everyday conditional reasoning in the new paradigm. *Thinking & Reasoning,* 19.
Oberauer, K. (2006). Reasoning with conditionals: A test of formal models of four theories. *Cognitive Psychology*, 53, 238–283.
Orenes, I., & Johnson-Laird, P. N. (2012). Logic, models, and paradoxical inferences. *Mind & Language*, 27(4), 357-377.
Osherson, D. N. 1974–1976. *Logical Abilities in Children*, Vols 1–4.

Partee, B. H. (1979). Semantics – mathematics or psychology? In Bäuerle, R., Egli, U., and von Stechow, A. (Eds.) *Semantics from different points of view*. Berlin: Springer-Verlag.

Peirce, C. (1955). 1931-1958. Collected papers. 8 vols. Harvard University, Cambridge, Mass.

Perner, J., & Roessler, J. (2012). From infants' to children's appreciation of belief. Trends in Cognitive Sciences, 16.

Phillips, J., & Cushman, F. (2017). Morality constrains the default representation of what is possible. *Proceedings of the National Academy of Sciences,* 114, 4649-4654.

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

Portner, P. (2009). *Modality.* Oxford University Press: Oxford.

Quelhas, A. C., & Johnson-Laird, P. N. (2017). The modulation of disjunctive assertions. *Quarterly Journal of Experimental Psychology*, 70(4), 703-717.

Quelhas, A. C., Rasga, C., & Johnson-Laird, P. N. (2019). The analytic truth and falsity of disjunctions. *Cognitive Science*, 43.

Radvansky, G. A., & Zacks, J. M. (2014). *Event cognition*. Oxford University Press.

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

Renoult, L., Irish, M., Moscovitch, M., & Rugg, M. (2019). From knowing to remembering: The semantic-episodic distinction. *Trends in Cognitive Sciences,* 23, 1041-1057.

Richard, M. (1983). Direct reference and ascriptions of belief. *Journal of Philosophical Logic, 12,* 425-452.

Singmann, H., Klauer, K. C., & Over, D. (2014). New normative standards of conditional reasoning and the dual-source model. *Frontiers in Psychology*, *5*, 316.

Schwanenflugel, P. J., Henderson, R. L., & Fabricius, W. V. (1998). Developing organization of mental verbs and theory of mind in middle childhood: evidence from extensions. *Developmental Psychology*, 34.

Schtulman, A., & Tong, L. (2013). Cognitive parallels between moral judgment and modal judgment. *Psychonomic Bulletin & Review,* 20, 1327-1335.

Squire, L. R. (2004). Memory systems of the brain: a brief history and current perspective. *Neurobiology of Learning and Memory*, 82.

Stalnaker, R. C. (1999). *Context and content: Essays on intentionality in speech and thought.* Clarendon Press.

Trickett, S.B., & Trafton, J.G. (2007). "What if…": The use of conceptual simulations in scientific reasoning. *Cognitive Science, 31,* 843-875.

Wason, P. C., & Johnson-Laird, P. N. (1972). *Psychology of reasoning: Structure and content* (Vol. 86). Harvard University Press.

Wason, P. C., & Shapiro, D. (1971). Natural and contrived experience in a reasoning problem. *Quarterly Journal of Experimental Psychology*, 23, 63-71.

Wilkins, M. C. (1929). The effect of changed material on ability to do formal syllogistic reasoning. *Archives of Psychology*, 102, 83.