Decomposing modal thought

Jonathan Phillips[1,3] and Angelika Kratzer[2,3]

[1]Dartmouth College, Hanover NH
[2]University of Massachusetts Amherst, MA
[3]Centre for Advanced Studies in the Humanities “Human Abilities”, Berlin

Abstract

Cognitive scientists have become increasingly interested in understanding how natural minds represent and reason about possible ways the world may be. However, there is currently little agreement on how to understand this remarkable capacity for ‘modal thought’. Drawing on formal frameworks for reasoning about possibilities from logic, philosophy, computer science, and linguistics, we argue that this capacity is built from a set of relatively simple component parts, centrally involving a basic ability to consider possible extensions of a piece of the actual world. Natural minds can productively combine this basic ability with a range of other capacities, eventually allowing for the observed suite of increasingly more sophisticated ways of reasoning about what is possible. We demonstrate how this (de)compositional account can accurately predict both what has been observed in the trajectory of children’s developing capacity to reason about possibilities and what has been observed in how modal thought is expressed within and across natural languages. Our hope is that this framework will provide cognitive scientists with a more systematic way of understanding variation in actuality-directed modal thought and talk, which will serve as the beginnings of a common language that allows researchers across disciplines to better understand each other.

Keywords: modality; possibility; anchor semantics; counterfactual reasoning; modal cognition; metacognition

Decomposing modal thought

Across the Cognitive Sciences, there's a newfound interest in studying modal cognition, our ability to represent possible ways the world we live in may be—whether it be representing what may happen in the future, reasoning about what might have happened in the past, figuring out what could have happened if things had been different, or even just guessing what else might be the case right now in parts of the world beyond our reach. Modal thought in this sense is central to much of high-level human thought, including for example, causal and counterfactual reasoning (Lewis, 1973, 1974; Pearl, 2009; Gerstenberg, Goodman, Lagnado, & Tenenbaum, 2021), or planning and decision making (Kaelbling, Littman, & Cassandra, 1998; Morris, Phillips, Huang, Cushman, 2021). Historically, work under the label of 'modality' has been the pursuit of philosophers, linguists, logicians, or computer scientists. But these days, an increasing number of researchers are studying how natural minds actually represent and reason about what is possible—from non-human animals (e.g., Engelmann, Völter, O'Madagain, Proft, Haun, Rakoczy, & Herrmann, 2021; Redshaw, & Suddendorf, 2016), to human infants (e.g., Téglás, Girotto, Gonzalez, & Bonatti, 2007) and young children (e.g., McCormack & Hoerl, 2020; Shtulman & Carey, 2007), to adult humans (e.g., Byrne, 2007; De Brigard, Addis, Ford, Schacter, & Giovanello, 2013; Phillips & Cushman, 2017).

Modal thought, in the sense understood here and in the literature on modal cognition, is directly reflected in modal expressions like the English auxiliaries *must*, *might*, or *can*, or modal adjectives like *fragile* or *breakable*. Modal thought thus relates to possibilities, but we want to emphasize at the outset that our account does not aim at accounting for all thought relating to possibilities. To illustrate, take a sentence like *there is a cat on my lap*. You know its meaning if you know which possible situations would make it true: not just some actual situation you may be talking about that happens to have a cat curled up in your lap, but any possible situation that has a cat on your lap. Understanding the meaning of even the simplest sentence, then, relies on our capacity to connect sentences to possibilities. Yet simply understanding this sentence is not an instance of the kind of modal thought targeted here. Or, to take another example, pigeons can famously be trained to associate pecking at a target with the subsequent production of a food reward, and with enough experience, can further learn to that this association only holds under certain conditions and not others (Skinner, 1965). This kind of reinforcement learning involves pairing a current action with an as-of-yet non-actual event—the future emergence of food—and thus relies on a capacity to relate actions to future possibilities. Yet simply coming to learn associations through repeated exposure is again not a phenomenon we are interested in. But then, what is the sort of modal thought we are concerned with?

The notion of modal thought we are interested in is actuality-directed: it involves taking some piece of actuality and entertaining possible extensions of it. We can point to simple but clear examples of this capacity: you may wonder what could fit into an empty box when packing for a move; you may see clouds gathering on a cold day and guess that snow will be on your doorstep by tomorrow morning; or you may examine a puddle on the street and determine that it must have rained. In each of these cases, you take some part of the actual world (the *modal anchor*, to adopt terminology of Hacquard, 2006) such as the empty box, the actual clouds, or the puddle, and then consider possible extensions of that anchor situation. A possible future extension of the box situation may have a pot packed in the box. A possible future extension of the cloud situation may include snow on your doorstep. And a possible past extension of the puddle situation may feature a downpour.

We think this basic capacity of considering possible extensions of an actual anchor situation is a common component of many apparently different types of modal cognition. Some uses of this capacity will seem quite humble. However, the basic capacity of considering possible extensions of an actual anchor situation can be refined further or productively combined with other capacities. This can yield more sophisticated abilities such as the capacity to consider possibilities that were still live options at some point in the past, but that we now know to be only counterfactually possible, the capacity to compare, rank, weight or quantify over possibilities, or the capacity to design an optimal plan of action in light of a range of possibilities. It is these more complex capacities that ultimately give rise to the sophisticated range of modal reasoning found in adult humans.

In the empirical work thus far, there has largely been agreement on the kind of phenomena of interest when studying 'modal cognition', but there has also been much less agreement on what modal cognition is, or how it works (Bell & Johnson-Laird, 1988; Carey, Leahy, Redshaw, & Suddendorf, 2020; Cesana-Arlotti, Téglás, & Bonatti, 2012; Engelmann, et al., 2021; Johnson-Laird & Ragni, 2019; Leahy & Carey, 2020; Phillips, Morris, & Cushman, 2019; Redshaw & Suddendorf, 2020; Téglás & Bonatti, 2016). Our aim in this paper is to contribute to more clarity about those questions. Unifying insights across linguistics, philosophy, logic, and computer science, we argue that what has been discussed under the label 'modal cognition' in the literature is actuality-directed modal thought that can be built from a set of relatively simple component abilities, all involving the basic ability of considering possible extensions of an actual anchor. Our hope for this work is twofold. First, this (de)compositional approach should offer researchers a more systematic way of thinking about variation in observed instances of actuality-directed modal thought—whether it be variation across species, within human development, or in natural language. Second, we aim to lay the foundation for a common language for researchers studying modality across disciplines: allowing, for example, developmental psychologists to make claims more easily understood and modeled by logicians, or comparative cognition researchers to test ideas originally formulated by philosophers and semanticists.

Simple worked examples

Going back to the empty box example, imagine again that you are deliberating on what to pack in the box in view of an upcoming move. Such an ability can be decomposed into a set of component processes (see Fig 1). You begin with the actual empty box. This serves as the *modal anchor*, that is, a part of the actual world, a situation. We can then consider possible four-dimensional extensions of the anchor situation: situations that include matches of the empty box at the present time, but now extend into possible futures where various things go into the box. You might, for example consider a future extension of the anchor situation where books are put in the box, or one where it is a ceramic pot, or perhaps even one where it is a pair of shoes. The process of considering possible extensions of an actual modal anchor can be understood as an ability for *factual domain projection*, that is, a process that implements a function, f_{act}, which takes a part of the actual world, the anchor, and returns possible situations containing matches of the anchor (Kratzer, 2013).[1] The result of factual domain projection is a set of possible situations bound to the actual world by matches of the anchor at its current time. Collectively, these possibilities make up the *modal domain*, which one could go on to reason about by comparing, ranking, weighting, or quantifying over the possibilities in it. In our example, you might go on to reason that, while you could pack a book in that box, the best thing to pack would actually be the ceramic pot.

[1] Of course, there are many ways that such a function may be realized at an algorithmic level, and even more at the level of physical implementation; our proposal is not an account at these levels of analysis (Marr, 1982).

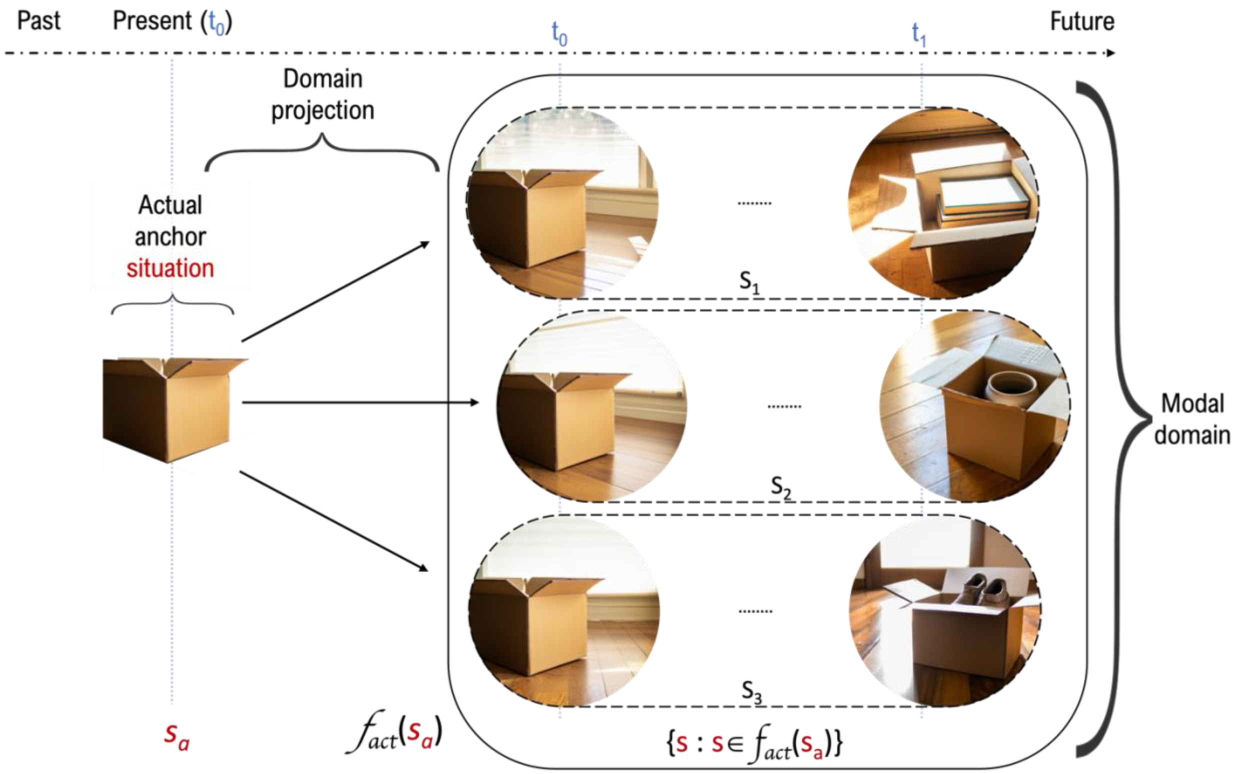

Figure 1. Schematic depiction of the core processes of future-oriented modal thought when considering what to pack inside a box. A domain projection function f_{act} takes as input the modal anchor s_a (the situation in the actual world involving the box you are packing) and returns a set of situations, $\{s: s \in f_{act}(s_a)\}$, with matches of the anchor at the present time but different extensions into the future. Among these are situations in which you pack books, a ceramic pot, or shoes in the box, s_1-s_3, respectively.[2]

To have another example, imagine again that you are looking at a puddle of water on the ground and are wondering where it came from (Fig. 2). The process you use is much the same as in the case with the box: you begin with the actual puddle at the current time as your anchor, and then generate possible past extensions of that puddle-situation, this time including earlier events leading to that puddle: a storm passing through, a bucket of water being spilled, or a fire hydrant being tested. The puddle example differs in an important way from the box example. While the anchor is a part of the actual world at the present time in both, the possible extensions considered stretch forward into the future in the case of the box, but backwards into the past in the case of the puddle. That is, the two examples illustrate a change in the *temporal orientation* of the extensions of the anchor situation that are being considered (Condoravdi, 2002). The basic ability to consider possible extensions of an actual anchor situation can thus be refined by distinguishing extensions according to their temporal orientation. They may stretch forward into the future or backwards into the past, and they could also stay within the present, with matches of the anchor situation appearing in varying possible spatial surroundings.

[2] The images in these schematic depictions of factual domain projection were all generated by JP using the new outpainting feature in DALL·E 2. It was a fortuitous coincidence that OpenAI developed their artificial intelligence tool for generating extensions of an input image at the same time we were developing our theory of how natural intelligence generates extensions of an anchor situation.

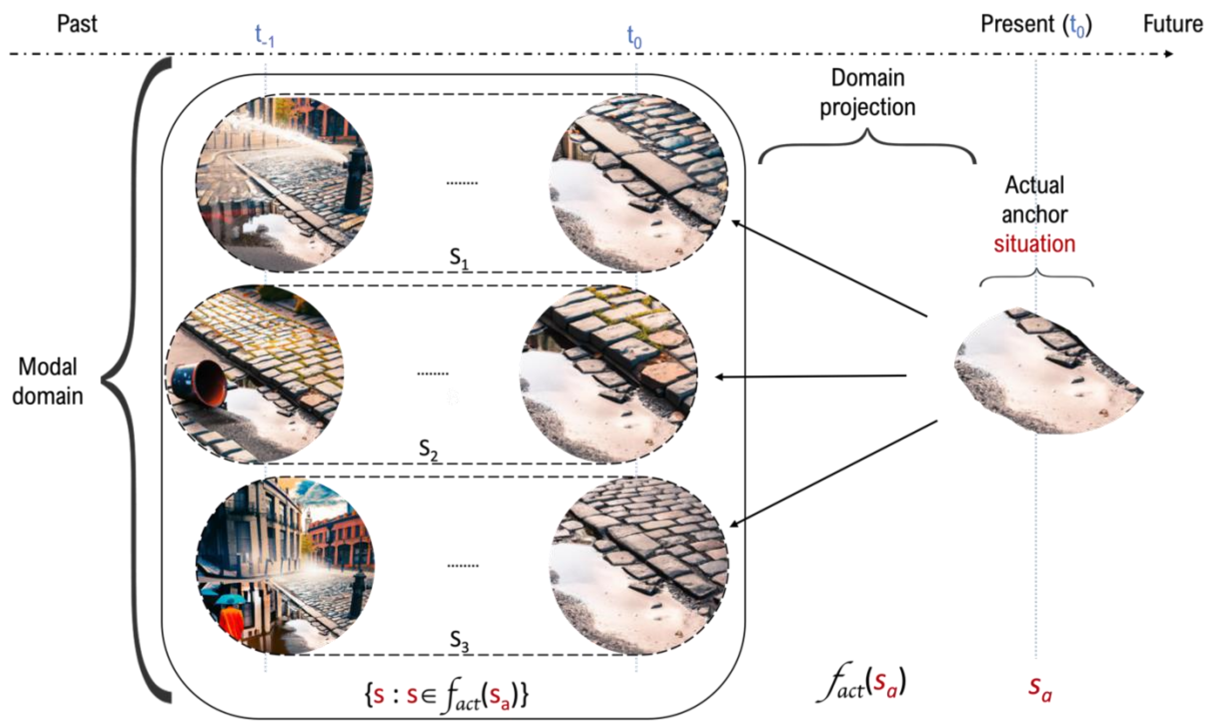

Figure 2. Schematic representation of the core processes of modal cognition when reasoning about the causes of a puddle on the ground. A domain projection function, f_{act}, takes as input the modal anchor s_a (the situation in the actual world involving the puddle you are looking at) and returns a set of past situations, $\{s: s \in f_{act}(s_a)\}$, which have matches of the anchor situation at the present time, but different past extensions. Among these are situations in which a fire hydrant, a bucket of water, or a rainstorm caused the puddle, s_1-s_3, respectively.

Finding evidence for the pieces of modal thought

The central question we now face is what kind of evidence there is for this way of thinking about modal cognition. After all, there are alternative accounts, which differ substantively from the version we have sketched (e.g., Leahy & Carey, 2020; Johnson-Laird & Ragni, 2019; Redshaw, & Suddendorf, 2020). One way to get traction on this question is to point out that our account makes predictions about developmental paths and cross-species variation. On our proposal, we expect to find more complex manifestations of modal cognition that are the result of combining several abilities but are all built on the foundation of factual domain projection. Many observed manifestations of modal thought should thus be analyzable as the result of a combinatorics of basic abilities, and these same combinatorics should be detectable in patterns of variation across phylogeny and ontogeny. For example, the process of factual domain projection would be expected to exhibit variation that corresponds to differences in the choice of modal anchors and the temporal orientation of projected possibilities. Likewise, when examining how modality is encoded in natural languages, we might expect to find evidence that the combinatorics of abilities we posit for actuality-directed modal thought are transparently reflected in the way the meanings of sentences expressing modal thoughts are compositionally constructed. In the sections to follow, we will provide first evidence for our proposed decomposition, first from cognitive development, and then from natural language. In three cases, we also provide empirical tests of some of the more novel predictions of our proposal.

Evidence from Cognitive Development

Future-oriented domain projection

Evidence for the ability to consider possible futures for present actual anchors can be found as early as twelve months. In a series of experiments by Téglás and colleagues (2007), infants watched as different objects bounced around a circular container that had a single opening at the bottom. Three of the objects were yellow, and one was blue. After infants watched the objects move around the container, an occluder covered the container, and a single object exited from its bottom. Twelve-month-old infants exhibited longer looking times when a blue rather than yellow objected exited, providing some initial evidence that they were representing future possibilities with yellow objects exiting the container. Critically, in a second study, the three yellow objects were blocked from exiting the container by an additional horizontal wall, and only the blue object could physically exit the container (see Fig. 3a). In this case, infants' looking time indicated the opposite pattern: they were more surprised when a yellow, rather than a blue, object exited, suggesting that they were no longer considering future possibilities where the yellow objects exited (Téglás et al., 2007).

These two studies serve to illustrate what has now been found across a range of studies: at a surprisingly early age, human children are able to project future possibilities from present anchor situations (e.g., Cesana-Arlotti, et al., 2012; Téglás & Bonatti, 2016). We next turn to examples that, from our perspective, illustrate how the more basic capacity of generating future possibilities from a present anchor can combine with other capacities to give rise to more complex capacities whose mastery shows a substantial degree of variation across phylogeny and ontogeny.

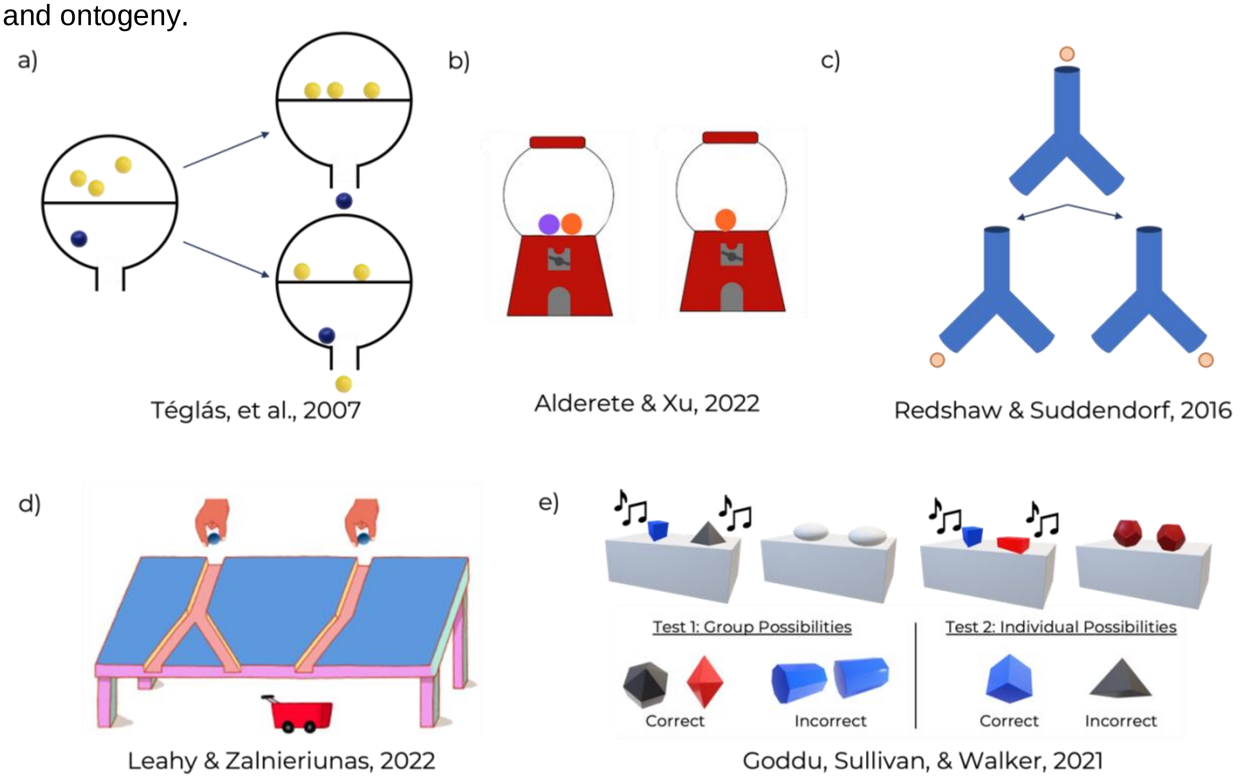

Figure 3: Illustrations of the experimental materials of five paradigms used to study the development of modal thought.

Making possible events actual

The basic ability to generate future possibilities from a present anchor can be recruited in decision making tasks. Those tasks require future-oriented factual domain projection, the ability to compare possibilities according to their desirability, and, finally, the ability to use the results of the comparison as the basis for choosing the most desirable action. A recent study suggests this more complex suite of abilities is present by two and a half years of age, if not earlier. Alderete and Xu (2022), presented children with two gumball machines, one that had a single gumball in it, and another that had two differently colored gumballs, as shown in Figure 3b. What they found is that when seeking to get an orange gumball, for example, young children would choose the machine with a single orange gumball and avoid the machine with both a purple and an orange gumball. This suggests that the children were aware that choosing the machine with two gumballs might lead to a situation where they receive the undesirable purple ball. By at least the age of two and half, then, children do not only have the ability to represent possible future courses of events. They can also represent and evaluate the possible outcomes of different possible future actions and consistently select the preferred option.

Selecting actions conditional on possible future events

The results of the Alderete and Xu (2022) study stands in stark contrast with several studies that feature what looks like a closely related task—the ‘forked-tube task’ (see Fig. 3c)—which children struggle with until much later (Beck, Robinson, Carroll, & Apperly, 2006; Leahy & Carey, 2020; Redshaw, & Suddendorf, 2020). In the forked-tube task, an object is dropped into a tube shaped like an upside-down Y, and subjects are incentivized to catch the object as it exits from the bottom of the forked tube (Redshaw & Suddendorf, 2016). Researchers have observed a remarkably robust pattern: non-human primates and young children (before the age of 4) fail to systematically cover both branches of the forked tube, instead alternating between covering one or the other of the two branching tubes (Redshaw & Suddendorf, 2016; Redshaw, Suddendorf, Neldner, Wilks, Tomaselli, Mushin, & Nielsen, 2019; Suddendorf, Crimston, & Redshaw, 2017).

At first glance this task appears quite similar to the gumball machine task. In both cases, children have a choice between an action that guarantees success and an action (or actions) that only has (or have) a 50% chance of success. In both cases, they have to represent possible future courses of events, evaluate them for desirability, and choose an action that leads to the most desirable outcome. But then given these similarities, why do children before the age of four robustly fail the forked-tube task while passing the gumball task with ease?

An important difference is that in the gumball task children are in control of which of the two possible future courses of events will end up being actual: whether or not they can be sure to receive the preferred gumball predictably depends on which action they take, and they are free to choose their action. The forked-tube task, in contrast, seems more demanding. Children are not in control of the two possible future courses of events they are presented with. Success now requires that they come up with an action that is optimal regardless of which of the two possible future courses of events will become actual. That is, they have to think of an action (or set of actions) that would allow them to catch the object regardless of which trajectory it may take. Covering the right side or the left side of the tube only sometimes results in catching the object, but only covering both guarantees success.

If this is the right way of thinking about why the forked-tube task is difficult, we can conclude (contra Leahy & Carey, 2020) that it’s not because young children have difficulty generating multiple incompatible future extensions of the falling object. We think they understand that the falling object may both go left or go right, but they fail to see that covering both sides simultaneously guarantees success, regardless of which trajectory becomes actual. Thus, the pattern observed in the forked-tube task in no way challenges the results of Téglás et al. (2007), and both involve future-oriented factual domain projection; it's just that the forked-tube task demands additional abilities as well.

Decomposing the abilities required by the forked-tube task in the way we have makes clear predictions about which kinds of tasks should be difficult and which should not be. For example, if young children are simply watching the object be dropped into the forked tube, it predicts (unlike other accounts) that they should not be surprised if the ball exits out of the right or left side of the fork, indicating that they are representing both possibilities. Further, if the difficulty does arise from having to select an action that succeeds regardless of which of the two uncontrollable future possibilities will turn out to be actual, then even relatively small modifications to the gumball task (Alderete and Xu, 2022) should make it equally difficult for children to pass. Suppose, for example, that the experimenter in the gumball task explains that they will turn the handle on both gumball machines at once and the child's job is to simply catch an orange gumball as it exits. Young children should now be expected to struggle with this task in much the same way that they struggle with the forked-tube task. Success once again would demand that children select an action (or set of actions) that will succeed regardless of which one of two uncontrollable future possibilities will become actual.

Helpfully, an experiment that is structurally identical to the proposed modification of the gumball task has recently been conducted by Leahy and Zalnieriunas (2022). In this task (see Fig. 3d), children are shown two slides, one that has a single exit, and one that is shaped like an upside-down Y and thus has two exits. Two round objects are placed at the top of the two slides and will be dropped by the experimenter at the same time. Children's task is to place a cart underneath one of the three exits to catch one of the falling objects. What Leahy and Zalnieriunas found is that children younger than 4 place the cart under the single-exit slide only about 50% of the time. That is, in contrast to the gumball task, which children much younger than 4 pass with ease, they fail the seemingly similar slides task. The essential difference is that passing the slides task, like passing the forked-tube task, requires figuring out that some particular action maximizes success regardless of which of multiple uncontrollable future courses of events will be actualized.

To sum up, rather than thinking that children's failure on the forked-tubes task reveals a broad inability for modal thought, as argued by Leahy and Carey (2020) and Redshaw and Suddendorf (2020), we think it would be a mistake to draw conclusions from the results of this task in isolation. This task must be evaluated alongside other tasks that involve modal cognition, such as the Téglás et al. task and Alderete and Xu's gumball task. The project then becomes to explain what the broader set of success and failures suggests about which component parts of modal cognition are or are not present at a particular age or in a particular species.

A subsequent joint in modal thought: Epistemic anchors

A major consequence of our assumption that modal thought in its various instantiations relies on factual domain projection is that it predicts a critical hurdle that needs to be overcome before children master certain more difficult modal reasoning tasks around age 6. In all of the examples we have discussed so far, subjects engaged in modal reasoning could choose anchors from their environment: situations with empty boxes, puddles, clouds, bouncing balls, gumball machines, and so on. Our account of modal thought in terms of factual domain projection predicts that there are certain kinds of scenarios where factual domain projection fails unless subjects take their own epistemic state as modal anchors. In what follows, we will first discuss an example of such a scenario and then move on to show that modal reasoning tasks that children do not master before around age six involve precisely those kinds of scenarios.

Imagine you are climbing up a hill and see someone standing near a grove of trees in the distance. You can see that it's a man wearing a Shaker-style straw hat, but you can't yet clearly discern his face. Remembering that your friends Matt and Raphael both bought that style of hat recently, you say to yourself: the man might be Matt, but it might also be Raphael. Unbeknownst to you, it happens to be Raphael. When we try to analyze your modal thought in the way we did before, we run into a serious problem. What actual situation should we take to be the modal anchor? Suppose we took the salient actual situation you are looking at to be the anchor. That happens to be a situation involving Raphael. But then, in every possible extension of that situation that person will, of course, remain Raphael. There would be no possibility in the projected modal domain where that person could somehow turn into Matt. But if that's right, then our account would seem to predict that you made a serious mistake when you thought that the man you saw might be Matt. That's absurd though. You did not make a mistake—for all you knew, the man in the Shaker-style hat *might* have been Matt! We need to find some way out of this dilemma.

The dilemma goes away as soon as we consider the possibility that there is another plausible choice of anchor for your modal thought. If modal anchors are contextually salient parts of the actual world, *you in your actual epistemic state* as you are looking at the man in the distance should also be an admissible anchor.[3] Possible extensions of that situation have you in your current epistemic state looking at a man standing near a grove of trees, and in at least some of those extensions, the man you are looking at is Matt. That is, when your perception of the man and what you know more generally about your two friends is compatible with the figure being Matt, there will be some possibilities in the projected domain where you are looking at Matt, in addition to possibilities where you are looking at Raphael.

What is significant about the example of the man standing in the grove of trees in the distance is that it demonstrates how epistemic anchors—situations involving one's own epistemic states—can be required for certain kinds of modal thought. Integrating this insight into our general account of factual domain projection, the upshot is that passing certain kinds of modal reasoning tasks will require the cognitive ability to represent epistemic anchors. Such an insight may be especially critical given that an important step in cognitive development is the relatively late emergence of the ability to reason about one's own epistemic state (Beran, Perner, & Proust, 2012; Kloo, Rohwer, & Perner, 2017; Rohwer, Kloo, & Perner, 2012). Moreover, the example of the man standing in the distance also gives us a critical clue about which kinds of modal reasoning tasks do or do not require epistemic anchors. Our approach thus makes novel predictions about which modal reasoning tasks can be passed before, and which ones can only be passed after, the emergence of the ability to reason about one's own epistemic state.

The contrast between modal reasoning tasks that do vs. those that do not require epistemic anchors is brought out impressively in two tasks from Robinson and colleagues (Robinson, Rowley, Beck, Carroll, & Apperly, 2006) that are remarkably similar to one another. Somewhat older children (5- to 6-years-old) were introduced to two different paper bags—one of which

[3] This way of construing epistemic domains is in the spirit of Lewis (1996). For Lewis, your epistemic alternatives at a given time are the sets of worlds where you have the exact same perceptual experiences and memories you actually have at that time.

contained only black building blocks, and one of which contained both orange and green building blocks. They were also introduced to a cardboard wall with three differently colored doors—orange, black, and green—and it was explained that building blocks would be pushed through the door that corresponded to their color. The children's task was to place trays below the doors to make sure that the building blocks pushed through the doors were caught (illustrated in Fig. 4).

On some trials, called 'unknowable' trials, children were told that the experimenter was going to pick a block out of the bag with both orange and green blocks, and that they needed to place trays to make sure that the building block was caught when it came through the door. On these trials, Robinson and colleagues found that four- to five-year-old children placed trays beneath both the orange and green doors, indicating that they succeeded in generating some future possibilities in which an orange block was drawn from the bag and other future possibilities in which a green block was drawn. Critically, the pattern of responses diverged sharply on other trials, called 'unknown' trials. The only difference on these trials was that the experimenter had already drawn the block from the bag and placed it behind the corresponding door, but the children had not seen which block was drawn or which door it was placed behind. Now, when prompted to make sure the block was caught, four- to five-year-old children only placed a tray beneath either the green or orange door, but not both.

Given the minimal difference between the conditions, what explains why children were passing one version of this task but not the other? That is, why would children be better at generating the relevant possibilities for events that are 'unknowable' rather than simply 'unknown'? The difference is captured once one realizes that the 'unknown' task does, but the 'unknowable' task does not, require an epistemic state as anchor. In the 'unknowable' trials, children can use a contextually salient situation involving the actual blocks still in the bag as the modal anchor. With this anchor, the future possibilities in the projected modal domain differ with respect to the location of the block that was drawn. In some future extensions of the anchor, a green block is drawn and pushed through the green door; in others, an orange block is drawn and pushed through the orange door. By the age of 5, children should be able to generate these two types of future possibilities from a present anchor situation and select actions that succeed regardless of which possibility will be the actual one (see Figure 4), which is exactly what Robinson and colleagues find.

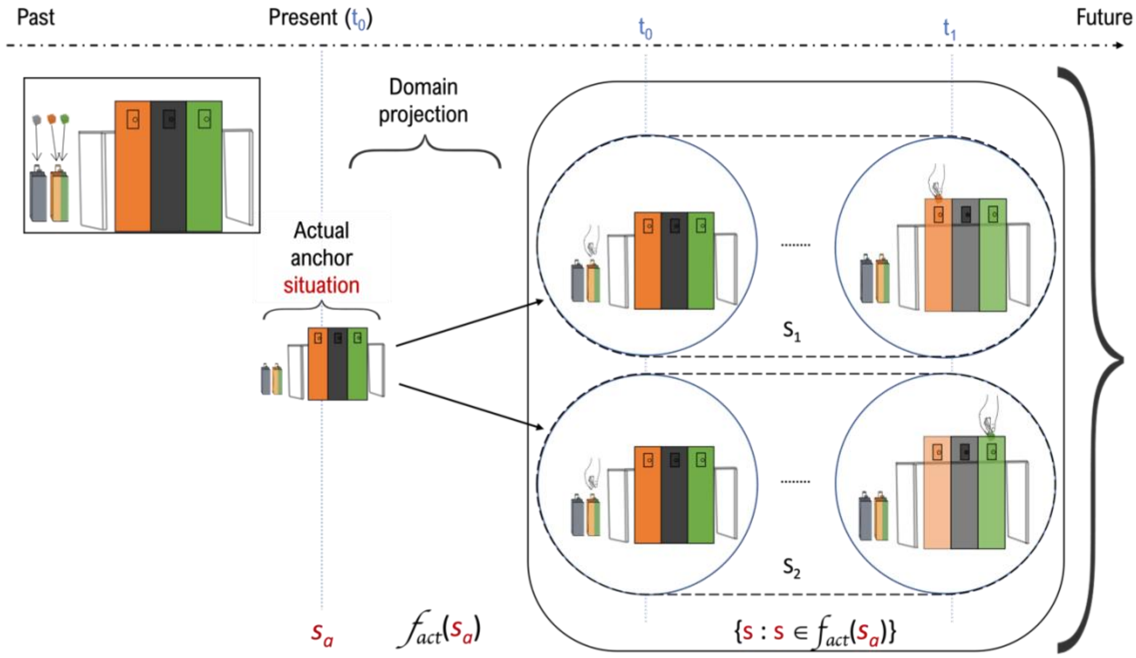

Figure 4. Schematic depiction of future-oriented factual domain projection used by children to pass the 'unknowable' trials in Robinson et al. (2006). The domain projection function, f_{act}, takes as input the modal anchor s_a (the situation in the actual world involving the experimental setup and the colored bags with blocks in them) and returns a set of situations, $\{s: s \in f_{act}(s_a)\}$, that have an exact match of s_a but also extend into the future. Among those situations are ones in which an orange block is drawn and placed behind the orange door (s_1) and ones in which a green block is drawn and placed behind the green door (s_2).

However, in the 'unknown' trials, this strategy won't work. It will not lead to a domain of possibilities that differ with respect to the location of the block that has been drawn. On these trials, a block has already been drawn and has been placed behind the corresponding door. If *that* situation were the modal anchor, all possibilities projected from it would contain that very same block sitting in that very same location. So, if 4- to 5-year-old children were attempting to use the actual present situation with the hidden block—whatever color it is and whatever location it is currently in—as the modal anchor, we can understand why they would only put down one tray. One shouldn't put down more than one tray because there is no possibility that the block that has already been drawn and is already behind one of the doors will come through more than one door. If the situation they are looking at is a situation in which the block is behind the green door, then in all future extensions of that situation, it will come through the green door. Alternatively, if the situation they are looking at is a situation in which the block is behind the orange door, then in all future extensions of that situation, it will come through the orange door. Since the children don't know where the block that was drawn is sitting, the best they can realistically do is pick a door randomly.

This picture of what young children may be doing when failing the 'unknown' version of this task aligns nicely with what has been observed in a series of related studies involving finding a previously hidden object in one of two locations (Mody & Carey, 2016). As argued for by Leahy and Carey (2020), a promising interpretation of the data is that children are essentially selecting randomly between one of the two locations, but then taking themselves to be surprisingly certain that all relevant possibilities are ones in which that object is in that location (see also Leahy, Huemer, Steele, Alderete, & Carey, 2022).

The remaining question then is how does anyone pass the ‘unknown’ version of the task? Generating a modal domain that includes possibilities where the block that was drawn is behind the orange door *as well as* possibilities where the block that was drawn is behind the green door, can be achieved by using an epistemic anchor. If you use yourself— including what you see and remember and everything else you know about the situation—as the anchor, and you yourself do not know whether the block that was drawn and placed behind the corresponding door is green or orange, then you can generate present-oriented extensions of that anchor situation, and in some of those possibilities an orange block is behind the orange door, while in others, a green block is behind the green door (see Figure 5).

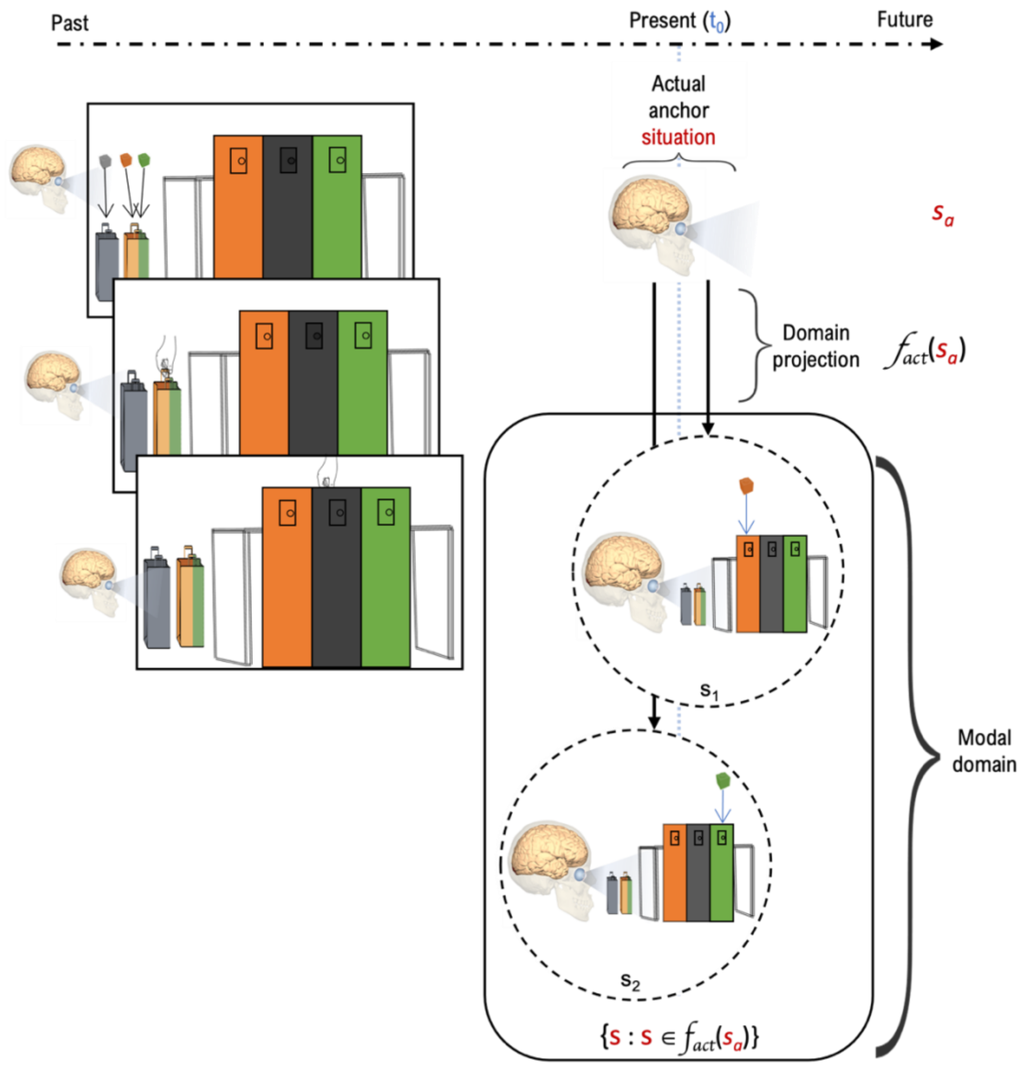

Figure 5. Schematic depiction of present-oriented factual domain projection required to pass the ‘unknown’ trials in Robinson et al. (2006). A domain projection function, f_{act}, takes as input the modal anchor s_a (the situation in the actual world involving the participant and their actual epistemic state at the present time) and returns a set of situations, $\{s: s \in f_{act}(s_a)\}$, that have an exact match of s_a but also extend spatially. Among those situations are ones in which an orange block is behind the orange door (s_1) and ones in which a green block is behind the green door (s_2).

Given our explanation of the difference between the ‘unknowable’ and ‘unknown’ tasks, it should not be surprising that children pass the ‘unknown’ condition from Robinson and colleagues (2006) later in development. To be able to use yourself in your current epistemic state as the modal anchor requires that you can represent your own epistemic state, and thus that you have the capacity for *metacognition*. Accordingly, the relatively late development of the ability to solve tasks such as the ‘unknown’ condition tracks the relatively late emergence of the ability for metarepresentation (Beran, et al., 2012; Kloo, et al., 2017; Rohwer, et al., 2012).

Researchers directly studying metarepresentation have uncovered a remarkably similar distinction in the development of children’s ability to reason about their own knowledge (Sodian & Wimmer, 1987; Rohwer, et al., 2012). In one study, Rohwer and colleagues hid toys in an opaque box and asked 3- to 7-year-old children to assess whether they knew which toy was hidden in the box. When children were able to directly see the toy that went into the box, they correctly assessed the fact that they knew which toy was in the box. Similarly, when they had not seen any of the toys before one was hidden, they correctly assessed the fact that they did not know which toy was hidden (for similar findings, see, Pratt & Bryant, 1990; Tardif, Wellman, Fung, Liu, & Fang, 2005). However, when children were first introduced to a couple of toys (e.g., a train and a ball) before one of them was hidden in the box out of view, young children exhibited a remarkable difficulty in assessing their own knowledge: 70% of 3- to 5-year-old children claimed to know which toy was in the box (Rohwer, et al., 2012; see also, Sodian & Wimmer, 1987). Passing this task, just as passing the ‘unknown’ condition in the Robinson et al. task requires the ability to correctly represent one’s own epistemic state. Interestingly, children begin to pass both tasks around the exact same age (around 6), and they also fail both tasks in a similar way before the age of 6: in both, they act as if one of the two possibilities were actual.

Carving epistemic modal thought at its joints

In the literature, cases like the one where you reason that a man standing in the distance on a hill might be some particular friend of yours tend to be lumped together with cases like the one where you come across a puddle and infer that it might have rained. Both types of inferences tend to be seen as instances of ‘epistemic’ reasoning. In English, the modal *might* can be used in both cases, and in the linguistics literature, English *might* is generally categorized as an ‘epistemic’ modal (see e.g. Papafragou, 1998; Cournane, 2020). On our account, the two kinds of modal inferences are importantly different. The puddle-inference, unlike the distant-man-inference, does not seem to *require* an epistemic modal anchor, hence does not require metacognition. By drawing a distinction between these two types of inferences, we make novel predictions about which kinds of modal inferences children will struggle with before the age of six. Modal inferences that require epistemic anchors, hence metacognition, to pass should be mastered around the age of six, while modal inferences that do not require epistemic anchors, such as determining possible causes of an event, should be mastered much earlier.

A relatively large body of literature suggests that young children show surprising proficiency in entertaining multiple hypotheses concerning possible causes of some outcome (Gopnik, Sobel, Schulz, & Glymour, 2001; Gopnik & Wellman, 2012; Gweon & Schulz, 2011). One beautiful set of experiments by Goddu and colleagues illustrates this important test of our account’s prediction (Goddu, Sullivan, & Walker, 2021). Children, between 18 and 30 months old, were given a causal reasoning task in which they had to determine what made a music box play music but were given insufficient evidence to identify a unique cause (see Fig. 3e top row). Specifically, children were shown evidence that was consistent both with (i) the box playing music whenever a mismatched pair of blocks was placed on top of it and (ii) the box playing music whenever one specific kind of block was placed on it (a blue cube in the illustration in Fig.

3e top row). All the evidence children received allowed for both of those two possible causes. After seeing the evidence, children were given two test trials that involved them selecting between blocks that could be placed on top of the music box to make it play music (Fig. 3e bottom row). On one test trial (Test 1: group possibilities in Fig. 3e), children were presented with two *pairs* of blocks they had never seen before. One of the pairs had matching blocks and the other had mismatching blocks. On a separate test trial (Test 2: Individual possibilities in Fig. 3e), the same children were asked to choose between two individual blocks which they had never seen previously activate the musical box on their own. One block had been part of both mismatched pairs that had activated the box, while the other had only been part of one of the mismatched pairs. The order of these trials was randomized and children did not receive feedback on whether their choice was correct. Goddu and colleagues observed that the children correctly selected the mismatching pair of blocks rather than the matching pair, and the same children also correctly selected the individual block that had been a part of both mismatching pairs rather than the block that had only been part of one of the pairs (Goddu, et al., 2021). In other words, Goddu and colleagues found that even toddlers seem to be able to simultaneously represent two possibilities containing different causes activating the box: What made the box play *might* have been a mismatched pairs of blocks, but it *might* also have been one particular block.

This task required two pieces of modal reasoning, both from anchor situations consisting of a music box. For the first piece, the box was playing or not playing depending on its immediate environment—pairs of blocks placed on top of it. The children had to infer what made the box play. The evidence they were given allowed for the possibility that one particular block alone had the power to get the box to play, but it also allowed for the possibility that any distinct two blocks together had that power. For the second piece of modal reasoning, children had to use their conclusions from the evidence they saw to generate future-oriented possibilities leading to success—future situations where the box would play music. Goddu et al. showed that eighteen months old children were able to draw the correct conclusions from the evidence they saw, even though this involved inferences about possible causes that we might unreflectively classify as ‘epistemic’ reasoning. While the reasoning required to pass the Goddu et al. music box task is complex, no epistemic anchors are necessary. On our account, the early success on even this relatively complex modal reasoning task is not surprising given the particular component pieces required to solve this task.

A late-emerging form of modal thought: Counterfactual reasoning

Perhaps the most well-studied form of modal thought in developmental psychology is the ability to reason *counterfactually*—an ability typically argued to not be fully operational until after the age of 8 and perhaps even as late as 13 (Harris, German, & Mills, 1996; Kominsky, Gerstenberg, Pelz, Sheskin, Singmann, Schulz, & Keil, 2021; Nyhout, Henke, & Ganea, 2019; Rafetseder, Cristi-Vargas & Perner, 2010; Rafetseder, Schwitalla, & Perner, 2013). Consider the task that is currently taken to be the litmus test for counterfactual reasoning. In this task, known as the ‘Muddy Shoes’ task, children are told a story in which two children, Susie and Max, have been playing outside and have gotten their shoes muddy. Susie and Max then both walk into the kitchen without taking their shoes off, and the kitchen floor becomes dirty. After hearing about what happened, children are asked whether the floor would have still been dirty if one of the children, e.g., Susie, had taken her shoes off. Using this kind of counterfactual reasoning task, researchers have found that children typically don’t ‘pass’ it until quite late in development; they say that the floor would not be dirty in that case (Rafetseder, et al., 2010; Rafetseder, et al., 2013).

The modal thought required by this task differs in a number of ways from the kinds of modal cognition we have considered thus far. First, there is a difference in the *time* of the modal anchor. In all the previous cases we discussed, the modal anchor was a contextually salient present part of the actual world. In the Muddy Shoes scenario, the anchor has to be a carefully chosen part of the past: We are going back to a time in the actual past when Susie and Max had already muddy shoes but had not yet walked into the kitchen. Second, in counterfactual reasoning, one needs to generate possible future extensions of the anchor situation where an assumption one knows to be actually *false* becomes true: Contrary to fact, Susie has to be assumed to not walk through the kitchen with muddy shoes. Thus, one needs to discard the fact that Susie walked in the kitchen (along with any other facts that are inconsistent with her not doing so). In the other forms of modal thought discussed thus far, the range of possibilities considered was always compatible with one's understanding of the actual world—they were possibilities that, for all one knew, might have been actual. Finally, the possibilities for counterfactual reasoning have to stay *close* to the actual world; they can't stray too much from actuality. In the Muddy Shoes scenario, the 'correct' judgment would require Susie to take off her shoes and not leave mud on the kitchen floor, but everything else should stay the same. In particular, Max would leave his shoes on and thus the floor would still get dirty. The pattern of apparent errors observed in the Muddy Shoes task is that before the age of 8, children tend to say the kitchen floor would *not* have been dirty if Susie had taken off her shoes. The question we face is, given the suite of coordinated abilities that are required to pass this task, why are children's responses differing from that of adults?

It is unlikely that the difficulty arises merely from representing past anchors. By this late in development, children have more than sufficient short-term memory to be able to recall the events in the presented narrative (Gathercole, 1999). Moreover, the studies typically include memory check questions to make sure that children correctly recall the past events that occurred, and those questions tend to be uniformly answered correctly (see, e.g., Rafetseder, et al., 2013).

Another possibility that has been pursued is that children may not be able to represent possibilities that are inconsistent with their understanding of the situation (Riggs, Peterson, Robinson, Mitchell, 1998; Peterson & Riggs, 1999). While this is possible, children by the age of 5 do seem to be able to successfully generate and reason over counterfactual states of affairs in other tasks. They can, for example, pass the false belief task, which requires predicting how an agent will act based on a representation of the world that is inconsistent with their own understanding (Perner, Sprung, & Steinkogler, 2004; Phillips & Norby, 2021; Wellman, Cross, & Watson, 2001), so it seems unlikely that this fully explains the surprisingly late development of counterfactual reasoning either.

A third way of explaining the observed pattern of 'errors' is to instead assume that younger children can successfully engage with all parts of counterfactual reasoning, except that they may have a different notion of what it may mean to stay close to the actual world after making a counterfactual assumption. For example, children might assume that Max and Susie were coordinating their actions. Maybe Max generally copies what Susie does, or maybe they were just playing *together*. If so, this should have a substantial impact on how one reasons counterfactually. If Max is a copycat, for example, then if Susi takes her shoes off, so does Max. In that case, then, staying close to the actual facts (including that Max is a copycat) means that the floor would not be dirty.

Kratzer (1981, 1989, 2012) discusses several cases where facts get 'lumped together' and thus stand and fall together when a counterfactual assumption is made. As in the Muddy Shoes case, this leads to apparent violations of the 'stay-close-to-the-facts' strategy: The more facts are lumped together, the more facts have to be discarded under a counterfactual assumption, and that means counterfactual conclusions may depart further and further from the way things actually are. Kratzer (1981) reports judgments about counterfactuals that are quite similar to those of the children in the Muddy Shoes task, but do not feel erroneous in any way. In one of her examples, she imagines that two friends—Hans and Babette—are spending the evening together. They go to a restaurant, 'Dutchman's Delight', sit down, order, eat, and talk. We are then asked to suppose counterfactually that Babette had gone to a different restaurant, 'Frenchman's Horror', instead. The question is where Hans would have gone in that case. Attaching importance to the fact that Hans and Babette are spending the evening together, a natural answer is that Hans, too, would have gone to Frenchman's Horror. Giving this answer, we no longer seem to hold on to the fact that Hans actually went to Dutchman's Delight.

In the light of Kratzer's example, a plausible explanation for children's apparent 'failures' on the Muddy Shoes task would be that they were indeed working under the assumption that Susie and Max coordinate their actions. Recent work by Nyhout, and colleagues (2019) supports this suggestion. They found that children performed significantly better in the Muddy Shoes task when Susie and Max walked on the kitchen floor at different times and for different reasons—that is, when it was made clear that there was no connection between their actions. The assessment that, in counterfactual reasoning, the primary difference between children and adults may be that children work with different generalizations is also in line with recent studies on children's counterfactual reasoning about the trajectories of physical objects (Kominsky et al., 2021). These studies found that when asked to engage in counterfactual reasoning, most 4-year-old children identified the correct situation in the past to begin reasoning from and did consider future possibilities that were inconsistent with what had actually occurred. However, the possibilities they considered deviated, often substantially, from what had actually occurred (Kominsky et al., 2021). This result would be consistent with the assumption that children may work with generalizations that connect facts that are not connected for adults: As in a lumping semantics (Kratzer, 1981), the smaller the lumps of facts that are connected, the smaller the deviation from the actual course of events when one of the facts in the lump is incompatible with a counterfactual assumption.

This account differs in an important way from accounts that argue that children before the age of 6 rely on "basic conditional reasoning" to pass some counterfactual reasoning tasks, but do not actually yet have the capacity for genuine counterfactual thought (Rafetseder, & Perner, 2014; Rafetseder et al., 2010; Rafetseder, O'Brien, Leahy, & Perner, 2021). To get a sense for the difference between the two kinds of proposals, consider a story used in an experiment by Rafetseder et al. (2010). In this story, the mother of two children (a boy and a girl) places candy either in a box on the top shelf of a cupboard in the kitchen or in a box on the bottom shelf of a cupboard. The boy is tall but the girl is short, so if there is candy on the bottom shelf, both can reach it, but if there is candy on the top shelf, only the boy can reach it. If children come into the kitchen and find the candy their mother has placed in the cupboard, they take it to their room. In the critical condition, the mother placed candy on the top shelf, and the boy came into the kitchen, found it and took it to his room. Subjects were then asked where the candy would be if the girl had come in the kitchen instead of the boy. Even 6-year-old children did not give the 'correct' answer—instead, saying that if the girl had come in the kitchen instead of the boy, the candy would be in the girl's room. Importantly, this 'failure' only occurs in this critical condition, but children 'passed' the three other variations: When the candy was described as being on the top shelf and the girl came into the kitchen for the candy, they correctly reasoned that if the boy had come instead, the candy would have ended up in his room. Moreover, when the candy was described as being on the bottom shelf and either the boy or girl actually took it, they correctly reasoned that if the other child had come into the kitchen instead the candy would have ended up in the other child's room. Rafetseder et al. explained the pattern across all four conditions by arguing that children are not actually engaging in counterfactual reasoning, but rather in "basic conditional reasoning". The idea is that the pattern can be predicted by assuming children are employing an unarticulated conditional generalization like, 'If child x comes for candy, the candy ends up in x's room.' This strategy fails in the critical condition but leads to correct responses in the other three conditions. Rafetseder et al. concluded that 6-year-olds have not mastered counterfactual reasoning, but only "basic conditional reasoning".

In contrast, to this proposal, we suggest that children could very well be engaging in genuine counterfactual reasoning but may lump facts of the world together in ways that adults do not (as we suggested for the muddy-shoes scenario). In the scenario of Rafetseder et al. (2010), a lumping account could assume that children see a connection between the candy being on the top shelf and the boy coming to get it. The two facts may be seen as connected, such that whenever there is candy on the top shelf, the boy comes and gets it. If children assumed such a generalization, they could conclude correctly that if the boy hadn't come and gotten the candy, there couldn't have been any on the top shelf. But if there *is* candy (as we should assume trying to stay as close to actuality as possible), and if the candy is *not* on the top shelf, then given the background story (telling us that if there is candy, it's either on the top shelf or on the bottom shelf), we can conclude that the candy would be on the bottom shelf. If that's the case, and the girl had come in the kitchen instead of the boy, she would have retrieved the candy and taken it to her room. Thus, like the Rafetseder et al. explanation, the lumping explanation predicts that subjects would give an apparently 'incorrect' answer in the critical condition, but correct answers in the three other conditions. The only difference between the children and adults would now be that children rely on a connection between the candy being on the top shelf and the boy coming to get it. That connection is not implausible, but also not completely warranted by the story they are told. If this is right, then—just as in the case of the muddy shoes example—the Rafetseder et al. data do not provide any reason to think that children younger than 6 lack an ability for counterfactual reasoning *per se*. We hope future work will provide a test between these approaches.

Stepping back from the details of these studies on counterfactual reasoning, the point we've sought to illustrate is how even quite late-emerging forms of modal thought, such as counterfactual reasoning, are still built around the basic capacity for factual domain projection, though they also require additional capacities. Mastering the counterfactual reasoning tasks we discussed requires subjects to locate a past anchor and to consider future-oriented possibilities projected from it. The special challenge presented by counterfactual reasoning is that the possibilities considered are known to be non-actual, but they can't diverge too far from actuality. What that means seems to at least partly depend on what generalizations about the actual world we are holding on to. That's a difficult issue, not just for a child. Counterfactual reasoning is known to be a slippery matter (see, e.g., Kratzer, 1981). As before, by decomposing the complex ability to reason counterfactually into its component parts, we can begin to pinpoint more precisely where children's capacities differ from that of adults.

The possibilities considered under counterfactual assumptions are restricted in a particular way: they can't stray too far from the actual world. We saw that connections and generalizations that are not explicitly mentioned might contribute to those restrictions. Implicit restrictions for domains of possibilities are a much more general phenomenon, though. They are not specific to counterfactual reasoning, or even modal thought more generally.

Default restrictions for modal domains

Factual domain projection returns a set of possibilities with a match of the anchor situation. This provides one kind of restriction on the kinds of possibilities in a modal domain—they all must have matches of the anchor situation. However, there will always be a very large number of far-fetched possibilities that could include a match of the anchor situation. For example, if one is considering what one could pack in a particular box when moving, there will be a truly enormous number of things that could be packed in that particular box; these may include a pet cat, another slightly smaller but equally empty box, and, almost certainly, a single lentil. For factual domain projection to be useful, the possibilities we consider need to be constrained to those that merit consideration (Phillips, et al., 2019).

This problem is solved by a general tendency for the possibilities we consider to be restricted by *normality*. This default is likely quite general, and can be found, not only in modal thought per se, but whenever possibilities are invoked in reasoning—whether one is processing the meaning of a sentence or simply making an inference about the shape of a partially occluded object (de Wit & van Lier, 2002). Kahneman (2011, p.77) provides an (unintended but beautiful) illustration of how normality constrains the possibilities we consider on the basis of verbal descriptions of scenarios. Kahneman mentions an item from Frederick's (2005) Cognitive Reflection Test (CRT) that goes as follows:

If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to take 100 widgets? 100 minutes OR 5 minutes?

In a footnote, Kahneman informs the reader that the "correct" answer is "5 minutes." However, to arrive at this apparently correct answer, we have to assume that the possibilities considered in this case are limited to those where each of the 100 machines produces 1 widget in 5 minutes, and where moreover all 100 machines start working at the exact same time. And we must exclude, for example, possibilities in which machines work sequentially, machines passing along a partially completed widget to the next machine. None of these assumptions logically follows from the verbal description we are given. They are expected to be presumed as 'normal.'

Of course, default normality constraints can also be shown to broadly constrain actuality-directed modal thought per se. It is well-established, for example, that the possibilities we generate tend not only to conform to the laws of physics, but also involve events that are statistically likely and actions that are morally good, rational, and conform to conventional norms (Baillargeon, 1987; Kahneman & Miller, 1986; McCloy & Byrne, 2000; Phillips & Knobe, 2018).

The presence of normality constraints on modal thought have been found as early in human life as has been tested (Baillargeon, 1987; Brown & Woolley, 2004; Chernyak, Kushnir, Sullivan, & Wang, 2013; Kalish, 1998; Lane, Ronfard, Francioli, & Harris, 2016; Phillips & Bloom, 2022; Shtulman, 2009; Shtulman & Carey, 2007; Shtulman & Phillips, 2018; Spelke, 2001; Stahl & Feigenson, 2015; Téglás, et al., 2007). For example, even very young infants' expectations about the movement of physical objects assume that physical generalizations concerning object continuity and solidity will not be violated (Baillargeon, 1987; Spelke, 2001; Stahl & Feigenson, 2015). Moreover, the previously discussed work by Téglás and colleagues has shown that infants' expectations about future events were guided by that event's probability. Infants were more surprised (measured by looking time) when an event with .25 probability occurred than when an event with .75 probability occurred (Téglás, et al., 2007). Moreover, when somewhat older children are asked explicitly about the possibility of different events happening, they judge that improbable events (e.g., being given all of the candy bars in a store for free) are impossible until well after 5 years of age. Remarkably, young children also explicitly judge that events in which agents violate prescriptive norms to be impossible, saying, for example, that it is impossible for someone to do something wrong, like taking a toy from another child. Importantly, these findings do not depend on children's interpretation of the word *possible*. They also judge that such events would require 'magic' to actually happen (Brown & Woolley, 2004; Phillips & Bloom, 2022; Shtulman & Carey, 2007; Shtulman & Phillips, 2018). And neither do those findings depend on something about English or western cultures, as similar findings have been reported across different languages and cultures (Chernyak, et al., 2013; Nissel & Woolley, 2022).

Normality-based restrictions on the possibilities considered have also been shown to persist into adulthood as a default (Phillips & Bloom, 2022). Recent work has found that the events that participants generate in open-ended decision problems are highly constrained to those that are likely to occur, involve rational and moral actions, and are believed to be normal (Hecht & Phillips, 2022; Srinivasan, Acierno, & Phillips, 2022). Moreover, these same features extend to adult's *default* sense of what is 'possible'. In one study, Phillips and Cushman (2017) asked adult participants to make judgments of possibility of various kinds of events, including ones that involved irrational and immoral actions. Importantly, participants were either forced to respond under time pressure, or were asked to respond after reflection. When participants were forced to make judgments of possibility quickly (and thus their judgments could not be adjusted from a 'default' understanding of the possibility of an event), they more tended to judge events involving immoral and irrational actions as impossible (Phillips & Cushman, 2017).

Investigating the defeasibility of normality constraints

An important aspect of our account is that there is no guarantee of the normality of all situations with an exact match of the anchor. Moreover, default normality constraints are not a core part of modal cognition; they serve as useful heuristics for constraining the domain of possibilities reasoned over but often are not able to withstand explicit challenge. To get a sense for this, consider the following scenario:

> **Scenario 1**: A child was born two years ago. The child was born from its mother's first pregnancy and its mother died a year later before becoming pregnant again.

And now, given that context, consider the modal claim (1):

(1) The child must not have any siblings.

Most likely, (1) will strike you as true. As (1) makes a universal claim, this means that all of the possibilities you represented were ones in which the child has no siblings. But such a homogenous domain is not given by factual domain projection; this homogeny is achieved instead a result of implicit normality constraints. And these implicit assumptions are defeasible: Consider the possibility that the mother gave birth to twins, triplets, or even more offspring from a single pregnancy. Once raised, such possibilities cannot continue to be excluded from the modal domain (despite their obvious abnormality) and once these possibilities have been included (1) should no longer strike you as true.

To systematically demonstrate the relationship between defeasible normality constraints and modal domains, we conducted a study in which half of participants were presented with Scenario 1 and asked to evaluate (1). They were subsequently asked to consider the possibility of twins, triplets, etc. and then asked to reevaluate (1). We expected relatively high agreement with (1) initially and comparatively lower agreement with (1) after we challenged the default normality assumptions by raising the possibility of twins, triplets, etc.

Critically, to show that this manipulation worked specifically because it challenged defeasible normality constraints, we would also want to compare agreement ratings in this case to those in a case where the possibility of twins, triplets, etc. would not have been implicitly excluded from the domain in the first place. To do so, the other half of participants were instead given the following scenario:

> **Scenario 2**: A dog was born two years ago. The dog was born from its mother's first pregnancy and its mother died a year later before becoming pregnant again.

After reading Scenario 2, participants were asked to rate their agreement with (2).

(2) The dog must not have any siblings.

As with Scenario 1, participants were subsequently asked to consider the possibility of twins, triplets, etc. and then asked to evaluate (2) again. In this case, we expected relatively low agreement with (2) initially, suggesting that the domain may already include possibilities involving multiple offspring from a single pregnancy. Moreover, we also predicted that explicitly raising such possibilities should therefore also have less of an impact on participants' reevaluation of (2). To allow us to further investigate whether this proposed mechanism, we additionally asked both groups of participants whether they had considered the possibility of twins, triplets, etc. before we raised it.

Results. Participants' agreement ratings with the two modal claims showed the expected pattern overall (see Fig. 6a). Statistically this pattern can be captured by the significant interaction effect between whether the scenario concerned a child or a dog and the impact of raising the possibility of twins, triplets, etc. This interaction was highly significant ($\chi^2(1) = 13.967$, $p < 0.001$). Specifically, in the case of the child, we found that participants largely agreed with (1) before the challenge ($M = 65.66$; $SD = 35.74$), but their agreement decreased markedly after the possibility of twins or triplets was raised ($M = 33.59$; $SD = 29.92$), $t(250.26) = 7.844$, $p < 0.001$, $d = 0.973$. By contrast, in the case of the dog, participants did not strongly agree with (2) even before the challenge ($M = 37.86$; $SD = 41.85$), and raising the possibility of multiple offspring from a single birth had a comparatively small effect on their reevaluation of the modal claim ($M = 26.67$; $SD = 36.68$), $t(246) = 2.24$, $p = 0.026$, $d = 0.284$.

We next considered whether these patterns could be explained by whether or not participants considered the option of multiple offspring from a single birth before we raised it. First, we found that when the scenario concerned a child, the majority of participants (72%) reported that they did *not* consider the possibility before we raised it. In contrast, when the scenario concerned a dog, the majority of participants (66%) reported that they *did* consider the possibility before we raised it. This difference was highly significant ($\chi^2 = 60.65$, $p < 0.001$). Second, combining the data from both conditions, we found that participants who had not thought of the possibility before we raised it overwhelmingly agreed with the modal claim, while those who had already considered this possibility, overwhelmingly disagreed with it. Accordingly, our explicit raising of this possibility only affected the agreement ratings of participants who had not already thought of the possibility (see Fig. 6b). Third, we conducted a mediation analysis that revealed that the differences in initial agreement with (1) vs. (2) were explained by differences in whether participants had considered the possibility of multiple offspring from a single pregnancy (95% CI of proportion mediated [0.43, 1.01], $p < 0.001$).[4]

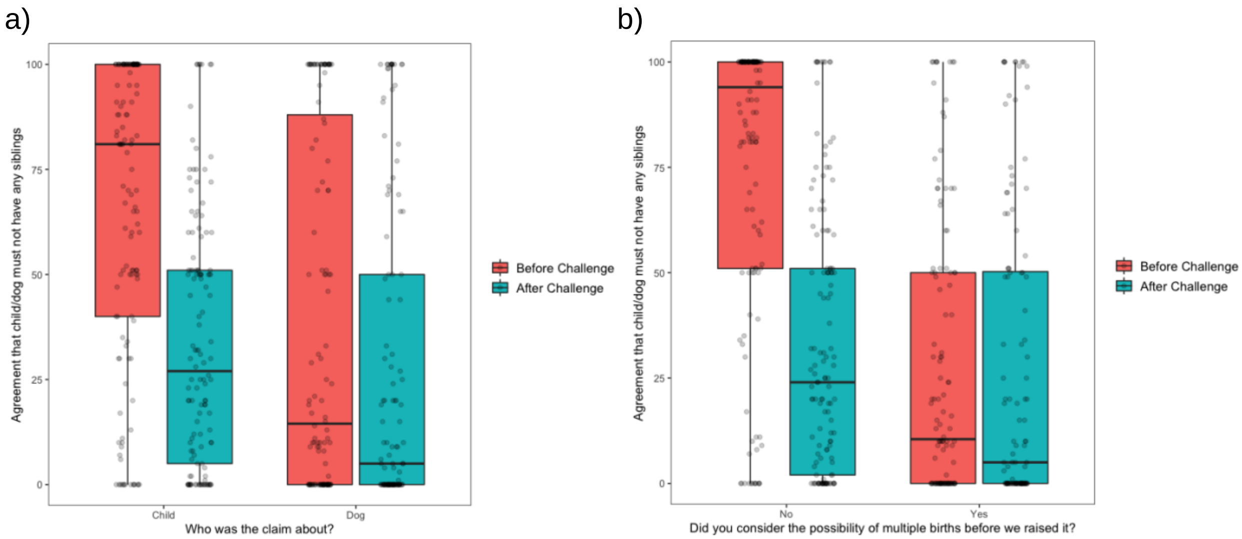

Figure 6. Boxplots of participants' agreement ratings with the modal claim, where small grey dots represent individual participants' agreement ratings. Red boxes depict agreement before we raised the possibility of twins, triplets, etc.; blue boxes depict agreement after we raised these possibilities. a) The bars grouped on the left depict agreement ratings when the scenario concerned a human child; the bars on the right depict agreement ratings when the scenario instead involved a dog. b) The bars grouped on the left depict agreement ratings of participant who indicated that they had not previously considered the possibility of twins, triplets, etc.; the bars on the right depict agreement ratings of participants who said they had considered such possibilities.

In short, what we find in this simple experiment conforms to the general contours of our account. On our approach, default normality constraints typically play a critical role in which possibilities are considered, but such default constraints are not special to modal thought. Moreover, these default constraints often do not stand up to challenge, suggesting that their influence does not come from the mechanism of factual domain projection itself.

Modality in natural languages

If the core cognitive mechanism of actuality-directed modal thought is factual domain projection from an anchor situation, we would expect that mechanism to be visible not only in developmental paths and cross-species variation, but also in variation in the way modal thought is expressed within and across languages. Most importantly, we would expect languages to have systematic ways of representing the core component of actuality-directed modal thought—domain projection from an anchor—and possibly also other components that combine with the core component to produce increasingly complex forms of modal thought.

In this section, we will give a brief overview of how complex actuality-directed modal thoughts are expressed in natural language. Within the limits of this article, our examples will be drawn from English, but our discussion will be informed by what is more generally known about how,

[4] These studies were approved by Dartmouth College's Committee for the Protection of Human Subjects (STUDY00032209). Additional experimental details, stimuli, materials, data, code, and a longer explication of the results can be found in the supplement to our paper: https://doi.org/10.7910/DVN/KUWNYK.

cross-linguistically, modals interact with other semantic building blocks, in particular with temporal and aspectual operators in their vicinity (Rullmann & Matthewson, 2018). What we hope will emerge from this short and preliminary overview is that the meaning components that we know natural languages rely on to compositionally construct expressions of modal thought from smaller pieces match up fairly well with the components of modal thought we identified on the basis of various non-linguistic tasks that have been used to establish milestones for cognitive development.

Locating anchors in modal expressions

An anchor-based semantics for modal expressions was first proposed by Hacquard (2006, 2010), and was developed further by Kratzer (2013). Kratzer hypothesized that modal anchors are generally provided by the arguments of modal words. The initial, still unrestricted, modal domain is then projected via factual domain projection from the anchor. Different kinds of modal words—e.g. modal adjectives vs. modal auxiliaries—take different kinds of arguments, hence select different kinds of anchors, and thus project different kinds of modal domains. This variation can be used to get a clear glimpse of the role of anchors in modal language. By way of illustration, consider the following scenario from Lewis (1997:14; also Kratzer, 2013).

"A sorcerer takes a liking to a fragile glass, one that is a perfect intrinsic duplicate of all the other fragile glasses off the same production line. He does nothing at all to change the dispositional character of his glass. He only watches and waits, resolved that if ever his glass is struck, then, quick as a flash, he will cast a spell that changes the glass, renders it no longer fragile, and thereby aborts the process of breaking."

Against the background of Lewis's scenario, look at the following two sentences.

(3) The glass is fragile.
(4) The glass could/might break.

(3) and (4) are close in meaning. A glass that is fragile could break easily, hence could break. Yet we judge (3) as clearly true on Lewis's story but are more hesitant about (4). A fragile glass remains fragile regardless of whether it is or isn't protected by a powerful sorcerer. But the presence of the sorcerer makes a difference with respect to whether or not the glass could break.

Modal adjectives such as *fragile* have an individual argument that is realized as its subject. That subject provides the modal anchor. Thus, in (3), the modal anchor is the glass at the current time. We are considering possibilities that have matches of the glass in its current state, but the surroundings of the glass may differ in whichever way: the glass may be on a shelf or packed safely away, for example. Crucially, in the projected possibilities, the glass may or may not be protected by the sorcerer, and if it isn't, it may break at some future time (see Fig. 7). Thus, (3) seems true as there are future possibilities in the domain in which the glass breaks.

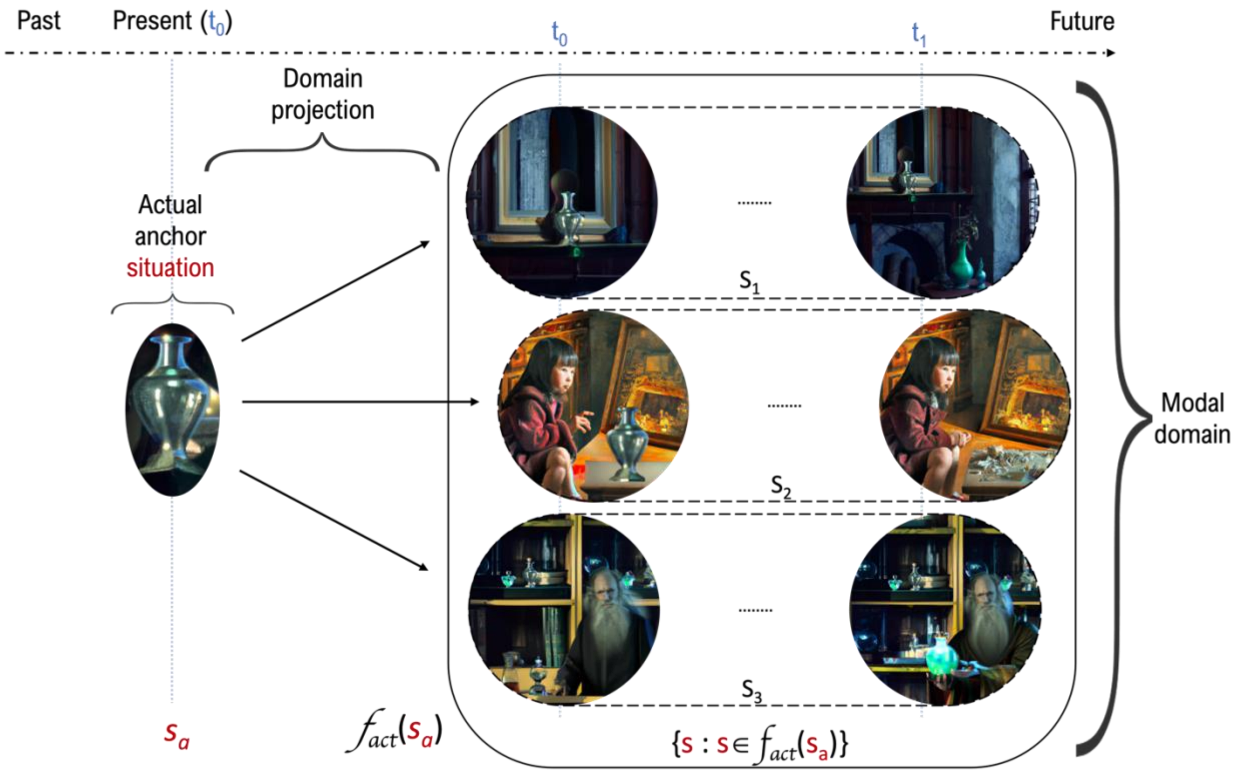

Fig. 7. Depiction of factual domain projection from an anchor that involves only the glass at the current time. Domain projection in this case returns a set of possibilites with exact matches of the glass at t_0. Among these may be situations in which it is not protected by a sorcerer but instead played with by a child (s_2), as well as ones in which it is protected by the sorcerer (s_3). In situations that do not include the sorcerer, the glass may go on to break at a subsequent time, as in s_2.

Modal auxiliaries like *could* rely on different kinds of modal anchors than modal adjectives like *fragile*. Modal auxiliaries are sentential operators. They do not have individual arguments at all. Contrary to first appearance, the grammatical subject in (4) is not an argument of *could*. To see this, consider (5):

(5) Three glasses could fit on this shelf.

(5) is ambiguous. There is scope interaction between *could* and the quantifier phrase *three glasses*. (5) might be understood as saying that there are three actual glasses that could fit on this shelf (wide scope of *three glasses*). Or it might say that the shelf can accommodate three glasses—not any particular ones (narrow scope of *three glasses*).[5] The two readings can be represented by the logical forms 6(a) and (b), where *could* is invariably a sentential operator. Being a sentential operator, *could* may or may not take scope over the subject *three glasses* (Bhatt 1999, Hackl 1998, Wurmbrand 1999).

---

[5] *Three glasses* in (5) is also ambiguous between a collective and a distributive interpretation. When we say that those three glasses fit on this shelf, we may mean that they fit there individually, or else collectively. Wide-scope and narrow-scope readings are available for both the distributive and the collective interpretations of *three glasses*.

(6) a. Three glasses λx (could (x fit on this shelf)).
'Three glasses have the property of being an x such that it is possible that x fits on this shelf'.
b. Could (three glasses fit on this shelf).
'It is possible that three glasses fit on this shelf'.

Because it is a sentential operator, the auxiliary *could* has only a situation argument, which provides the situation where the modal statement as a whole is evaluated. It has no other arguments. If modal domains are generally projected from a modal's arguments, then the modal domain for *could* has to be projected from its situation argument. Possible values of this argument may be smaller situations containing just the glass, or they may be larger situations containing the glass together with the sorcerer. There are no grammatical pressures favoring or militating against either possibility. Grammar doesn't tell us how to individuate or delimit the situations we are talking about.

Returning to (4), the upshot is that the modal anchor for *could* is more flexible than that for *fragile*. It may be a situation containing only the actual glass at the current time, or it may be the larger, contextually salient situation which contains the current stage of the glass together with the sorcerer protecting it. If the anchor situation for *could* includes the sorcerer, every projected possibility will have a match of that situation, and the sorcerer will thus be protecting the glass in each of those possibilities. So, if things go on to develop normally—the sorcerer doesn't quit or lose his powers—the sorcerer will protect the glass forever after, and the glass won't ever break (Fig 8).

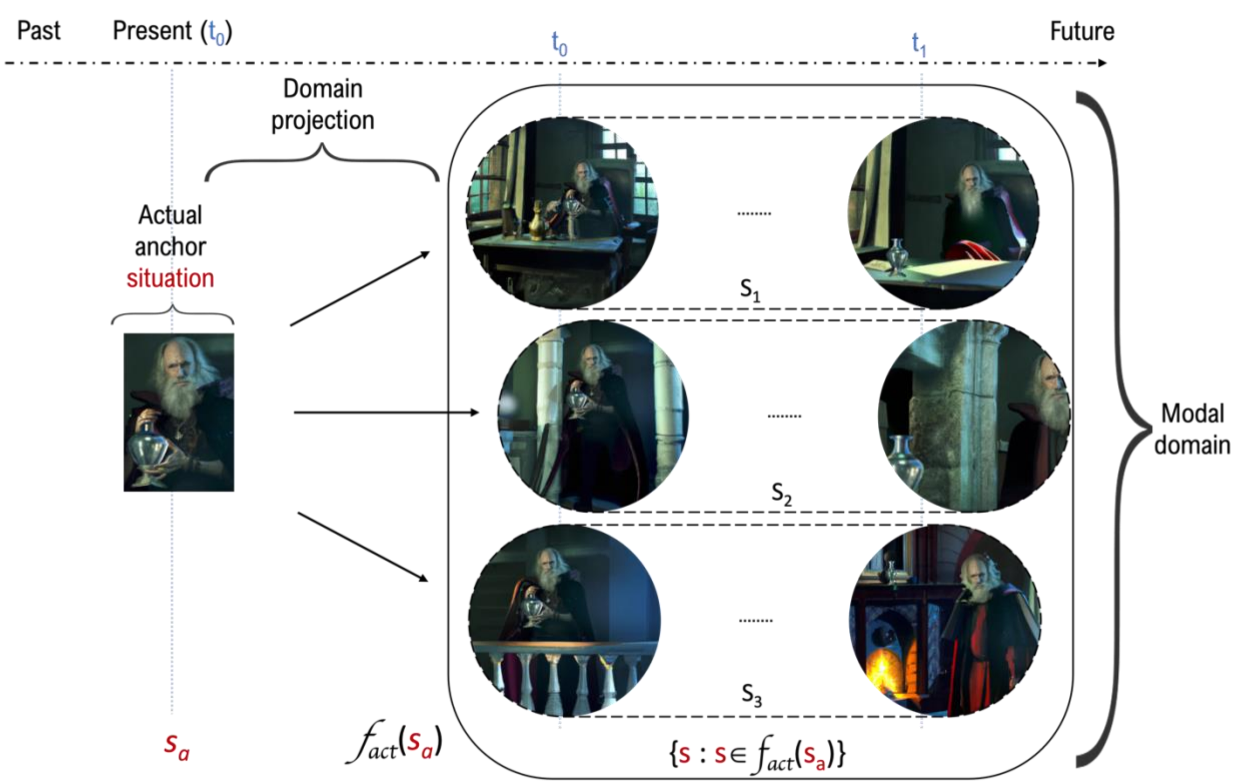

Fig. 8. Depiction of factual domain projection from an anchor that includes both the glass and the sorcerer at the current time. Domain projection in this case will return a set of situations with exact matches of the glass and sorcerer at t_0. In all of these situations, the

sorcerer protects the glass, and thus, if things go on to develop normaly, the sorcerer will prevent the glass from breaking in all possibilities in the domain.

Of course, as we've pointed out, the sorcerer doesn't HAVE to be included in the modal anchor for *could* in (4), and when he isn't, there will again be extensions of the anchor situation where no sorcerer is protecting the glass and it breaks, as illustrated in Figure 7.

Casting the difference between modal adjectives and modal auxiliaries in this way, we can explain the hesitancy felt when evaluating (4): we might feel pulled by the thought that the glass couldn't break because it is protected by the sorcerer, or we might feel pulled toward the thought that the sorcerer is protecting the glass because it is fragile and thus could break. Due to this indeterminacy, we would expect truth-value judgments for (4) to be more variable than those for (3), and they should vary according to whether or not the sorcerer is taken to be part of the anchor situation. In contrast to (4), the modal anchor for (3) is given by grammar, not contextual saliency, and we should thus expect relatively more agreement in truth-value judgments.

Experimentally demonstrating modal anchors

To illustrate and experimentally test these predictions, we gave participants Lewis's original story and then asked them to rate their agreement with either (3) or (4).[6] Subsequently, we asked them whether they were (a) only considering the glass or (b) considering the sorcerer and the glass when rating their agreement with the modal claim in the context of Lewis's scenario. This experiment allowed us to test five key predictions from the account we have just offered.

Results. First, we found that participants overall more agreed with the claim that the glass was fragile (M = 84.11; SD = 28.57), than with the claim that the glass could break (M = 61.58; SD = 39.29), confirming our original intuition, $t(514.08) = 7.626$, $p < .001$, $d = 0.633$ (see Fig. 9a). Second, participants' agreement ratings with (4) exhibited more variance than their agreement with (3) as predicted by the indeterminacy of the modal anchor for *could,* $F(328,201) = 1.892$, $p < .001$. Third, participants were less likely to report having considered the sorcerer when evaluating the *fragile* claim (31.5%) than when evaluating the *could* claim (64.5%), $\chi^2(1) = 52.91$, $p < .001$, which is predicted by the fact that the modal anchor for *fragile* is given by the grammatical subject, while *could* instead requires a contextually salient situation that may or may not include the sorcerer (Fig 9b). Fourth, whether or not participants included the sorcerer in the modal anchor was predictive of whether they agreed with the modal claim, $F(1,524) = 156.4$, $p < .001$, $\eta^2_p = 0.176$ (Fig 9c). And finally, a mediation analysis confirmed that the difference in agreement with (3) vs. (4) can largely be explained by whether or not the sorcerer was included in the modal anchor (95% CI of proportion mediated [0.32, 0,66], p< 0.001).[7]

[6] In this series of studies, we also asked participants to rate their agreement with modal claims about whether the glass *vulnerable* and whether it *might* break. Agreement with these claims also fit the predicted pattern and are reported in the supplement. We do not report them here for simplicity and brevity.
[7] Additional experimental details, stimuli, materials, data, code, and a longer explication of the results can be found in the supplement to our paper, https://doi.org/10.7910/DVN/KUWNYK

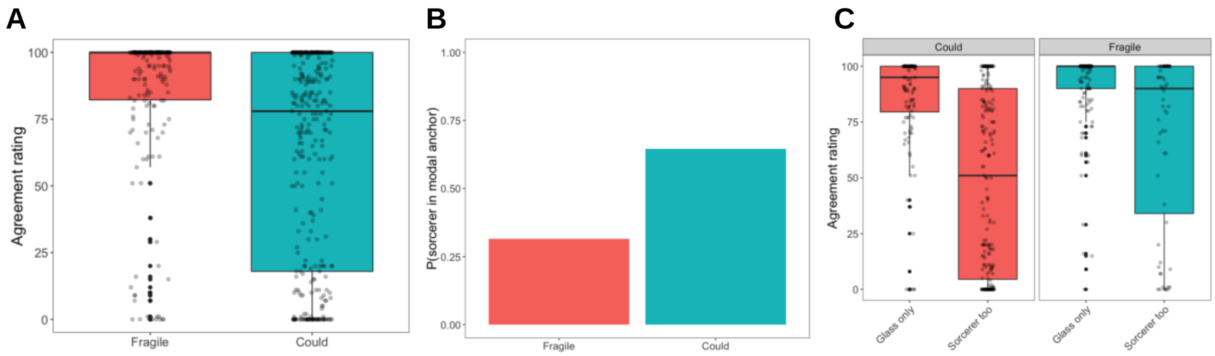

Fig. 9. **A** Boxplots of participants' agreement ratings with (3) (left box and points) and (4) (right box and points). Small grey dots depict individual participant responses, the colored boxes depict the middle two quartiles of responses, and the horizontal line depicts the median response. **B** depicts the probability that the sorcerer was included in the modal anchor for (3) (left bar) and (4) (right bar). **C** Boxplots of participants' agreement rating with (4) as a function of whether *only* the glass was included in the modal anchor (left boxes and points), or the sorcerer was included in the modal anchor along with the glass (right boxes and points), for both (3) right plot and (4) left plot.

Thus far in this section, what we've sought to demonstrate is how variation in truth-value judgments for different kinds of modal statements can be seen as arising from differences in modal anchors, where the choice of anchors is constrained by grammar. Modal domain projection from an actual anchor situation thus not only offers a way of accounting for variation in the development of modal reasoning, it also has the potential to explain variation in truth-value judgments for grammatically different modal statements.

As we argued in our review of the developmental literature, some of the observed variation in the development of modal thought can be explained by assuming that the core machinery of factual domain projection combines with other abilities. Here are the ones we discussed:

a. **Time of the anchor**: The ability to categorize anchors as present, past, or future.
b. **Temporal orientation of possible extensions of the anchor**: The ability to represent possible extensions of the anchor situation as stretching into the future or into the past, or as staying within the present.
c. **Epistemic anchors**: The ability to use one's own epistemic state as an anchor.
d. **Counterfactuality**: The ability to consider possible extensions of an anchor situation that one knows to be counterfactual.
e. **Restrictions for modal domains**: the ability to navigate defeasible normality restrictions for domains of possibilities.

In what follows, we will illustrate whether and, if so, how those five components of modal thought are reflected in the combinatorial semantic systems of natural languages.

Anchor time and temporal orientation of possible extensions

In our review of the development of modal thought, we saw that the first attested instances of modal thought involve future-oriented domain projection from a present anchor. Not surprisingly, then, the earliest modals acquired by children are also present tense instances of future-oriented modals of the kind illustrated in (5) below, which are referred to as 'root' modals, as opposed to ‘epistemic’ modals, in the linguistic literature (for overviews see Papafragou, 1998; Cournane, 2020).

(7) a. I **can** do this.
b. You **have to** help me.

In 7(a), the speaker says that there is a possible future extension of the current situation she is in where she does this action—whatever action it is she is referring to. In 7(b), the speaker conveys that in all possible—and acceptable—future extensions of the actual situation they are currently in, the addressee is helping her.

The modal anchors for the ‘root’ modals in 7(a) and (b) are located in the present. Natural languages also provide tools for talking about past modal anchors. For verbal modals with fully regular verbal paradigms the temporal location of the modal anchor is indicated by tense marking on the modal (Rullmann & Matthewson, 2018). (8) illustrates an instance of a future-oriented modal with a past anchor.

(8) I wasn’t sure whether I **could** take on this responsibility.

In (8), the modal *could* appears in its past tense form. The speaker is talking about a salient situation that occurred in the past. She is wondering whether there was a future extension of that situation where she takes on some particular responsibility. English thus uses tense morphology as a channel for specifying the time of the anchor situation.

The modals in 7(a), (b), and (8) are all future-oriented. So are all adjectival modals like *fragile*, *feasible*, *stealable*, and so on. A fragile glass is one that has a possible future instantiation that breaks, a feasible project has a possible future instantiation that is realized, and an stealable bike has a possible future instantiation that gets stolen. Strikingly, there do not seem to be adjectival modals that have a past orientation. For example, despite its potential usefulness, we do not have an adjective *kleptose* in English that is similar to *stealable*, but which applies to objects if they have possible past instantiations in which they were stolen. A bike would be kleptose if it is possible that in the past someone stole it, as in, *I wouldn’t recommend buying that that bike, it’s kleptose*. In natural languages, much like in cognitive development, future orientation seems to serve as a default. In both cases, a productive combinatorial system allows for variation away from this default.

Rullmann and Matthewson’s (2018) cross-linguistic study showed that, when expressed overtly, the temporal orientation of modals is encoded, not by the modals themselves, but by aspectual operators in the sentential constituent they scope over. (9) illustrates.

(9) a. She might **be** wait**ing** for you.
b. She might **have** depart**ed**.

In 9(a), the temporal orientation of the possibilities considered is present, which is indicated by the progressive form of the embedded verb *wait*. 9(a) talks about possibilities where she is presently waiting for you. In 9(b) the temporal orientation of the possibilities considered is past, which is indicated by the past perfect form of the embedded verb *depart*. Future orientation of modals is not marked in English, but this is not universally so. For example, a modal’s future orientation must be marked overtly on the verb embedded under the modal in Gitksan, an endangered Tsimshianic language of northwestern British Columbia (Matthewson, 2013).

To summarize, natural languages may explicitly encode both the time of the anchor situation and the temporal orientation of possible extensions of the anchor situation. The relevant meaning components are carried by tense and aspect markers that compositionally combine with the modal expressions themselves. The result is a combinatorics of meanings that allows us to talk about present-, past-, or future-oriented possibilities projected from present, past, or future modal anchors.

Epistemic anchors and the 'epistemic gap'

In our review of the cognitive development of modal reasoning, we argued that a major developmental achievement is acquiring the capacity to use one's own epistemic state as an anchor for projecting possibilities. A natural question is then whether this distinction maps onto the much-discussed linguistic distinction between 'root' and 'epistemic' modals we mentioned earlier (Papafragou, 1998; Cournane, 2020). (10) and (11) illustrate.

(10) 'Root' modals
I can (must, may, will, should) leave the room.

(11) 'Epistemic' modals
She could (must, might, will, should) have left the room.

'Root' modals are always future oriented and relate to abilities, potentials, intentions, or obligations, while 'epistemic' modals need not be future-oriented, but can also be past- or present-oriented, and have traditionally been understood as relating to information states or knowledge. Children produce so called 'root' modal verbs from around age 2, and so-called 'epistemic' modal verbs from around age 3 (for overviews see again Papafragou, 1998; Cournane, 2020). In the Bristol Language Development Study, for example, agent-oriented uses of *can* and instances of *will* used for communicating intentions are the earliest uses of modals by children (Wells, 1979). 'Epistemic' *may* and *might* appear at age 3;3 (Wells, 1985). This difference in the time of acquisition for 'root' versus 'epistemic' modals is called the 'epistemic gap' in the language acquisition literature.

Papafragou (1998) entertained the hypothesis that the epistemic gap occurs because 'epistemic' modals require metacognition, that is, the capacity for representing epistemic states. This hypothesis does not fit with the results of Kloo, et al. (2017), or Rohwer, et al. (2012), establishing that a capacity for metarepresentation of one's own epistemic states emerges quite late and is perhaps only fully realized by age 6. Papafragou's hypothesis has also been challenged by Cournane (2021), which presents corpus data from 17 English learning children showing that they are using contextually appropriate sentences with 'epistemic' adverbs like *maybe* from before age two, when they are still only using root modal verbs. For her, the later appearance of epistemic modal auxiliaries, as opposed to root modals, thus has to have linguistic explanation, possibly related to syntactic complexity.

Our approach to modal semantics via factual domain projection has the potential to resolve the conflict between Papafragou (1998) and Cournane (2021), while also being consistent with the results of Kloo, et al. (2017) and Rohwer, et al. (2012). We saw that not all instances of so-called 'epistemic' modals require epistemic anchors, that is, the metacognitive ability to represent one's own epistemic state. For example, among the early cases of 'epistemic' adverbs mentioned by Cournane (2021: 221) are examples like 12(a) and (b).

(12) a. Maybe grandma made this.
b. Maybe that's a fish.

A possible context for 12(a) may be a situation where we are looking at a hand-knit sweater. We can easily imagine possible past-oriented extensions of the present actual sweater situation where a whole variety of people might have knit that sweater. No metacognition is required. 12(b) might be about a sketchy drawing. Here, we can easily imagine possible past-oriented extensions of the actual situation with the drawing in it, but where the drawing was intended to represent a sock or a potholder. Again, no metarepresentation of one's own epistemic state is required. Put in general terms, our factual domain projection approach predicts that metacognitive capacities are only required with those instances of so-called 'epistemic' modals where the choice of a non-epistemic anchor would yield modal domains that are not diverse enough, that is, domains of possibilities that all give the same answer to the question of our inquiry. The modal domains projected for 12(a) and (b) do not suffer from this problem. Even if grandma actually made that sweater, that's not a property that the sweater has in every possible past-oriented extension of the anchor situation. It could have been made by somebody else. Likewise, even if this is actually a drawing of a fish, this is not a property that the drawing has in every possible past-oriented extension of the anchor situation. It could have been created to represent something else.

On our account, Papafragou (1998) is right in assuming that there is a connection between mastering 'epistemic' modals and the capacity for metacognition. Certain instances of so-called 'epistemic' modals do indeed require metacognition. Interestingly, the standard tasks in experimental studies on the acquisition of 'epistemic' modals are hidden-object tasks, and the modals figuring in those tasks are thus precisely the kind of modals that require epistemic anchors. In Noveck et al. (1996), for example, children were introduced to two open boxes and a third, closed, box. They saw that one of the open boxes contained a bear and a parrot, and the other open box only had a parrot. They were told that the closed box had the same content as one of the open boxes, and then heard each of two puppets articulate a modal statement. One of the statements was true, the other was false. The children's task was to say which of the two puppets was right. Among the contrasting statements that Noveck et al. tested were 13(a) and (b).

(13) a. There might be a bear in the box.
b. There has to be a bear in the box.

Given the information the children were presented with, 13(a) is true, but 13(b) is false. Yet even 5-year olds' choice between 13(a) and (b) was at chance. On our approach, this result is expected. To generate a modal domain where a bear is in the box in some, but not all possibilities, children would have had to use their own epistemic states as anchors. But the capacity for metacognition about one's own epistemic state only emerges around 6 years of age. If children instead use the actual box itself as the modal anchor, then they will either generate a domain in which all relevant possibilities are ones in which there is a bear in the box, or they will generate a domain in which none of the relevant possibilities are ones in which there is a bear in the box. In the first case, both sentences are true, and children should choose at chance if forced to choose between 13(a) and (b); In the second case, both sentences are false, and children should again choose at chance between 13(a) and (b). In short, our account predicts that they will not be able to distinguish between possibility and necessity statements in tasks that require an epistemic anchor.

While hidden object tasks allow us to detect children's inability to use epistemic anchors, there are many more situations where epistemic anchors are required: An approaching man might or might not be your friend Matt, a bird you are looking at might be a sparrow or a female house finch, and the plants you see on the shelf over there might or might not be fake. Do children use 'epistemic' modals in those situations? And if they do, how do they understand them? That last question is critical, but we don't have the data to answer it. On our account, we would expect that children younger than 6 might very well produce modal statements like *maybe the man is Matt*, *maybe the bird is a sparrow*, or maybe *the plants are fake*, but, in line with Noveck et al. (1996), we wouldn't expect young children to be able to distinguish between possibility and necessity interpretations for those statements.

Our approach is thus entirely compatible with the corpus data presented in Cournane (2021) and Wells (1985), which showed that children produce contextually appropriate instances of so-called 'epistemic' modals (adverbs as well as auxiliaries) well before age 6, hence well before the age they can be assumed to have the necessary metacognitive capacities. On our approach, many so-called 'epistemic' uses of modals do not require epistemic anchors. And in contexts where epistemic anchors would be required, corpus data alone do not allow us to infer whether children are, in fact, relying on them.

To conclude, from our perspective, the traditional linguistic category of 'epistemic' modal covers two uses that can be distinguished, depending on whether or not they REQUIRE an epistemic anchor. The distinction is crucial for understanding a critical juncture in the development of modal thought, as we have seen. However, while the distinction is critical in cognitive development, we do not know of any language that marks it overtly. There may be good reasons for generally using our own epistemic states as anchors for past- or present-oriented modals. To see this, consider the puddle case again. Wouldn't the presence of particular amounts of particular kinds of ions exclude yesterday's rainstorm as having caused the puddle without major departures from what's normal in the actual world? Wouldn't even this case be a case that 'forces' us to use an epistemic anchor? Upon reflection, wouldn't almost any case be that way?

Evidentials. One type of linguistic expressions whose mastery seems to invariably require metacognitive capacities about one's own knowledge are so-called 'evidentials'. Evidentials mark how a person acquired evidence bearing on the truth of a proposition (for overviews see Murray, 2021a, 2021b). In some languages, evidentiality is obligatorily marked in every sentence. Tariana, an endangered Arawak language spoken in Northwestern Brazil, for example, has obligatory verbal inflections that mark whether the evidence supporting an assertion was direct visual evidence, direct auditory evidence, indirect visual evidence, evidence via common knowledge inference, or hearsay (Aikhenvald, 2004).

Mastery of evidentials requires the capacity to reflect on one's own epistemic state and to represent the source of some of the information it contains. It should thus not come as a surprise that evidentials are indeed acquired late. Ozturk and Papafragou (2016), for example, found that the semantics and pragmatics of evidential morphology in Turkish are only acquired at around age 6 or 7, which again aligns with the age when one begins to succeed on tasks that require reasoning about one's own mental states.

Counterfactuality

The expression of counterfactuality across languages is complicated and far from transparent (von Fintel & Iatridou, forthcoming). To illustrate, consider (14) against the background of the Muddy Shoes scenario discussed previously:

(14) If Susie had taken off her shoes, Max would have, too.

In the context of Muddy Shoes, we readily understand (14) as making a counterfactual assumption. After all, we were told that Susie and Max did NOT take their shoes off. Where does the counterfactual interpretation come from? It seems that English does not have dedicated morphology contributing a counterfactual interpretation.[8] (15) illustrates.

(15) Copycat
When I came into the kitchen, I noticed that Susie had taken off her muddy shoes. I couldn't see Max, but knowing what a copycat he is, I knew that **if Susie had taken off her muddy shoes, Max would have, too**.

(15) provides a context for (14) where it doesn't have a counterfactual interpretation. Whether a sentence like (14) does or does not have a counterfactual interpretation, then, depends on the context where it is used. (14) acquires a counterfactual interpretation only in contexts where it is presumed that Susie didn't take off her muddy shoes. In the context of Muddy Shoes, the *if*-clause (antecedent) of (14) is known to be false, in the context of Copycat, it is known to be true. The same conditional (14) can also be used in contexts where it's unknown whether its antecedent is true, as in (16).

(16) Uncertainty
Susie and Max were playing outside and I saw them enter the house with their muddy shoes. I feared that they might have entered the kitchen without taking their shoes off. Knowing Susie as the more reasonable kid, I thought there was a slight chance that she might have taken her shoes off before going into the kitchen. And, knowing what a copycat Max is, I was sure that **if Susie had taken off her muddy shoes, Max would have, too**.

While English doesn't have to morphologically distinguish between conditionals whose antecedent is known to be false, known to be true, or not known to be true or false, it does single out counterfactuals negatively. In contexts like Copycat or Uncertainty, but, crucially, not in Muddy Shoes, English can also use a plain indicative conditional like (17).

(17) If Susie took off her muddy shoes, Max did, too.

Mastering the full range of conditionals in English, then, does not only require the capacity to assess the status of a conditional antecedent with respect to what is known in a context. It also requires the capacity to navigate a complex linguistic situation where none of the three relevant factors for the interpretation and appropriate use of a simple conditional construction—whether

---

[8] The only remnant of what might be taken to be counterfactual morphology in English is the form *were* in cases like *if I were here*, *you would be, too*. Yet even this form doesn't necessarily signal that an assumption is false. It may also signal uncertainty about the truth of an assumption, as in: *If I were chosen for the position, I would turn your company around*. This could be a promise made in a job interview by someone who still hopes to be chosen for the position. See von Fintel & Iatridou (forthcoming) for more discussion.

the antecedent is known to be true, known to be false, or not known to be true or false—has a unique morphological expression. A conditional like (14) can be used in all three conditions, and a conditional like (17) in two. Not surprisingly, English conditionals corresponding to (14) are acquired late (Crutchley, 2004, 2013).

While English has no distinctive marking of counterfactuality, one can still find some of the ingredient pieces for counterfactual reasoning. English does mark the pastness of the modal anchor and the future orientation of the possibilities considered. Interestingly, these are precisely the components that all three possible interpretations of sentences like (14) have in common. The future-oriented modal *will* appears in its past tense form *would* in (14), and the pastness of the modal anchor is also indicated by the past tense of the antecedent.

The expression of counterfactuality shows considerable variation, even among closely related languages. German, for example, uses indicative conditionals corresponding to (17) in the same contexts English does, but uses past subjunctive mood marking obligatorily in contexts where the antecedent is known to be false, and optionally in contexts where it's not known whether the antecedent is true or false. Pastness of the modal anchor is expressed by the fact that past, rather than present subjunctive mood is used in German counterparts of (14). Future orientation of the possibilities considered is not expressed. This is no surprise, given that future orientation of modals is a default, as we have seen. However, in contrast to English, subjunctive mood marking prevents the German counterparts of (14) from being used in situations where the antecedent is known to be true.

Given substantial differences in the linguistic expression of counterfactual thought, we might expect different trajectories for the acquisition of counterfactual conditional constructions in different languages.[9]

Domain restriction

We saw in the developmental section that modal domains, and domains of possibilities more generally, are constrained by normality restrictions. Those restrictions come in by default and do not have to be indicated explicitly. However, the modal vocabularies of natural languages also include items that explicitly direct us away from default normality restrictions (Rubinstein, 2012, 2014; Phillips & Cushman, 2017; Phillips, et al., 2019; von Fintel & Iatridou, forthcoming). (18) is one of Rubinstein's examples.

(18) [Preparing a company's tax report.]
a. We have to report all of our revenue.
b. We should report all of our revenue.
Rubinstein (2014, 538).

18(a) relies on default normality restrictions for the possibilities considered, which are presumed to be shared without need for prior negotiation. In this particular case, the possibilities considered all conform to current tax laws. In contrast, as Rubinstein observes, the *should*-claim in 18(b) conveys that tax evasion is among the possibilities considered, albeit judged to be non-

---

[9] Linguistic and psycholinguistic investigations of counterfactuals usually look at constructions where the counterfactual assumption is expressed by a subordinate sentence corresponding to an *if*-clause in English. But there are, of course, also counterfactuals like (i):
(i) Without a coat you would be freezing now.
These constructions may offer researchers in Cognitive Development a promising method for investigating counterfactual reasoning independently of a demanding syntactic construction.

optimal. On this view, both *have to* and *should* are necessity modals, but they differ in what they indicate about how their domains are restricted: *Have to* relies on default normality restrictions, while the more specialized *should* indicates the presence of restrictions that depart from the default.

The existential modals *could* and *might* differ in a similar way: all-purpose *could* relies on default normality restrictions, whereas the more specialized *might* signals a departure from the default. In this case, departing from the default means that considerations of what is morally acceptable can no longer play a role. The Evil Ship Captain scenario illustrates.

(19) Evil Ship Captain

FitzRoy is a notoriously ruthless pirate who decided to pose as a captain of a ship in order to steal several expensive sculptures that a museum needs to transport across the sea. After he posed as an ordinary ship captain who was taking some passengers across the sea, FitzRoy got the job.

While sailing on the sea, a large storm came upon FitzRoy and his small ship. As the waves began to grow larger, FitzRoy realized that his small vessel was too heavy and the ship would flood if he didn't make it lighter. The only things on FitzRoy's small boat were the expensive art sculptures that he was stealing and the passengers he was transporting. He knew he had to throw something overboard to keep the ship from capsizing.

In the context of Evil Ship Captain, compare 20(a) and (b).

(20) a. FitzRoy could throw the sculptures overboard.
b. FitzRoy might throw the sculptures overboard.

The possibilities we consider for *could* in 20(a) are influenced by what would be morally acceptable. This should not be surprising. Moral considerations are independently known to provide a default constraint on the possibilities considered and the way they are ranked. This has been separately observed for the possibilities considered in causal reasoning (e.g., Alicke, 2000; Knobe & Fraser, 2008; Kominsky & Phillips, 2019; Kominsky, Phillips, Gerstenberg, Lagnado, &. Knobe, 2015), the possibilities considered in making judgments of force and freedom (e.g., Phillips & Knobe, 2009; Phillips & Young, 2011; Bernhard, LeBaron, & Phillips, 2022), the possibilities entertained in counterfactual reasoning (e.g., McCloy & Byrne, 2000; N'gbala & Branscombe, 1995; Byrne & Timmons, 2018), the possibilities entertained when asking explicitly about what is possible (e.g., Phillips & Cushman, 2017; Shtulman & Phillips, 2018), and in many other cases as well (e.g., Pettit & Knobe, 2009; Phillips, Lugri, & Knobe, 2015). In the Evil Ship Captain scenario, FitzRoy throwing the sculptures overboard is clearly among the highest ranked possibilities on a moral scale, and thus it is not surprising that we agree with 20(a). For *might*, on the other hand, moral considerations are prevented from influencing the possibilities considered. It is this departure from the default that seems to be lexically signaled by using *might*, rather than *could*. Without these moral considerations, the possibilities for *might* are restricted and ranked in a way that more clearly reflects what is likely to happen. Given FitzRoy's evil nature, the highest ranked possibilities for *might* are less likely to include possibilities in which the sculptures, rather than the passengers, are thrown overboard, hence our comparative disagreement with 20(b). An opposing pattern can be found when comparing 21(a) and (b).

(21) a. FitzRoy could throw the passengers overboard.
b. FitzRoy might throw the passengers overboard.

Here, the comparative reliance on default normality constraints of *could* but not *might*, should now lead us to agree with 21(b) more than with 21(a).

We confirmed these predictions experimentally. Participants were first asked to read the Evil Ship Captain context, and then were asked to rate their agreement with all four modal claims 20(a), (b) and 21(a), (b) in random order. For simplicity we focus here on participants' first responses, though similar patterns are found when analyzing all of participants' responses. Overall, we found the predicted interaction between the modal term used (*could* vs. *might*) and whether the modal claim concerned throwing passengers or sculptures overboard, $F(1,185) = 54.730$, $p < .001$, $\eta^2_p = 0.228$ (Fig 10). Specifically, participants more agreed that FitzRoy *could* throw the sculptures overboard ($M = 92.17$; $SD = 14.29$) than they agreed that he *might* ($M = 46.60$; $SD = 30.67$), $t(56.28) = 8.828$, $p < .001$, $d = 1.948$. And, in contrast, they less agreed that FitzRoy *could* throw the passengers overboard ($M = 65.13$; $SD = 37.04$) than they agreed that he *might* ($M = 79.89$; $SD = 22.97$), $t(90.80) = -2.422$, $p = .017$, $d = 0.469$.[10]

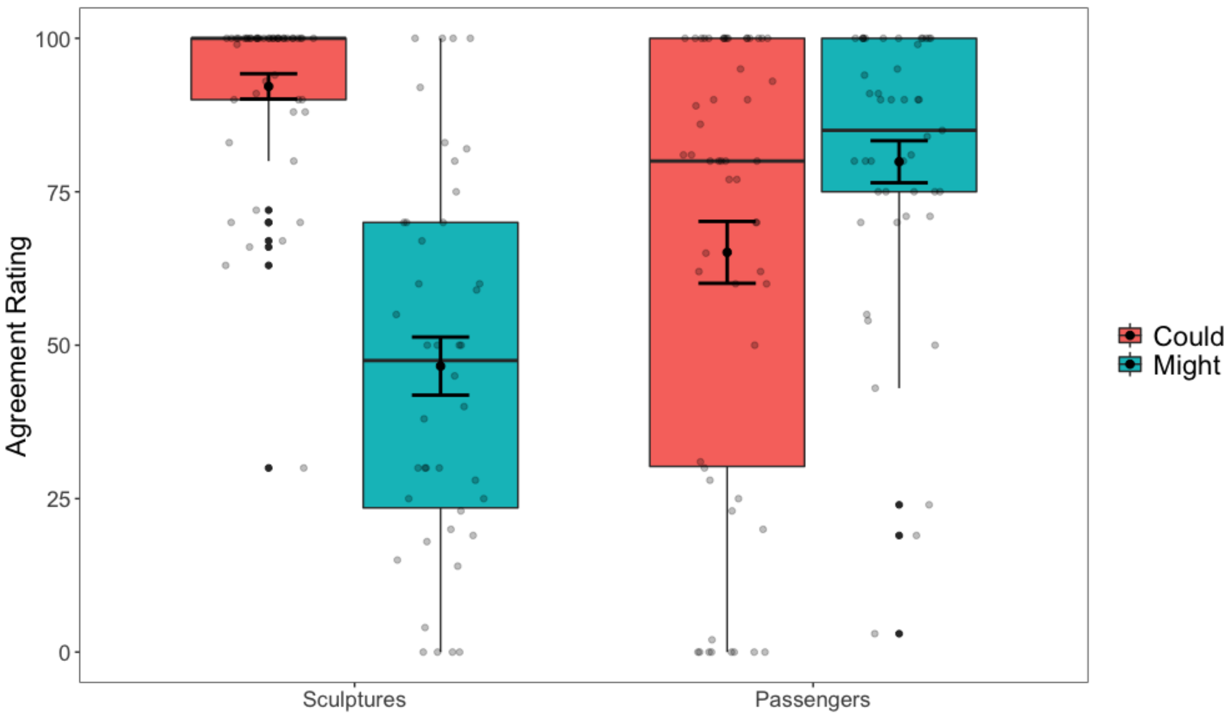

Fig. 10. Boxplots of participants' agreement the existential modal claims when they concerned ratings FitzRoy throwing the passengers overboard (left boxes and points) or the sculptures overboard (right boxes and points), both when the existential modal used was *could* (red boxes) and when it was *might* (blue boxes). The colored boxes depict the middle two quartiles of responses, the horizontal line depicts the median response, and the small grey dots depict individual participant responses. Thick black dots depict mean agreement and error bars represent +/- 1 SEM.

[10] Additional experimental details, stimuli, materials, data, code, and a longer explication of the results can be found in the supplement to our paper: https://doi.org/10.7910/DVN/KUWNYK

Further evidence for taking more specialized modal vocabulary items to indicate departures from default normality restrictions can be found in studies that require participants to make a range of different modal judgments either very quickly or after taking a moment before responding (Phillips & Cushman, 2017, Acierno, Mischel, & Phillips, 2022). If normality-based domain restriction is a default which is only moved away from with additional specific lexical information, then we should expect that the differences between the kinds of domain restriction indicated by different modal terms (e.g. *could* vs. *might*) would require some additional, non-default, processing. Accordingly, we should then also expect that if people were not allowed enough time to complete that processing before having to respond, their judgments concerning these two different modal terms will begin to look more similar to one another, as they would all be forced to rely on default, normality-based, domain restriction. This prediction of our account of default normality restrictions is also borne out. Phillips and Cushman (2017) presented participants with different background contexts and asked them to make truth-value judgments of modal statements that concerned various events occurring in those contexts. Participants were either forced to respond quickly or were asked to reflect before responding. Phillips and Cushman then computed the similarity between (among others) modal judgments with *could* vs *might*, both when participants were responding quickly and when they were responding slowly. This analysis revealed that these modal judgments became more similar to one another when participants were forced to answer quickly but were significantly less similar when participants were allowed additional time before responding (Phillips & Cushman, 2017). Studies focusing on the processing required when making different modal judgments thus provide support for a view on which normality-based domain restriction serves as a default that can be moved away from with additional processing when the lexical information of a more specialized modal requires it.

Fitting our pieces together

We have now outlined what we take to be the main components of actuality-directed modal thought, how these pieces can be seen in cognitive development, and how they may be expressed in natural languages. Before concluding this paper, we want to consider one further question: when we utter (or understand) a sentence involving a modal expression, how do the linguistic components we've illustrated relate to the non-linguistic components of modal thought we began with? How do modal thought and talk fit together? The beginning of an answer to this question, we think, can be found in some relatively routine observations. It is probably not by mere accident that similar component pieces can be posited to explain variation in the development of actuality-directed modal thought and the linguistic expression of actuality-directed modal thought across languages. Our non-linguistic modal thoughts can best be shared with others when we can compress them into a linguistic representation that reflects their structure. To do this optimally, linguistic representations of modality would be expected to encode the component pieces that are required for other minds to generate a representation of possibilities that corresponds to our own. We cannot share our non-linguistic modal thoughts with others directly, but we may have a common recipe for recreating them. Modal language provides a way of encoding the necessary ingredients in a linguistic form that can be shared.

With these observations in hand, we can now formulate a more precise question: Which ingredient pieces do linguistic representations of modality encode, and which ingredient pieces are left to non-linguistic cognition for sorting out? How do modal thought and talk fit together? The account we've offered provides some cases in which the division of labor begins to become clearer. Linguistic representations of modality do not attempt to encode all of the relevant assumptions required to restrict the domain of possibilities; such work is left to non-linguistic cognition, and linguistic representations instead tend to mark deviations from default normality-based domain restriction. In contrast, linguistic representations of modality often do explicitly encode the modal anchor, as in the case of modal adjectives like *fragile*, as well as marking the time of the anchor and indicating when temporal orientation of domain projection is past or present. There are also many cases in which the labor is more integrated: determining the anchor situation for modal auxiliaries like *could* may require some collaborative work on the part of non-linguistic cognition in picking out the contextually salient situation.

At the same time, natural languages vary in what is and is not encoded. For example, some languages require explicit encoding of the force of a modal claim (e.g. existential vs. universal), while others do not (Rullmann, et al., 2008; Deal, 2011). Some languages have lexically specified deontic modals (Rullmann, et al., 2008), others – like English – do not. Not all components of modal thought we identified have to be morphologically encoded. English, for example, does not morphologically mark counterfactuality, nor does it seem to lexically require anchors to be epistemic states. Thus, while there may be relatively clear examples of a division of labor within a given language, it seems unlikely that there will be any general recipe for how modal thought and talk fit together. Rather, the generalization we are led to by natural language is that there are many different ways in which enough of the component pieces of modality can be encoded to allow for non-linguistic cognition to infer the necessary remaining ingredients and faithfully recover modal thoughts; each language offers a blueprint for how this can be done.

Conclusion

We've set out to offer a unified way of decomposing actuality-directed modal thought into its component pieces. Actuality-directed modal thought is not a monolithic capacity that emerges fully formed. It is a complex capacity built from simpler parts that can be productively combined to allow for an increasingly impressive range of abilities. Our account of how this decomposition might go provides a starting point for understanding how, starting from pieces of actuality, natural minds represent possibilities. As researchers across the cognitive sciences continue to study modal thought, we suspect that new discoveries will require aspects of our decomposition to be refined or abandoned altogether. But our hope in offering this proposal is that it will serve as the beginnings of a common language for thinking about the component pieces of actuality-directed modal thought, allowing, for example, developmental psychologists to make claims more easily understood and modeled by logicians, or comparative cognition researchers to test ideas originally formulated by philosophers and semanticists.

Author note. These studies were approved by Dartmouth College's Committee for the Protection of Human Subjects (STUDY00032209). All experimental details, stimuli, materials, data, code, and a longer explication of the results can be found in the supplement to our paper: https://doi.org/10.7910/DVN/KUWNYK

Acknowledgments

We'd like to thank the institutions that made our collaboration possible: The Radcliffe Institute for Advanced Study of Harvard University, The Centre for Advanced Studies in the Humanities "Human Abilities" in Berlin, and the Diesel Café in Somerville. For detailed comments, discussions, and other forms of support we thank Fiery Cushman, Mariel Goddu, Joshua Knobe, Daniel Lassiter, Brian Leahy, and Barbara Vetter.

References

Acierno, J, Mischel, S., & Phillips, J. (2022). Moral judgements reflect default representations of possibility. *Phil. Trans. R. Soc. B*, *377*: 20210341.

Aikhenvald, A. (2004). *Evidentiality*. Oxford University Press.

Alderete, S. & Xu, F. (2022, January). Preschoolers' ability to think about alternative possibilities. Presentation at the 12th Annual Meeting of the Budapest CEU Conference on Cognitive Development.

Alicke, M. D. (2000). Culpable control and the psychology of blame. Psychological Bulletin, 126(4), 556.

Baillargeon, R. (1987). Object permanence in 3½-and 4½-month-old infants. *Developmental Psychology*, 23(5), 655.

Beck SR, Robinson EJ, Carroll DJ, & Apperly IA. (2006). Children's thinking about counterfactuals and future hypotheticals as possibilities. *Child Development*, 77(2):413-26.

Beran, M. J., Perner, J., & Proust, J. (Eds.). (2012). Foundations of metacognition. Oxford University Press.

Bernhard, R. M., LeBaron, H., & Phillips, J. (2022). It's not what you did, it's what you could have done. *Cognition*, 228, 105222.

Bell, V. A., & Johnson-Laird, P. N. (1998). A model theory of modal reasoning. *Cognitive Science*, 22(1), 25-51.

Bhatt, R. (1999). Covert Modality in Non-Finite Contexts. University of Pennsylvania PhD dissertation.

Browne, C. A., & Woolley, J. D. (2004). Preschoolers' magical explanations for violations of physical, social, and mental laws. *Journal of Cognition and Development*, 5(2), 239-260.

Byrne, R. M. (2007). *The Rational Imagination: How people create alternatives to reality*. MIT press.

Byrne, R. M., & Timmons, S. (2018). Moral hindsight for good actions and the effects of imagined alternatives to reality. *Cognition*, 178, 82-91.

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.

Cesana-Arlotti, N., Téglás, E., & Bonatti, L. L. (2012). The probable and the possible at 12 months: Intuitive reasoning about the uncertain future. *Advances in child development and behavior*, 43, 1-25.

Chernyak, N., Kushnir, T., Sullivan, K. M., & Wang, Q. (2013). A comparison of American and Nepalese children's concepts of freedom of choice and social constraint. *Cognitive Science*, 37(7), 1343-1355.

Condoravdi, C. (2002): Temporal Interpretation of modals: Modals for the present and for the past. In D. Beaver, S, Kaufmann, B. Clark and L. Casillas (eds.). *The Construction of Meaning*, 59–88. CSLI Publications.

Cournane, A. (2020). Learning modals: A grammatical perspective. *Language and Linguistics Compass*, *14*(10), 1-22.

Cournane, A. (2021). Revisiting the epistemic gap: It's not the thought that counts. *Language Acquisition*, 28(3), 215-240.

Crutchley, A. (2004). 'If she had of shutted the cage, the rabbit wouldn't escape': Past counterfactuals elicited from 6- to 11-year-old children. *First Language,* 24(2), 209-240.

Crutchley, A. (2013). Structure of child and adult past counterfactuals, and implications for acquisition of the construction. *Journal of Child Language,* 40, 438-468.

De Brigard, F., Addis, D. R., Ford, J. H., Schacter, D. L., & Giovanello, K. S. (2013). Remembering what could have happened: Neural correlates of episodic counterfactual thinking. *Neuropsychologia*, *51*(12), 2401-2414.

Deal, A. R. (2011). Modals without scales. *Language*, 87(3), 559-585.

Engelmann, J.M., Völter, C.J., O'Madagain, C., Proft, M., Haun, D.B.M, Rakoczy, H., & Herrmann, E. (2021). Chimpanzees consider alternative possibilities, *Current Biology*, 31(20): R1377-R1378.

von Fintel, K. & Iatridou, S. (forthcoming). Prolegomena to a theory of X-marking. *Linguistics & Philosophy*.

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives*, 19, 25–42.

Gathercole, S. E. (1999). Cognitive approaches to the development of short-term memory. *Trends in Cognitive Sciences*, *3*(11), 410-419.

Gerstenberg, T., Goodman, N. D., Lagnado, D. A., & Tenenbaum, J. B. (2021). A counterfactual simulation model of causal judgments for physical events. *Psychological Review*, *128*(5), 936.

Goddu, M. K., Sullivan, J. N., & Walker, C. M. (2021). Toddlers learn and flexibly apply multiple possibilities. *Child Development*, *92*(6), 2244-2251.
Gopnik, A., Sobel, D. M., Schulz, L. E., & Glymour, C. (2001). Causal learning mechanisms in very young children: two-, three-, and four-year-olds infer causal relations from patterns of variation and covariation. *Developmental Psychology*, 37(5), 620.
Gopnik, A., & Wellman, H. M. (2012). Reconstructing constructivism: causal models, Bayesian learning mechanisms, and the theory theory. *Psychological Bulletin*, *138*(6), 1085.
Gweon, H., & Schulz, L. (2011). 16-month-olds rationally infer causes of failed actions. *Science*, 332(6037), 1524-1524.
Hackl, Martin. 1998. On the semantics of ability attributions. Manuscript, MIT.
Hacquard (2006): *Aspects of Modality*. MIT PhD dissertation.
Hacquard, Valentine. (2010). On the event-relativity of modal auxiliaries. *Natural Language Semantics* 18, 79-114.
Harris, P. L., German, T., & Mills, P. (1996). Children's use of counterfactual thinking in causal reasoning. *Cognition*, 61(3), 233-259.
Hecht, E. & Phillips, J. (2022). Domain-general nodal spaces play a foundational role throughout high-level cognition. *Proceedings of the 44th Annual Meeting of the Cognitive Science Society.*
Johnson-Laird, P. N., & Ragni, M. (2019). Possibilities as the foundation of reasoning. *Cognition*, *193*, 103950.
Kaelbling, L. P., Littman, M. L., & Cassandra, A. R. (1998). Planning and acting in partially observable stochastic domains. *Artificial Intelligence*, 101(1-2), 99-134.
Kahneman, D. (2011). *Thinking, Fast and Slow*. Macmillan.
Kahneman, D., & Miller, D. T. (1986). Norm theory: Comparing reality to its alternatives. *Psychological Review*, 93(2), 136.
Kalish, C. (1998). Reasons and causes: Children's understanding of conformity to social rules and physical laws. *Child Development*, 69(3), 706-720.
Kloo, D., Rohwer, M., & Perner, J. (2017). Direct and indirect admission of ignorance by children. *Journal of Experimental Child Psychology*, 159, 279-295.
Knobe, J., & Fraser, B. (2008). Causal judgment and moral judgment: Two experiments. *Moral Psychology*, 2, 441-8.
Kominsky, J. F., Gerstenberg, T., Pelz, M., Sheskin, M., Singmann, H., Schulz, L., & Keil, F. C. (2021). The trajectory of counterfactual simulation in development. *Developmental Psychology*, 57(2), 253.
Kominsky, J. F., Phillips, J., Gerstenberg, T., Lagnado, D., & Knobe, J. (2015). Causal superseding. *Cognition*, 137, 196-209.
Kratzer, A. (1981). Partition and revision: The semantics of counterfactuals. *Journal of Philosophical Logic*, 10(2), 201-216.
Kratzer, A. (1989). An investigation of the lumps of thought. *Linguistics and Philosophy*, 607-653.
Kratzer, A. (2012). *Modals and conditionals: New and revised perspectives* (Vol. 36). Oxford University Press.
Kratzer, A. (2013). Modality for the 21st Century. In S. R. Anderson, J. Moeschler, F. Reboul (eds.): *The Language Cognition Interface / L'Interface Language-Cognition*. Geneva-Paris (Librairie Droz). 179-199.
Lane, J. D., Ronfard, S., Francioli, S. P., & Harris, P. L. (2016). Children’s imagination and belief: Prone to flights of fancy or grounded in reality?. *Cognition*, 152, 127-140.
Leahy, B. P., & Carey, S. E. (2020). The acquisition of modal concepts. *Trends in Cognitive Sciences*, *24*(1), 65-78.
Leahy, B., Rafetseder, E., & Perner, J. (2014). Basic conditional reasoning: How children mimic counterfactual reasoning. *Studia Logica*, 102(4), 793-810.
Leahy, B., Huemer, M., Steele, M., Alderete, S., & Carey, S. (2022). Minimal representations of possibility at age three. Unpublished manuscript.
Leahy, B., & Zalnieriunas, E. (2022). Might and might not: Children's conceptual development and the acquisition of modal verbs. In *Semantics and Linguistic Theory* (Vol. 31, pp. 426-445).
Lewis, D. (1973). Counterfactuals and comparative possibility. In *Ifs* (pp. 57-85). Springer, Dordrecht.
Lewis, D. (1974). Causation. *The Journal of Philosophy*, *70*(17), 556-567.
Lewis, D. (1996). Elusive knowledge. *Australasian Journal of Philosophy*, 74(4), 549-567.
Lewis, D. (1997). Finkish dispositions. *The Philosophical Quarterly*, 47, 143-158.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. MIT press.
Matthewson, Lisa. (2013). Gitksan Modals. *International Journal of American Linguistics,* 79, 349-394.
McCormack, T., & Hoerl, C. (2020). Children's future-oriented cognition. In *Advances in Child Development and Behavior,* 58, 215-253.
McCloy, R., & Byrne, R. M. (2000). Counterfactual thinking about controllable events. *Memory & Cognition*, 28(6), 1071-1078.
Mody, S., & Carey, S. (2016). The emergence of reasoning by the disjunctive syllogism in early childhood. *Cognition*, 154, 40-48.
Morris, A., Phillips, J., Huang, K., & Cushman, F. (2021). Generating options and choosing between them depend on distinct forms of value representation. *Psychological Science*, 32(11), 1731-1746.
Murray, S. (2021a). Evidentiality, modality, and speech acts. *Annual Review of Linguistics* 7. 213-233.
Murray, S. (2021b). Evidentials. In *The Wiley Blackwell Companion to Semantics*, ed. by Daniel Gutzmann, Lisa Matthewson, Cécile Meier, Hotze Rullmann, & Thomas Ede Zimmermann.
N'gbala, A., & Branscombe, N. R. (1995). Mental simulation and causal attribution: When simulating an event does not affect fault assignment. *Journal of Experimental Social Psychology*, 31(2), 139-162.
Nissell, J. & Woolley, J. D. (2022) Why wearing a yellow hat is impossible: Chinese and U.S. children's possibility judgments. Presentation at the Annual Meeting of the Cognitive Development Society. Madison, WI.
Noveck, I., Ho, S., & Sera, M. (1996). Children's understanding of epistemic modals. *Journal of Child Language,* 23, 621-643.
Nyhout, A., Henke, L., & Ganea, P. A. (2019). Children's counterfactual reasoning about causally overdetermined events. *Child Development*, 90(2), 610-622.
Ozturk, O. & Papafragou, A. (2016). The acquisition of evidentiality and source monitoring. *Language Learning and Development,* 12(2), 199-230.
Papafragou, A. (1998). The acquisition of modality: Implications for theories of semantic representation. *Mind & Language*, *13*(3), 370-399.
Pearl, J. (2009). *Causality*. Cambridge university press.
Perner, J., Sprung, M., & Steinkogler, B. (2004). Counterfactual conditionals and false belief: A developmental dissociation. *Cognitive Development*, *19*(2), 179-201.
Peterson, D. M., & Riggs, K. J. (1999). Adaptive modelling and mindreading. *Mind & Language*, 14(1), 80-112.
Pettit, D., & Knobe, J. (2009). The pervasive impact of moral judgment. *Mind & Language*, 24(5), 586-604.
Phillips, J., & Bloom, P. (2022). Do children believe immoral events are impossible? Unpublished Manuscript.
Phillips, J., & Cushman, F. (2017). Morality constrains the default representation of what is possible. *Proceedings of the National Academy of Sciences*, *114*(18), 4649-4654.
Phillips, J., & Knobe, J. (2009). Moral judgments and intuitions about freedom. *Psychological Inquiry*, 20(1), 30-36.
Phillips, J., & Knobe, J. (2018). The psychological representation of modality. *Mind & Language*, *33*(1), 65-94.
Phillips, J., Luguri, J. B., & Knobe, J. (2015). Unifying morality's influence on non-moral judgments: The relevance of alternative possibilities. *Cognition*, 145, 30-42.
Phillips, J., Morris, A., & Cushman, F. (2019). How we know what not to think. *Trends in Cognitive Sciences*, *23*(12), 1026-1040.
Phillips, J., & Norby, A. (2021). Factive theory of mind. *Mind & Language*, 36(1), 3-26.
Pratt, C., & Bryant, P. (1990). Young children understand that looking leads to knowing (so long as they are looking into a single barrel). *Child Development*, 61(4), 973-982.
Rafetseder, E., Cristi-Vargas, R., & Perner, J. (2010). Counterfactual reasoning: Developing a sense of "nearest possible world". *Child Development*, 81(1), 376-389.
Rafetseder, E., O'Brien, C., Leahy, B., & Perner, J. (2021). Extended difficulties with counterfactuals persist in reasoning with false beliefs: Evidence for teleology-in-perspective. *Journal of Experimental Child Psychology*, 204, 105058.
Rafetseder, E., & Perner, J. (2010). Is reasoning from counterfactual antecedents evidence for counterfactual reasoning? *Thinking & Reasoning*, 16(2), 131-155.

Rafetseder, E., Schwitalla, M., & Perner, J. (2013). Counterfactual reasoning: From childhood to adulthood. *Journal of Experimental Child Psychology*, 114(3), 389-404.
Redshaw, J., & Suddendorf, T. (2016). Children's and apes' preparatory responses to two mutually exclusive possibilities. *Current Biology*, *26*(13), 1758-1762.
Redshaw, J., & Suddendorf, T. (2020). Temporal junctures in the mind. *Trends in Cognitive Sciences*, *24*(1), 52-64.
Redshaw, J., Suddendorf, T., Neldner, K., Wilks, M., Tomaselli, K., Mushin, I., & Nielsen, M. (2019). Young children from three diverse cultures spontaneously and consistently prepare for alternative future possibilities. *Child Development*, *90*(1), 51-61.
Riggs, K. J., Peterson, D. M., Robinson, E. J., & Mitchell, P. (1998). Are errors in false belief tasks symptomatic of a broader difficulty with counterfactuality?. *Cognitive Development*, 13(1), 73-90.
Rohwer, M., Kloo, D., & Perner, J. (2012). Escape from metaignorance: How children develop an understanding of their own lack of knowledge. *Child Development*, 83(6), 1869-1883.
Robinson, E. J., Rowley, M. G., Beck, S. R., Carroll, D. J., & Apperly, I. A. (2006). Children's sensitivity to their own relative ignorance: Handling of possibilities under epistemic and physical uncertainty. *Child Development,* 77(6), 1642-1655.
Rubinstein, A. (2012). Roots of Modality. University of Massachusetts at Amherst PhD dissertation.
Rubinstein, A. (2014). On necessity and comparison. *Pacific Philosophical Quarterly*, 95, 512–554.
Rullmann, H. & Matthewson, L. (2018). Towards a theory of modal-temporal interaction. *Language* 94(2), 281-331.
Rullmann, H., Matthewson, L., & Davis, H. (2008). Modals as distributive indefinites. *Natural Language Semantics,* 16(4), 317-357.
Shtulman, A. (2009). The development of possibility judgment within and across domains. *Cognitive Development*, 24(3), 293-309.
Shtulman, A., & Carey, S. (2007). Improbable or impossible? How children reason about the possibility of extraordinary events. *Child Development*, *78*(3), 1015-1032.
Shtulman, A., & Phillips, J. (2018). Differentiating "could" from "should": Developmental changes in modal cognition. *Journal of Experimental Child Psychology*, *165*, 161-182.
Skinner, B. F. (1965). *Science and Human Behavior*. Simon and Schuster.
Sodian, B., & Wimmer, H. (1987). Children's understanding of inference as a source of knowledge. *Child Development*, 424-433.
Spelke, E. S., Breinlinger, K., Macomber, J., & Jacobson, K. (1992). Origins of knowledge. *Psychological Review*, 99(4), 605.
Srinivasan, G., Acierno, J. & Phillips, J. (2022). The shape of option generation in open-ended decision problems. *Proceedings of the 44th Annual Meeting of the Cognitive Science Society.*
Stahl, A. E., & Feigenson, L. (2015). Observing the unexpected enhances infants' learning and exploration. *Science*, 348(6230), 91-94.
Suddendorf, T., Crimston, J., & Redshaw, J. (2017). Preparatory responses to socially determined, mutually exclusive possibilities in chimpanzees and children. *Biology Letters*, 13(6), 20170170.
Tardif, T., Wellman, H. M., Fung, K. Y. F., Liu, D., & Fang, F. (2005). Preschoolers' understanding of knowing-that and knowing-how in the United States and Hong Kong. *Developmental Psychology*, 41(3), 562.
Téglás, E., & Bonatti, L. L. (2016). Infants anticipate probabilistic but not deterministic outcomes. Cognition, 157, 227-236.
Téglás, E., Girotto, V., Gonzalez, M., & Bonatti, L. L. (2007). Intuitions of probabilities shape expectations about the future at 12 months and beyond. Proceedings of the National Academy of Sciences, 104(48), 19156-19159.
Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: The truth about false belief. *Child Development*, 72(3), 655-684.
Wells, Gordon. (1979). Learning and Using the Auxiliary Verb in English. In V. Lee (ed.), *Cognitive Development: Language and Thinking from Birth to Adolescence*. London (Croom Helm), 250–70.
Wells, Gordon. (1985). *Language Development in the Pre-School Years.* Cambridge University Press. Cambridge.
de Wit, T. C., & van Lier, R. J. (2002). Global visual completion of quasi-regular shapes. *Perception*, *31*(8), 969-984.
Wurmbrand, S. (1999). Modal verbs must be raising verbs. *Proceedings of WCCFL,* 18, 599-612.

Young, L., & Phillips, J. (2011). The paradox of moral focus. *Cognition*, 119(2), 166-178.