



Finding the force: How children discern possibility and necessity modals

Anouk Dieuleveut¹ · Annemarie van Dooren² · Ailís Courname³ · Valentine Hacquard⁴

Accepted: 28 April 2022

© The Author(s), under exclusive licence to Springer Nature B.V. 2022

Abstract

This paper investigates when and how children figure out the force of modals: that possibility modals (e.g., *can/might*) express possibility, and necessity modals (e.g., *must/have to*) express necessity. Modals raise a classic subset problem: given that necessity entails possibility, what prevents learners from hypothesizing possibility meanings for necessity modals? Three solutions to such subset problems can be found in the literature: the first is for learners to rely on downward-entailing (DE) environments (Gualmini and Schwarz in *J. Semant.* 26(2):185–215, 2009); the second is a bias for strong (here, necessity) meanings; the third is for learners to rely on pragmatic cues stemming from the conversational context (Dieuleveut et al. in *Proceedings of the 2019 Amsterdam colloquium*, pp. 111–122, 2019a; Rasin and Aravind in *Nat. Lang. Semant.* 29:339–375, 2020). This paper assesses the viability of each of these solutions by examining the modals used in speech to and by 2-year-old children, through a combination of corpus studies and experiments testing the guessability of modal force based on their context of use. Our results suggest that, given the way modals are used in speech to children, the first solution is not viable and the second is unnecessary. Instead, we argue that the conversational context in which modals occur is highly informative as to their force and sufficient, in principle, to sidestep the subset problem. Our child results further suggest an early mastery of possibility—but not necessity—modals and show no evidence for a necessity bias.

Keywords Modals · Negation · Language acquisition · Corpus study · Human Simulation Paradigm experiment

1 Introduction

Modals are used to talk about possibilities and necessities, that is, non-actual states of affairs. This paper investigates how children figure out the force of the modals in their language: that words like *can*, *may*, or *might* in (1a) express possibility, whereas words like *must*, *should*, or *have to* in (1b) express necessity.

Extended author information available on the last page of the article

- (1) a. You *can/may/might* / . . . go this way. possibility (◇)
 b. You *must/should/have to* / . . . go this way. necessity (□)

The experimental literature on children's modal comprehension suggests that they struggle with modal force until at least age 4: they tend to both accept possibility modals in necessity situations and necessity modals in possibility situations (e.g., Noveck 2001; Ozturk and Papafragou 2015). Typically, these errors are attributed to reasoning difficulties: children over-accept possibility modals in necessity situations because of difficulties reasoning about when a stronger modal would be more appropriate (i.e., they have trouble with scalar implicatures); they accept necessity modals in possibility situations because of difficulties reasoning about open possibilities (Acredolo and Horobin 1987). Usually, these studies take for granted that children already know the underlying force of modals. However, children's difficulties could reflect a lack of knowledge of their underlying force. In this paper, we address more directly the questions of when and how children figure out modal force by investigating modal talk to and by young children with a corpus study of the Manchester Corpus of UK English (CHILDES database, MacWhinney 2000; Theakston et al. 2001) and four experiments based on a modified version of the *Human Simulation Paradigm* (HSP; Gillette et al. 1999), testing how well adult participants can guess the force of modals uttered by either children or their mothers from the conversational context alone.

Imagine a child who hears a new modal, *sig*: 'You *sig* go this way.' How can she determine whether *sig* expresses necessity or possibility? The modal's syntactic position, before a verbal complement, might help narrow candidate meanings to expressing some kind of modal meaning (in the spirit of Landau and Gleitman's 1985 syntactic bootstrapping hypothesis), but it cannot help distinguish force, since possibility and necessity modals can appear in all the same syntactic environments. Cues from the physical context are also bound to be limited, since modals express non-actual concepts, with few physical correlates (Landau and Gleitman 1985). To learn the force of modals, children might thus need to rely heavily on cues from the conversational context. But how informative is the context about modal force?

One issue that might make this mapping of modal form to force particularly challenging is that necessity entails possibility: If you *must* go this way, considering the range of passable routes, then you *can* go this way, given that same set of options. Likewise, if you *must* eat with your right hand, given the rules of etiquette, then those same rules imply you *can* eat with your right hand. So, if you think that *sig* means 'possible' but in fact it means 'necessary', it is unclear how you can discover that in fact, *sig* has a stronger meaning: in situations where a necessity modal is used, a possibility statement is also systematically true. What then prevents learners from postulating possibility meanings for necessity modals like *must* or *have to*? This kind of *subset* or *entailment problem* arises whenever two words' meanings enter into a set/subset relationship, and has been discussed for content words, like *dog/animal* (e.g., Xu and Tenenbaum 2007), and quantifiers, like *some/every* and numerals (e.g., Piantadosi 2011; Piantadosi et al. 2012a; Rasin and Aravind 2020). In this paper, we focus on this issue for modals.

Different types of solutions have been proposed in the literature for how learners resolve or sidestep subset problems.¹ The first one is for them to rely on downward-entailing (DE) environments, which reverse patterns of entailment, as Gualmini and Schwarz (2009) suggest as a general solution to subset problems. The second is a bias toward strong (here, necessity) meanings, in the spirit of Berwick (1985). The third one is for the conversational context in which modals occur to be rich enough for learners to infer their force, without having to rely on either DE environments or a necessity bias (Dieuleveut et al. 2019a; Rasin and Aravind 2020 for *every*).

According to the first solution, all that children need to solve the subset problem is to observe necessity modals in DE environments, for instance under negation, as these environments reverse patterns of entailment (*not possible* entails *not necessary*). If children hear ‘You *don’t have to* go this way’ in a situation where it is clear that there are other ways to go, they should be able to infer that *have to* doesn’t express possibility: if it did, its negation would mean *impossible*, and wouldn’t allow for other ways. We argue that this is not a viable solution for modals. First, our corpus results show that necessity modals rarely occur with negation, let alone in other DE environments, in the actual input to children. Second, problems arise from the fact that scope relation between modals and negation are idiosyncratic. Necessity modals do not uniformly scope under negation: *have to* does, but *must* and *should* do not (Iatridou and Zeijlstra 2013). Third, instances where necessity modals do occur with negation are the least informative about their force: our experimental results show that participants had the most difficulty guessing the force of necessity modals in negative contexts.

If learners cannot clearly rely on DE environments, they may need a bias toward necessity meanings. According to this second solution, children would assume necessity meanings by default and revise their hypothesis only for possibility modals, when hearing them used in situations of non-necessity. This kind of solution, proposed for other instances of the *subset problem*,² has been criticized by many authors, both on conceptual (Gualmini and Schwarz 2009) and empirical grounds (Musolino 2006; Pi-

¹In this paper, we put aside two other possibilities. First, children could learn force on the basis of explicit negative feedback: they would be corrected by adults when they produce necessity modals to describe mere possibilities. We leave a systematic investigation of whether children receive such feedback, but, anecdotally, we haven’t encountered any. In their study on *every*, Rasin and Aravind (2020) report only one case of explicit negative feedback out of 72 *every* uttered by children. Second, children could exploit the principle of contrast (Clark 1987) (e.g., knowing a possibility modal could lead children to hypothesize a necessity meaning for the next modal they encounter). However, as we discuss in Sect. 2.1, learners cannot expect two modals to necessarily express different forces, since several modals express the same force. If children heard necessity claims explicitly contrasted with possibility claims (e.g., “you *can*, but you *don’t have to*”), they could infer that the former expresses possibility, and the latter necessity, based on logical compatibility. While we did not systematically examine the transcripts for such cases, we did extract all cases where two modal claims were related by the logical connectives *or*, *but*, and *and*. We found only 85 such cases (out of 18,853 utterances). Among these, only three were possibly useful to figure out the force of possibility modals (e.g., “it *might* be in your box or it *might* be lost”), but we found no informative cases for necessity modals. This may be an underestimation, since we did not look for the occurrence of modals across different utterances. However, this very low frequency makes it unlikely that children can learn from such contrasts.

²Many variants of this idea can be found in the literature. The Subset Principle (Baker 1979; Berwick 1985; Dell 1981; Manzini and Wexler 1987; Pinker 1979, a.o.) was originally proposed for the acquisition of syntactic phenomena. Later on, the Semantic Subset Principle (SSP) was introduced by Crain and

antadosi 2011; Piantadosi et al. 2012a; Rasin and Aravind 2020; Xu and Tenenbaum 2007; for a summary, see Musolino et al. 2019). But such a bias could be necessary in the case of modals: there may, for instance, be fewer visual cues about their meanings than for concrete objects or even quantifier meanings, since they express abstract concepts about the non-actual.

In this paper, we argue that a necessity bias may not be necessary, even in the case of modals, and that the subset problem is in principle solvable based solely on cues stemming from the conversational context in which modals occur (Dieuleveut et al. 2019a). Rasin and Aravind (2020) reach a similar conclusion for *every*: while truth-conditional evidence alone may not allow the child to block an existential meaning (e.g., *some*) for *every*, pragmatic evidence can play an important role in sidestepping it. Rasin and Aravind consider potential sources of truth-conditional evidence against an existential meaning: DE environments, non-monotonic environments, and environments where the existential quantifier cannot be used (NPI-licensing, *almost*, *nearly*, and exceptive *but*). They show that such cases are very rare in the input (1.75% of *every* utterances), but that informative pragmatic evidence for rejecting an existential meaning for *every* is systematically available. Indeed, often enough (17%) when *every* is used in questions or assertions, its existential counterpart is already part of the common ground (as determined by the two authors). Hence, if *every* had an existential meaning, its use would result in a trivial contribution (e.g., asking “is *someone* here?” or asserting “*someone* is here” in a context where someone has arrived). They conclude that if children assume that speakers do not make vacuous contributions, they could infer that *every* does not have an existential meaning and thus sidestep the subset problem without special learning principles.

In this study, we probe the role of context in a different way, by having naïve speakers guess the force of modals based on the context of use, expanding on Dieuleveut et al. (2019a).³ Our experimental results show that this conversational context is informative about modal force: for the most part, participants were able to accurately recover the force of both possibility and necessity modals from mere snippets of conversation. Thus, in principle, learners should be able to figure out modal force based on cues from the conversational context alone and solve the subset problem without having to rely on DE contexts or on a necessity bias. But while a necessity bias is in principle not necessary on the basis of the input, children may still make use of one in practice. However, we find no evidence for a necessity bias in young children’s modal productions.

Our current understanding of young children’s modal force use is limited. Comprehension studies (e.g., Hirst and Weil 1982; Byrnes and Duff 1989; Noveck 2001; Ozturk and Papafragou 2015, a.o.) tend to focus on older children. Corpus studies (e.g., Kuczaj and Maratsos 1975; Papafragou 1998; Courname 2015, 2021; van

Thornton (1998) to account for *semantic set/subset* problems at the sentential level (see also Crain et al. 1994; Crain 2012).

³We were unaware of Rasin and Aravind’s methodology when we started this project with Dieuleveut et al. (2019a) and thus approached the problem for modals differently. We leave an application of Rasin and Aravind’s methodology to modals for future research, though we suspect that it might not always be easy to tell whether a possibility is already part of the common ground when a necessity statement is made, as all kinds of possibilities may be open at any given point in a conversation, even if they may not be salient.

Dooren et al. 2017) tend to focus on modal flavor, and while they note when particular lexemes first appear in children's productions, to date no study systematically examines modal force in naturalistic productions (but see Jeretič 2018; Dieuleveut et al. 2019a). In this paper, we provide the first large-scale study of the development of modal force by examining the modal production of 12 children between the ages of 2 and 3. Our corpus and experimental results on children's modals indicate an asymmetry in force acquisition. Children seem to master possibility modals early: at age 2, children use them frequently and productively, both with and without negation. And they use them in an adult-like way: crucially, they do not use them in necessity situations. However, they seem to struggle with necessity modals. They produce these much less frequently and, often, in a non-adult-like way: they use them in situations where adults prefer possibility modals. If this difficulty with necessity modals persists into the preschool years, it could explain children's tendency in prior comprehension studies to both accept possibility modals in necessity contexts (they may lack a relevant stronger alternative) and necessity modals in possibility contexts (they may not be sure that these modals express necessity).

Together, our results from mothers' and children's productions seem to lead to a puzzle: if the conversational context is informative about both forces, why should children particularly struggle with necessity modals? The early advantage for possibility modals could be due to a combination of factors. First, possibility modals are more frequent than necessity modals in children's input (about 3:1). Second, situations in which possibility modals occur with negation seem to be particularly informative (e.g., prohibitions, impossibilities), while negation may be particularly misleading with necessity modals. Whatever the reason for children's difficulty with necessity modals, their successes with possibility modals and relative failures with necessity modals provide no evidence for a necessity bias. Given that a necessity bias is neither necessary, in view of the information available in the input, nor evidenced in children's productions, we suggest that it is dispensable, even in the case of modals.

The rest of this paper is structured as follows. In Sect. 2, we provide some general background on modal force and its acquisition. We first give a brief overview of the semantics and pragmatics of modals in English and beyond, particularly as they relate to force, and discuss the possible learnability implications that these cross-linguistic considerations engender. We then turn to how modals interact with negation, and what this might entail for force acquisition. We then review the main relevant findings from the modal acquisition literature. In Sect. 3, we present our input study. We first provide a descriptive, quantitative assessment of the modals children hear: which modals occur and how often, and when they appear with negation and in other DE environments. We then present three input-based experiments that assess the general informativity of natural conversational contexts about modal force, by asking adult participants to guess a modal blanked out from a dialogue extracted from the corpus, following a modified version of the HSP (Gillette et al. 1999). In Experiment 1, the blanked modal statement is presented in context (seven preceding lines of dialogue); in Experiment 2, it is presented without the dialogue, with only the target sentence; and in Experiment 3, all content is removed, as the sentence is presented with content words replaced by nonce 'Jabberwocky' words. Our results show that the conversational context modals are used in is highly informative about both forces. We then

probe what aspects of the conversational context might be helpful and identify one feature in particular for root modals, namely the desirability of the prejacent. A fourth experiment confirms that necessity—but not possibility—modals are typically used with undesirable prejacentes. In Sect. 4, we turn to children’s productions. We first provide a quantitative assessment of the modals they produce and then present a fifth experiment that assesses the extent to which children use their modals in an adult-like way, by asking adult participants to guess the force of modals used by children. Our results suggest that children master possibility modals early but struggle with necessity modals. In Sect. 5, we discuss implications of our findings for how modal force acquisition might unfold in English and beyond. Section 6 concludes.

2 Background

2.1 Modal force in English and beyond

English modals come in two main forces: possibility and necessity. This is standardly captured by treating modals as existential or universal quantifiers over possible worlds, following the modal logic tradition. Further force distinctions can be found: necessity modals can be split into strong (*must*) vs. weak (*should*) necessity (von Stechow and Iatridou 2008);⁴ nouns (*slight possibility*) and adjectives (*likely*) can encode even finer-grained strength distinctions. Here we focus on the main contrast between possibility and necessity modals and the learnability issues that it gives rise to.

Modals can be used to express different flavors of modality: epistemic modals (as in (2)) express possibilities and necessities given some evidence; deontic modals express possibilities and necessities given some relevant rules (as in (1)). We use the term ‘root’ modality (Hoffmann 1966) for all non-epistemic flavors. This distinction matters for us in that root modals tend to pattern together and differently from epistemic modals in their interactions with scope-bearing elements, notably negation.

- | | | | |
|-----|----|----------------------|----------------------------|
| (2) | a. | It might be raining. | possibility (\diamond) |
| | b. | It must be raining. | necessity (\square) |

In English, a modal always expresses the same force (possibility or necessity). However, it can be used for different flavors: *Jo must draw* can express an epistemic necessity (‘Jo is likely to draw’) or a teleological, bouletic, or deontic necessity (‘Jo needs/wants/is required to draw’). This is captured in a classical Kratzerian framework (Kratzer 1981, 1991) by having modals be lexically specified for force but not for flavor. Flavor gets determined by conversational backgrounds which specify the set of worlds that the modal quantifies over, as the lexical entries, slightly modified from Kratzer (1991), illustrate in (3).

⁴The difference between weak and strong necessity is illustrated in the following example: ‘Employees must wash their hands. Everyone else should’ (von Stechow and Iatridou 2008). Weak necessity modals are still treated as necessity modals, but quantify over a smaller domain than their strong counterparts.

- (3) For any world w , conversational background f :⁵
- $[[\text{can}]]^{w,f} = \lambda q_{<\text{st}>}. \exists w' \in \cap f(w): q(w')=1$
 - $[[\text{must}]]^{w,f} = \lambda q_{<\text{st}>}. \forall w' \in \cap f(w): q(w')=1$

According to Horn (1972), modals form scales ($\langle \text{can}_{\text{deontic}}, \text{have to}_{\text{deontic}} \rangle, \langle \text{might}_{\text{epi}}, \text{must}_{\text{epi}} \rangle, \dots$),⁶ and as such, they give rise to scalar implicatures. The use of (1a), for instance, can implicate that you don't have to go this way; the use of (2a) can implicate that it doesn't have to be raining. In the Gricean tradition (Grice 1975), this implicature arises from the assumption that the speaker is trying to be maximally informative but is not in a position to assert the relevant stronger statement in (2b). Speakers should prefer to use *must p* whenever they believe it to be true and relevant: listeners can then infer from the fact that the speaker did not chose the stronger (more informative) sentence that she must not believe it.

In Indo-European languages like English, possibility and necessity duals are common. However, various languages seem to lack such pairs. Instead, the same 'variable force' modals can be used in situations where English speakers would either use a possibility or necessity modal. Analyses vary in how to capture these variable force behaviors (see Yanovich 2016 for a summary). In St'át'imcets and Washo, modals have been analyzed as underlyingly necessity (universal) modals that can be weakened by contextually restricting their domain of quantification (Rullmann et al. 2008; Bochnak 2015). In Nez Perce, the modal *o'qa* has been analyzed as a possibility (existential) modal whose apparent variable force is due to the lack of a lexicalized stronger necessity dual in the language: *o'qa* does not belong to a Horn-scale, therefore its use is never associated with a scalar implicature (Deal 2011). Gitksan *ima* is similarly analyzed as a possibility modal (Peterson 2010; Matthewson 2013).⁷

Turning back to our learning problem, the range of cross-linguistic variation we find suggests that there may be few constraints on the space of hypotheses learners have to entertain for modals. They can't expect modals to come in duals, or that their language must have a possibility modal or a necessity modal. And even in a language with duals like English, knowing the force of one modal doesn't guarantee that the next modal will express a different force, given that several lexemes can express the same force (e.g., *can*, *might*, and *may*): children will thus have to figure out force for each modal anew.

⁵We ignore the ordering source here, which can derive further gradability and flavor differences among root modals.

⁶Logical entailment relations hold within flavor only: for example, epistemic necessity (e.g., 'given what we know, he *must* be upstairs') does not entail deontic possibility (e.g., 'given the rules, he *can* be upstairs'). Horn scales are thus defined within a flavor. Because of flavor variability, this means that the same lexeme can appear in different scales. We leave aside debates about ability modals, often argued to have no necessity counterpart (Horn 1972; Hackl 1998) (e.g., 'Jo *can* speak German; in fact, he *has to*' leads to oddity, or forces a switch in flavor interpretation). It is also argued that ability modals do have duals, *compulsion* modals, but they are just extremely infrequent (e.g., 'I *have to* sneeze') (Mandelkern et al. 2015).

⁷Other analyses take variable modals to neither be underlying possibility nor necessity. Kratzer (2012) analyzes them as upper-end degree modals, roughly equivalent in meaning to 'it is somewhat probable (desirable) that p'.

One aspect of the English modal system that could indirectly help the learner is that speakers may refrain from using possibility modals in necessity situations, since necessity modals would be more informative. If the situations in which possibility modals are used never overlap with those in which necessity modals are used, this could help English learners distinguish possibility from necessity modals. However, the extent to which adults always choose necessity modals over possibility modals in necessity situations is not entirely clear. Speakers do not always aim for maximal informativity: other conversational principles intervene. Possibility modals can be used, for instance, to soften statements in a polite way: ‘You *could* be a little quieter’ can be used as an order to be quiet, or ‘It *might* be too late’ to convey that it *is* too late (Searle 1969; Austin 1975; Grice 1975; Brown and Levinson 1987, a.o.). Note that these politeness considerations are peculiar to modals and do not arise, for instance, with quantifiers over individuals. If frequent enough, they could blur the distinction between possibility and necessity modals and be particularly misleading. One of our main goals here is to find out how clear the input is about the underlying force of modals in speech to children.

We now turn to the interaction of modals with negation and discuss the extent to which negative environments can help or hinder learners to figure out modal force.

2.2 Modals and negation

Sentences containing modals and negation can in principle receive two interpretations: a ‘strong’ interpretation (*not* > *possible*, logically equivalent to *necessary* > *not*), and a ‘weak’ interpretation (*possible* > *not*, logically equivalent to *not* > *necessary*). Cross-linguistically, epistemic possibility modals tend to be interpreted above negation and root possibility modals below (Coates 1988; Cinque 1999; Drubig 2001; Hacquard 2011; for a typological overview, see De Haan 1997; van der Auwera 2001). This is illustrated for English in (4a), (4b), and (4c): (root) *can* is interpreted below negation, and (epistemic) *might* is interpreted above negation; *may* is interpreted under negation with a root interpretation and over negation with an epistemic interpretation.

- (4) a. Jo *can*'_{root} draw. ¬ ◇ * ◇ ¬
 b. Jo *might*_{epistemic} *not* draw. *¬ ◇ ◇ ¬
 c. Jo *may*_{root/epistemic} *not* draw.
 root: ‘it is **not possible** that Jo draws’ ¬ ◇ * ◇ ¬
 epistemic: ‘it is **possible** that Jo does **not** draw’ *¬ ◇ ◇ ¬

Necessity modals, on the other hand, seem to keep the same scopal behavior with respect to negation, regardless of flavor: they either systematically scope over negation, like *must/should* in (5a) (Dutch *moeten*, German *müssen*) (a behavior Iatridou and Zeilstra 2013 attribute to their being Positive Polarity Items), or under negation, like *need* in (5b) and *have to* in (5c). English *need*, Dutch *hoeven*, and German *brauchen* are commonly analyzed as a Negative Polarity Items (NPIs).

- (5) a. Jo *must not/should not* draw. □ ¬ * ¬ □
 epistemic/root: ‘it is **necessary** that Jo does **not** draw’

- b. Jo *needn't* draw. *□ ¬ ¬ □
epistemic/root: 'it is **not necessary** that Jo draws.'
- c. Jo *doesn't have to* draw. *□ ¬ ¬ □
epistemic/root: 'it is **not necessary** that Jo draws.'

Thus, modals are not uniform in their interaction with negation, force-wise or flavor-wise. This means that for at least some of the modals children have to learn, using negation to infer their force will be problematic. If they expect negation to scope over all modals by default (regardless of force and flavor), cases like (4b) and (4a) will be problematic: (4b) could suggest a necessity meaning for *might* (*need not* ~ *might not*), and (4a) could suggest a possibility meaning for *must* (*can't* ~ *mustn't*). If learners expect negation to scope over root modals but under epistemic modals (given some more general assumptions about flavor and scope),⁸ (4b) is no longer problematic, but (5a) still is. Alternatively, if learners initially assume strong interpretations for any negated modal sentence (following Crain and Thornton's 1998 *Semantic Subset Principle*; see Moscati and Crain 2014),⁹ cases like (5b), (5c), and (4b) will be problematic. For negation to be helpful in figuring out a modal's force, learners would need to have already figured out how the modal scopes relative to negation and expect negation to scope differently based on force and flavor. However, it is not clear how they would figure out the right scope relations between modals and negation without knowing the force of the modals.

In the next section, we briefly review findings about children's understanding of modal force and its interaction with negation from the acquisition literature.

2.3 Modal force acquisition

Possibility modals like *can* are found early in child productions, by age 2. The literature reports an asymmetry, with root modals appearing earlier than epistemics (Kuczaj and Maratsos 1975; Papafragou 1998; Cournane 2015, 2021; van Dooren et al. 2017).¹⁰ Experimental work on children's comprehension usually targets older children (age 4 and up) (Hirst and Weil 1982; Byrnes and Duff 1989; Noveck et al. 1996; Noveck 2001; Ozturk and Papafragou 2015, a.o.) and focuses on epistemic flavor, using felicity judgment tasks, where children have to judge whether a possibility or a necessity statement is true in scenarios where a toy is hidden in one of two boxes. By age 4, children seem to be sensitive to the relative force of modals, when the contrast is made salient by the experimental design, but they still do not behave like adults. First, they tend to over-accept possibility modals when necessity modals are more appropriate (Noveck 2001; Ozturk and Papafragou 2015). This is traditionally blamed on general difficulty with *scalar implicatures* (Chierchia et al. 2001; Barner and Bachrach 2010; Barner et al. 2011; Skordos and Papafragou 2016, a.o.): children have trouble accessing the relevant alternatives that the speaker takes

⁸See, for example, Cinque's (1999) hierarchy.

⁹Such considerations might explain why necessity but not possibility modals tend to be PPIs.

¹⁰This asymmetry has been attributed to conceptual and grammatical factors, but it might instead reflect a frequency asymmetry in the input. For how children learn that modals can be used to express various flavors, see van Dooren et al. (2017, 2022).

for granted and using them to understand the implicature when asked to judge sentences in isolation. Second, children also tend to accept necessity modals in possibility situations (Ozturk and Papafragou 2015; Koring et al. 2018), a perhaps more surprising result from an adult's perspective: whereas possibility modals are underinformative but logically true in necessity situations, necessity modals are false in possibility situations. Ozturk and Papafragou (2015) argue that children's difficulty with necessity modals stems from (non-linguistic) difficulty reasoning about indeterminate events: in reasoning tasks that introduce indeterminacy, children tend to commit to a possible conclusion before decisive evidence is available and arbitrarily select one possibility over the other (a tendency sometimes referred to as *premature closure*; Bindra et al. 1980; Piéraut-Le Bonniec 1980; Acredolo and Horobin 1987; Robinson et al. 2006).

A few experimental studies focus on children's interpretation of sentences containing negated modals (Moscati and Gualmini 2007 [*can*]; Gualmini and Moscati 2009 [*need* and Italian *dovere* 'must']; Moscati and Crain 2014 [Italian *potere* 'can']; Koring et al. 2018 [Dutch *hoeven* 'need']). Children tend to prefer strong interpretations of negated modal sentences (*not*>*possible* or *necessary*>*not*), even when adults prefer weak ones (*possible*>*not* or *not*>*necessary*). These studies assume that children already know the underlying force of their modals and focus on their scope relative to negation. However, children's non-adult-like responses could, in principle, be explained by not knowing the force of the modals involved. For instance, one predicts the same responses for Italian *potere non* (where the possibility modal scopes over negation, leading to a weak interpretation) if children assume that *potere* expresses possibility and negation scopes over the modal or if they assume that *potere* expresses necessity and negation scopes under the modal.

We now turn to our studies, which probe more directly the questions of when and how children figure out modal force by investigating their modal input (Sect. 3) and their early productions (Sect. 4).

3 Children's modal input

The goal of this study is to provide an analysis of the modals children are exposed to. We first present quantitative results from a corpus study: how are possibility and necessity modals distributed in actual speech to children? How frequently do they occur with negation? We then present four experiments, using corpus data, aimed at assessing the informativity of the conversational context as to force. In Experiment 1, based on the HSP (Gillette et al. 1999), participants have to guess the force of a missing modal in dialogues extracted from the corpus, allowing us to assess the general informativity of conversational contexts depending on force, negation, and flavor (epistemic vs. root). Experiment 2 isolates the role of context from possible biases toward possibility or necessity meanings, by showing participants the blanked modal sentence without the dialogue, while Experiment 3 shows the same sentence with all content words replaced by 'Jabberwocky' nonce words. Experiment 4 focuses on a particular feature of the context, namely the desirability of the prejacent as a cue to force for root modals.

3.1 Corpus study

3.1.1 Methods

We used the Manchester Corpus (Theakston et al. 2001) of UK English (CHILDES database, MacWhinney 2000), which consists of 12 child-mother pairs (six females; age range: 1;09-3;00) recorded in unstructured play sessions. We chose this corpus for its relative density and uniformity of sampling and early age range. We focused on the period between ages 2;00 and 3;00. All utterances containing modal auxiliaries and semi-auxiliaries (26,598 of 564,625 total utterances; adult: 20,755; child: 5,842; excluding repetitions [6.6%]: adult: 19,986; child: 4,844) were coded for force (possibility vs. necessity) (6), presence of negation (7), and flavor (epistemic vs. root) (8).¹¹ We did not include *will*, *would*, *shall*, or *going to* as they primarily express future, for which force is a matter of debate (Stalnaker 1978; Cariani and Santorio 2017, a.o.).

- (6) Modal lemmas by force:
 Possibility: *can*, *could*, *might*, *may*; *able to*
 Necessity: *must*, *should*, *need*; *have to*, *got to*, *be supposed to*, *need to*
- (7) Negation:
 No negation: ‘I **can** go to the pub now.’
 Negation:
 on main verb: ‘I **can’t** get it.’/‘I **must not** forget Whispy.’
 on higher auxiliary: ‘we **don’t** have to play with your toys.’
 on embedding verb: ‘I **don’t think** you have to look for it.’
 other negative quantifier: ‘**nobody** can reach it.’
- (8) Flavor:
 Root:
 MOTHER: we won’t do that.
 CHILD: I want her.
 CHILD: I want her.
 MOTHER: well you **must** treat her nicely then. (Aran, 2;07,14)
 Epistemic:
 MOTHER: oh.
 MOTHER: somebody’s done a neat pattern, haven’t they?
 MOTHER: goodness me.
 MOTHER: that **must** have taken a long time. (Anne, 2;02.10)

3.1.2 Results

We found that, overall, possibility modals were more frequent than necessity modals in adult speech: they represent 72.5% of all adults’ modal utterances (Table 1). This is mostly driven by the highly frequent modal *can*.

¹¹We do not differentiate among subtypes of root flavors (ability, teleological, deontic, ...).

Table 1 Counts and percentages of modal uses by force for adults, ordered by lemma frequency, with and without negation (repetitions excluded: 3.7% of the data)^{a,d}

	ADULT (<i>n</i> = 19,986)		ADULT (<i>n</i> = 18,853) ^b			
	all		no negation		negation	
POSSIBILITY	14,491	72.5%	10,672	79.1%	2,828	20.9%
<i>can</i>	11,472	57.4%	8,383	77.7%	2,396	22.2%
<i>could</i>	1,449	7.3%	1,116	96.6%	39	3.3%
<i>might</i>	1,216	6.1%	1,005	82.8%	208	17.1%
<i>able</i>	315	1.6%	134	42.5%	181	57.4%
<i>may</i>	39	0.2%	34	89.5%	4	10.5%
NECESSITY	5,495	27.5%	4,814	89.9%	539	10.1%
<i>have to</i>	2,398	12.0%	2,290	95.5%	108	4.5%
<i>got to</i>	940	4.7%	926	98.8%	11	1.1%
<i>should</i> ^d	793	4.0%	537	77.1%	159	22.8%
<i>need (to)</i> ^c	493	2.5%	409	82.9%	84	17.0%
<i>must</i> ^d	452	2.3%	346	84.1%	65	15.8%
<i>supposed to</i>	335	1.7%	230	68.6%	105	31.3%
<i>ought to</i> ^d	84	0.4%	76	91.5%	7	8.4%

^aWe considered as repetitions cases where the speaker repeated a sentence uttered right before by herself or another speaker with no significant change

^bExcluding tags and repetitions. Tag questions (e.g., ‘you can wash it later, *can't you?*’) are very frequent in this corpus (4.7% of all modal utterances). We decided to exclude modals in tags, as they do not directly matter for our purposes

^cThere are only five occurrences of the NPI *need* (e.g., ‘you needn't whisper’)

^dIndicates necessity modals that outscope negation

Turning to negation and other DE contexts, we found that possibility modals occur with **negation** more frequently than necessity modals (possibility: 20.9% vs. necessity: 10.1%). For necessity modals, negation occurs proportionally more frequently with modals that out-scope negation (*should*: 22.8%; *must*: 15.8% vs. *have-to*: 4.5%; *got-to*: 1.1%). Modals rarely occur with other **negative quantifiers** (e.g., *nothing/never*), with no difference between possibility and necessity (possibility: 0.2%; necessity: 0.1%), or under a negated embedding verb (e.g., *don't think*), again with no difference between possibility and necessity (possibility: 1.5%; necessity: 2.1%). Details of negative environments are provided in online Appendix A (Table 10). We further found that modals are extremely rare in **antecedents of conditionals** (0.6% of adults' modal utterances). Necessity modals almost never occur in such environments: we found only 15 occurrences in the whole corpus (106 possibility modals), with seven of them corresponding to ‘if you must’. As a point of comparison, 135 necessity modals occur in consequents of conditionals vs. 432 possibility modals. A breakdown by modal is provided in online Appendix A (Table 11).

Overall, epistemic uses of modals are rare: they represent only 8.8% of all adults' modal utterances (Table 2). Negation is significantly more frequent on root than on

Table 2 Counts and percentages of modal uses, by force, flavor, and negation, for adults (excluding tags and repetitions)

	ADULT (<i>n</i> = 18,853)					
	all		no negation		negation	
root	17,190	91.2%	13,896	80.9%	3,293	19.1%
possibility	12,175	64.6%	9,414	77.3%	2,761	22.6%
necessity	5,015	26.6%	4,482	89.4%	533	10.5%
epistemic	1,662	8.8%	1,590	95.4%	73	4.6%
possibility	1,324	7.0%	1,257	94.9%	67	5.0%
necessity	341	1.8%	332	97.3%	6	2.6%

epistemic modals (epistemic: 4.6% vs. root: 19.1%). A breakdown by modal is provided in online Appendix A (Table 12).

3.1.3 Interim discussion

Overall, possibility modals are more frequent than necessity modals in mothers' speech: children thus have more opportunities to learn them. The relative rarity of necessity modals may be due to the alternative ways speakers can express necessity (e.g., using imperatives for deontic necessity or asserting the prejacent directly for epistemic necessity).

Necessity modals rarely appear in DE environments. Negation is infrequent with necessity modals: only 10.1% of all necessity modals cooccur with negation (vs. 20.9% of possibility modals), and necessity modals are exceedingly rare in antecedents of conditionals.

3.2 Experiment 1: Adults' modal productions

To assess the general informativity of natural conversational contexts about force, we implemented a variant of the HSP (Gillette et al. 1999) using dialogue contexts extracted from the corpus. The goal of the original HSP (Gillette et al. 1999; see also Snedeker 2000; Snedeker et al. 1999; White et al. 2016) is to compare the effect of different kinds of contextual information on the ability to recover a word's meaning: extralinguistic scenes, associated words and morphemes, or syntactic-frame information. The accuracy with which participants can recover the actual word given the context is taken as a general measure of informativity of properties of that context. Following Orita et al. (2013), we use the paradigm in a slightly different way: participants were given only written transcripts from the corpus (with no visual or acoustic information) and had to choose between a possibility and a necessity modal.¹² This allows us to, first, give a general measure of the informativity of conversational context about force: can naïve subjects guess the force of a blanked-out modal based

¹²Results from another HSP study where participants had to fill in the blank (instead of making a forced choice) are reported in Dieuleveut et al. (2019b).

Table 3 Summary of experimental conditions

Test condition (between participants)		Modal lemmas	
		possibility	necessity
EPI-AFF (epistemic affirmative)		<i>might</i>	<i>must</i>
ROOT-AFF (root affirmative) ^a	ROOT-AFF-1	<i>can</i>	<i>must</i>
	ROOT-AFF-2	<i>can/able</i>	<i>have to</i>
ROOT-NEG (root negative)		<i>can't/not able</i>	<i>not have to</i>

^aWe implemented two versions of the **ROOT-AFF** condition. **ROOT-AFF-1** (*can* vs. *must*) allowed us to keep the syntactic category of both options identical, while **ROOT-AFF-2** (*can/able to* vs. *have to*) allowed us to avoid concerns related to the formality of *must* for US English speakers. In cases where *have to* was tensed, we used *able to* as the alternative to avoid losing tense information: for example, participants had to choose between [*will have to*] and [*will be able to*]. We extracted the same number of contexts from *able to* to avoid having the *able to* option always be the wrong answer. Same principles applied for the **ROOT-NEG** condition: participants had to choose between [*didn't have to*] and [*wasn't able to*] when *have to* was tensed

solely on excerpts of conversations in which it appears? Second, we can test directly for interrelationships between force and negation: are contexts equally informative for both necessity and possibility modals? Are negative contexts more informative than positive contexts?

3.2.1 Methods

Procedure. The experiment was run online on IBEX Farm.¹³ Participants recruited via Amazon MechanicalTurk were asked to guess a redacted modal in a dialogue between a child and mother by choosing between two options, corresponding either to a possibility (e.g., *might*) or a necessity (e.g., *must*) modal, as illustrated in Fig. 1a. All dialogue contexts consisted of the modal sentence with a blank and seven preceding utterances, with two options displayed at the bottom of the screen. There was first a short training where participants had to choose between definite and indefinite articles (*the* vs. *a*) (three examples with feedback), followed by the test phase without feedback. Overall, each participant had to judge 40 different dialogues (20 trials: 10 possibility, 10 necessity; 20 controls using tense: 10 past, 10 future), presented in random order. The 20 trials were selected randomly for each participant from a list of 40 contexts originally extracted from the corpus; the 20 controls were the same for all participants. Further details of the instructions and material are provided in online Appendix B.

Conditions. We tested force (possibility vs. necessity) within participants, and flavor (root vs. epistemic) and negation (present vs. absent) between participants. Negation was tested only for root flavor because negated epistemics were too rare in the corpus (Table 2). Table 3 summarizes the experimental design.

¹³An example of the experiment can be accessed at <https://frm.pcibex.net/r/HDtxJP/>.

MOTHER: don't wake her up.
 MOTHER: she's fast asleep.
 CHILD: only tickling her.
 MOTHER: no.
 MOTHER: don't.
 MOTHER: she's fast asleep, look.
 MOTHER: isn't she?
 MOTHER: she _____ have had a good night last night wherever she went.



Fig. 1a Experiment 1 stimuli: example trial (EPI-AFF, *must*)

Fig. 1b Experiment 1 stimuli:
 control trial (*saw*)

CHILD: what did we do to park?
 MOTHER: what did we do at the park?
 CHILD: yeah.
 MOTHER: we had a walk by the lake.
 CHILD: yeah.
 MOTHER: didn't we?
 CHILD: yeah.
 MOTHER: and we _____ some ducks.



Material. Extraction procedure—160 contexts (2*20 per condition) were randomly extracted from the corpus for each modal (*can*, *able*, *might*, *must*, *have to*). **Exclusion criteria**—We excluded contexts where the adult or child used the target modal in preceding utterances. Contexts were not excluded when the adult or the child used another non-target modal. Criticisms, such as *willn't*, were removed from the dialogue and replaced with the American English equivalent (e.g., *won't*). We didn't exclude contexts where there were tag questions (e.g., '... mustn't she?'), but we removed tags when they occurred in the target sentence. **Controls**—Participants had to choose between future and past (e.g., [*saw*] vs. [*will see*]; see Fig. 1b). Importantly, the correct answer was not always guessable based on the target sentence alone: participants were required to read the entire dialogue. Extraction procedure and data cleaning were the same as for the targets. We excluded participants that were less than 75% accurate on controls.

3.2.2 Results

Participants. We recruited 289 participants on Amazon Mechanical Turk (four groups among participants: ROOT-AFF-1: 73, ROOT-AFF-2: 72; ROOT-NEG: 73; EPI-

AFF: 71; language: US English; 156 females, mean age = 40.6-years-old). We removed from analysis eight participants (2.8%) who were less than 75% accurate on controls. We thus present results for 281 participants (ROOT-AFF-1: 71, ROOT-AFF-2: 69; ROOT-NEG: 70; EPI-AFF: 71).

Analysis. All data analyses were conducted using R (R Core Team 2013) and the lme4 package (Bates et al. 2014). All R scripts for analysis are available at <https://osf.io/v9ure/>. Figure 2 summarizes the mean accuracy for each condition.¹⁴ Overall, participants were highly accurate at guessing modal force (general mean accuracy: 79.9%). We first ran binomial tests to see whether they differ from chance for each condition (Table 4). Participant accuracy significantly differs from chance in each condition. The lowest performance was found for ROOT-NEG necessity modals (e.g., *not have to*) (61.3%). **Force**—To test whether there was an effect of Force, we fitted a generalized linear mixed effects model, built with a maximal random effect structure, testing Accuracy (dependent variable, binomial), with Force as a fixed effect and Subject and Item as random factors (following Barr et al. 2013) (glmer syntax: $\text{Accuracy} \sim \text{Force} + (\text{Force} | \text{Subject}) + (1 | \text{Item})$),¹⁵ first overall and then subsetting the data for each condition. We compared these models with a reduced model without Force as a fixed effect ($\text{Accuracy} \sim 1 + (\text{Force} | \text{Subject}) + (1 | \text{Item})$).¹⁶ We found a general effect of Force, in the direction of a higher accuracy for possibility contexts (Model Comparison: $\chi^2(1) = 20.49, p = 5.9e-6^{***}$).¹⁷ Restricting to each comparison group, we found a significant effect in ROOT-AFF-1 ($\chi^2(1) = 61.1, p = 5.5e-15^{***}$) and ROOT-NEG ($\chi^2(1) = 15.6, p = 7.8e-05^{***}$), again in the direction of a higher accuracy for possibility contexts. In EPI-AFF, the effect of Force almost reached significance ($\chi^2(1) = 3.73, p = .053$ (NS)). In ROOT-AFF-2, it was not significant ($\chi^2(1) = 6e-04, p = 0.98$ (NS)). **Negation**—We compared ROOT-AFF-2 and ROOT-NEG, as these conditions included the same lemmas. We used the same method as above, comparing a model with Negation as a fixed effect and Subject and Item as random factors to a model without Negation as a fixed effect (Maximal model: $\text{Accuracy} \sim \text{Negation} + (1 | \text{Subject}) + (1 | \text{Item})$; for the interaction: $\text{Accuracy} \sim \text{Force} * \text{Negation} + (1 | \text{Subject}) + (1 | \text{Item})$). We found a strong interaction effect (**Interaction Force*Neg**: $\chi^2(1) = 7.9, p = 0.0047^{**}$). We found a significant effect of negation on necessity modals, which led to lower accuracy (*have to* vs. *not-have to*: $\chi^2(1) = 6.5, p = 0.011^*$). On possibility modals, negation led to higher accuracy, but the effect was not significant (*can* vs. *can't*: $\chi^2(1) = 2.29, p = 0.13$ (NS)). **Flavor**—There was no effect of flavor ($\chi^2(1) = 0.11, p = 0.74$ (NS)) (Maximal Model: $\text{Accuracy} \sim \text{Flavor} + (1 | \text{Subject}) + (1 | \text{Item})$).

¹⁴Accuracy for controls was very high (94.6%). There was no difference between groups in accuracy.

¹⁵We sometimes had to step back to random-intercepts-only models when the model failed to converge with the full random-effects specification.

¹⁶We used the same procedure (based on model comparisons) for all subsequent experiments reported in this paper, but don't systematically report the reduced model.

¹⁷Answers were coded as 1 if the response was accurate, and 0 otherwise. Effect sizes (β), standard errors (SE), z -values, and p -values for the logistic models are reported in online Appendix C for Experiment 1 and all following analysis.

Table 4 Accuracy rates and binomial tests by condition, testing for significance vs. chance (Experiment 1: adult productions) ($n = 281$, 10 observations per cell)

	Mean accuracy (se) ^a		Exact binomial tests (two-sided)	
	possibility	necessity	possibility	necessity
ROOT-AFF-1	91.7% (0.027)	71.7% (0.054)	$p < .001^{***}$ 95% CI [0.90, 0.94]	$p < .001^{***}$ 95% CI [0.68, 0.75]
ROOT-AFF-2	81.5% (0.053)	82.0% (0.052)	$p < .001^{***}$ 95% CI [0.79, 0.85]	$p < .001^{***}$ 95% CI [0.79, 0.84]
ROOT-NEG	89.5% (0.031)	61.3% (0.065)	$p < .001^{***}$ 95% CI [0.88, 0.92]	$p < .001^{***}$ 95% CI [0.56, 0.64]
EPI-AFF	87.2% (0.028)	74.3% (0.049)	$p < .001^{***}$ 95% CI [0.84, 0.90]	$p < .001^{***}$ 95% CI [0.71, 0.77]
Total	87.5% (0.018)	72.3% (0.028)		
ALL	79.9% (0.018)			

^aAccuracy corresponds to the mean accuracy (how good participants were to guess correctly the force of the modal, e.g., to answer *can* in a possibility context) across 20 contexts initially extracted from the corpus for each condition of force and flavor. Each participant saw only 10 out of the 20 contexts (10 for possibility, 10 for necessity). On average, each context was thus seen by 34.7 participants (ranging between 24 and 47). See online Appendix B, Table 12, for details

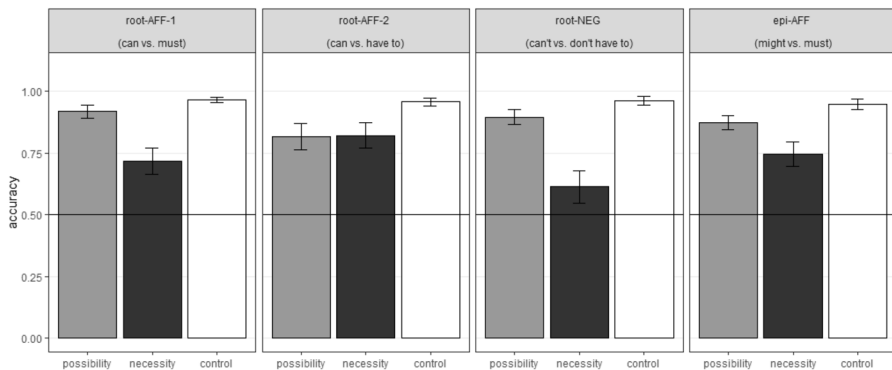


Fig. 2 Accuracy by condition (adult, $n = 281 * 10$)

Analysis by contexts (post-hoc). To get a sense of the kinds of contextual cues that were particularly helpful, we examined the contexts that led to the lowest and highest accuracy for both root and epistemic flavors. We focused on necessity modals, as there was more variability in accuracy for necessity modals. This informal analysis revealed two factors, depending on flavor: for root modals, cases where the proposition expressed by the prejacent seemed undesirable (e.g., going to the hospital) or effortful (e.g., lifting a heavy object) seemed to lead to high accuracy for necessity modals (see (9)). For epistemic modals, we found high accuracy for necessity modals in contexts that made salient robust evidence for the prejacent (see (10)). Our post-hoc analysis also pointed out a particularly high accuracy for possibility root modals

in interrogative sentences (e.g., *___ you see?*) (mean accuracy for root possibility modals in interrogative: 96.0%).¹⁸ In this case, accuracy may not reflect pure informativity, as participants may rely on idiomatic turns of phrases. However, they were still very accurate in contexts that did not involve interrogatives (mean accuracy for root possibility in declaratives: 76.3%).

- (9) CHILD: Mummy.
CHILD: Mummy.
MOTHER: Mummy?
CHILD: that Mummy.
MOTHER: what... what happened to Mummy?
CHILD: poorly.
MOTHER: she's poorly, is she?
MOTHER: she... she _____ go to the hospital.
(*have to*, 'undesirable'; mean accuracy: 96.6%)
- (10) MOTHER: ...
CHILD: yeah.
MOTHER: but Bertie was very close behind, wasn't he?
MOTHER: it was a near thing I think.
CHILD: he's lost his hat.
MOTHER: he has.
MOTHER: yeah.
MOTHER: it _____ have been windy eh?
(*must*, 'strong justification'; mean accuracy: 92.1%)

3.2.3 Interim discussion

Results from Experiment 1 show that the conversational context is informative about force: participants were able to guess the force of the modal accurately, just based on short conversation transcripts, and for both forces (general mean accuracy: 79.9%; possibility modals: 87.5%; necessity modals: 72.3%). This means that the information is there, at least in principle, for learners to figure out modal force based on context alone. If children are sensitive to the same cues as adults, they may not need to rely on negation or on a bias toward necessity meanings to figure out force.

Multiple factors may play a role in making the conversational context useful for guessing the right modal force: situational cues (e.g., who the interlocutors are), cues from world knowledge (e.g., what is allowed or prohibited), or pragmatic cues (what the speaker is trying to achieve, in particular performing orders, permissions, or prohibitions). Our post-hoc exploration suggests that cues may vary based on modal flavor. It appears that the (un)desirability and effortfulness of the prejacent could be particularly useful for roots, and some explicit justification for epistemics. We probe the effect of desirability more directly in Experiment 4 below.

¹⁸Contexts involving interrogative sentences appeared almost exclusively in ROOT-AFF-1 and ROOT-AFF-2, as epistemic and negated modals are rare in interrogatives. Out of 80 contexts for root-AFF, there were 21 interrogative sentences (19 involving possibility modals; 2 involving necessity modals).

Of course, some of the cues available to adults in this experiment might not be usable by children: for instance, children might lack some world knowledge. This limitation is intrinsic to any paradigm where adults are used to simulating word learning (the task asked of adults is to guess a word they already know, whereas children have to guess the meaning of a *new* word from the context in which it is used) (see Orita et al. 2013; White et al. 2016 for discussion). That said, children also have access to a substantially richer context than participants in our experiment, who had no visual or prosodic information and no common ground with the child and the mother.

We found a general effect of force, with participants being more accurate with possibility modals. This could be interpreted as possibility contexts being more informative than necessity contexts. However, the effect was carried by only two sub-conditions (ROOT-AFF-1 and ROOT-NEG; it was near-significant in EPI-AFF ($\chi^2(1) = 3.73$, $p = .053$), and not significant in ROOT-AFF-2). It was not significant once we took into account the effect of interrogative sentences, which led to a very high accuracy for root possibility modals (if we restrict to declarative contexts only, participants don't perform significantly better in possibility contexts).¹⁹

Lastly, turning to negation, we found that negated necessity modals are rare: our corpus results show that overall, modals scoping under negation were negated 7.4% of the time (vs. 22.6% for root possibility modals).²⁰ The results from Experiment 1 show that they are also less informative. We found opposite effects of negation on possibility and necessity modals: while negation led to a slightly higher accuracy for possibility modals (*can't*: 89.5% vs. *can*: 81.5% (NS)), it led to lower accuracy for necessity modals (*don't have to*: 61.3% vs. *have to*: 82.0%, $p = 0.011^*$) (significant interaction effect Force*Negation: $p = 0.0047^{**}$).

Why is that so? First, the low frequency of negated necessity modals may come from a competition with the use of a bare possibility modal, which can convey non-necessity via a scalar implicature (Horn 1972).²¹ We found a few cases that could be informative for children, like (11), where the context makes clear that the impossibility interpretation does not hold.

¹⁹This higher accuracy in possibility contexts might also reflect a general tendency to answer with possibility modals by default, maybe because of their relative frequency. To test for the effect of frequency, we compared accuracy for *can* and *able-to* (used in ROOT-AFF-2 and ROOT-NEG), which are both root possibility modals but strongly differ in frequency (3 *able* vs. 100 *can* in the Manchester corpus). The general accuracy on *able* was not significantly lower than on *can* (overall: *able*: 80.8% vs. *can*: 89.8% vs. *have to*: 71.7%).

²⁰Few other corpus studies address the distribution of possibility and necessity modals with respect to negation, and fewer look at child-directed speech, but they also suggest that negated necessity modals are not frequent. De Haan (2011) reports that negation is very rare with *must*: 2.5% in the Brown corpus (written English) and 1.4% in the Switchboard corpus (spoken English). Thornton and Tesan (2013) report the frequencies of some negative auxiliary verbs in the input to children in the Providence corpus, but don't specify their frequency relative to the positive forms. Jeretič (2018) reports that negation on necessity modals is infrequent in the input to French and Spanish children (French: 15.5%; Spanish: 6.2%).

²¹Horn focuses on a different but related problem, namely the fact that cross-linguistically, the 'O' corner of the Aristotelian square of opposition (corresponding to negated universals, e.g., non-necessity meanings) never seem to be lexicalized, whereas the other three corners (possibility, necessity, and impossibility) can be. Horn argues that this follows from the fact that there is no functional pressure to lexicalize non-necessity meanings: speakers already have a way to express non-necessity, using *scalar implicatures*.

- (11) MOTHER: now we don't throw things, do we?
 MOTHER: now.
 CHILD: I don't want that anymore.
 MOTHER: well then.
 MOTHER: we won't play with it anymore.
 MOTHER: you **don't have to** play with it.
 MOTHER: you can play with something else. (Aran, 2;6.10)

However our results suggest that most adult negated necessity modals are cases like (12), where the conveyed meaning is close to impossibility, which illustrate 'polite' uses of negated necessity modals. Here, *don't have to* is used to perform a prohibition.

- (12) CHILD: break.
 MOTHER: you want me to break it?
 CHILD: yeah.
 MOTHER: no.
 MOTHER: we **don't have to** break these things.
 MOTHER: oh.
 MOTHER: you've broken it.
 CHILD: yeah. (Aran, 2;0.28)

From this, we conclude that it is unlikely that children rely on negation to figure out the force of necessity modals. First, as discussed earlier, negation is potentially misleading for a number of necessity modals: *mustn't* is truth-conditionally equivalent to *can't*, which might drive children to infer that they express possibility, if children assume that negation scopes over root modals by default. Second, necessity modals that can scope under negation (e.g., *have to*, *got to*) are rare in the input, and their use is particularly misleading about their force because they often can be used to convey impossibility. Children will therefore need other strategies to solve the subset problem. However, our findings suggest that negation could be more helpful to figure out the force of possibility modals: they cooccur frequently in the input (22.6% of root possibility modals are negated), and Experiment 1 shows that impossibility contexts are highly informative (mean accuracy for *can't*: 89.5%). Children may be able to infer from these occurrences the force of possibility modals if they expect negation to scope over modals.

3.3 Experiments 2 and 3: Isolating the role of context

Experiment 1 shows high accuracy for both possibility and necessity. We take these results to mean that the context is informative as to force. But could it be that participants succeed at the task not by relying on the context, but through biases, which could also be at play in children's modal learning? In particular, could their high accuracy be due to a necessity bias that allows them to correctly guess necessity meanings?²² To isolate the contribution of the dialogue context, we ran two related follow-up experiments, presenting either only the target sentence without its discourse context (Experiment 2, 'sentence-only') or only the target sentence, but replacing all

²²We thank an anonymous reviewer for pointing this issue out.

she _____ have had a good night last night wherever she went.



Fig. 3a Experiment 2 stimuli (EPI-AFF, *must*)

content words with nonce words (Experiment 3, ‘Jabberwocky’). We expected that participants’ performance should decrease in Experiments 2 and 3, if their successes in Experiment 1 were due to a reliance on context.

3.3.1 Methods

Procedure. Experiments 2 and 3 were identical to Experiment 1, except that in Experiment 2, participants only saw the target sentence (without the preceding dialogue; see Fig. 3a), and in Experiment 3, they saw only the target sentence with all content words replaced by Jabberwocky (see Fig. 3b).²³ Experiment 2 lets us isolate the specific contribution of the dialogue. However, it does not remove all contextual information, since the content of the prejacet contributes to context (e.g., desirability of the event it describes). This motivates Experiment 3, which further removes any semantic information in the sentence (see Gillette et al. 1999; White et al. 2016, a.o.). As the task was shorter, participants judged all 40 contexts (Experiment 2: 60 trials: 20 possibility; 20 necessity; 20 controls using tense; Experiment 3: 40 trials: 20 possibility; 20 necessity). To make sure that participants kept paying attention, we also had eight attention checks (simple additions and subtractions, e.g., $1 + 3 = \underline{\quad}$). We removed from target sentences any repetitions (e.g., ‘dolly... dolly _____ use her potty’ was corrected to ‘dolly _____ use her potty’), as well as phatic words (e.g., *oh, yeah*). For Experiment 3, we replaced all content words (nouns, verbs, adjectives, adverbs) with nonce Jabberwocky words (e.g., *shink, gumbly*). We kept all function words (determiners, prepositions, complementizers, connectives, personal pronouns, temporal adverbs, and locatives), auxiliaries (*be, have*), modals, semi-modals, other than the target modal, which was replaced by a _____), plural morphology, and tense and aspect marking. Conditions were the same as in Experiment 1. Instructions are provided in online Appendix B.

3.3.2 Results

Participants. We recruited 252 participants on Amazon Mechanical Turk (Experiment 2: ROOT-AFF-1: 31, ROOT-AFF-2: 33; ROOT-NEG: 30; EPI-AFF: 29; 66 females, mean age: 44.0 years; Experiment 3: ROOT-AFF-1: 31, ROOT-AFF-2: 29; ROOT-NEG: 30; EPI-AFF: 38; 67 females, mean age: 38.8 years). In Experiment 2, we removed

²³Examples of the experiments can be accessed at <https://farm.pcibex.net/r/sohaoF/> (Exp. 2) and <https://farm.pcibex.net/r/ZbVcQT/> (Exp. 3).

she _____ have had a gumbly shink sest shink wherever she greft.

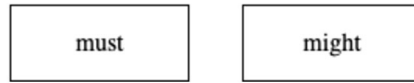


Fig. 3b Experiment 3 stimuli (EPI-AFF, *must*)

from the analysis one participant who was less than 75% accurate on attention checks and six participants who were less than 75% accurate on tense controls (5.7%).²⁴ In Experiment 3, accuracy on attention checks was very high (99.1%). No participant was excluded.

Analysis. Tables 5a and 5b report mean accuracy in each condition for Experiments 2 and 3. Like Experiment 1, we first ran binomial tests to see whether they differ from chance for each condition. In Experiment 2, participants were overall still good at guessing force (Table 5a). In Experiment 3, participants were above chance for possibility modals, but didn't differ from chance for necessity modals, except in EPI-AFF (Table 5b). We used generalized linear mixed effect models to test the effect of Experiment on Accuracy, first overall and then subsetting the data to test by each subcomparison group, comparing Experiments 2 and 1, Experiments 3 and 2, and Experiments 1 and 3. We tested Accuracy (dependent variable), with Experiment as a fixed effect and Subject and Item as random factors (Maximal model: $\text{Accuracy} \sim \text{Experiment} + (1|\text{Subject}) + (1|\text{Item})$). We compared these models with a reduced model without Experiment as a fixed effect ($\text{Accuracy} \sim 1 + (\text{Force}|\text{Subject}) + (1|\text{Item})$). First comparing Experiments 1 and 2, we found that participant performance is overall lower without the dialogue ($\chi^2(1) = 48.2, p = 3.9e-12^{***}$). Looking at the eight subcomparison groups, we found decreased performance for necessity contexts in ROOT-AFF-1, ROOT-AFF-2, and EPI-AFF and for possibility contexts in ROOT-AFF-2 and ROOT-NEG. We found no difference for possibility ROOT-AFF-1 and EPI-AFF and necessity ROOT-NEG. Detailed results of the comparisons are given in online Appendix D. Comparing Experiments 2 and 3, we found a significantly lower accuracy for Experiment 3 overall ($\chi^2(1) = 188, p < 2.2e-16^{***}$) and looking at each subcomparison group. Last, comparing Experiments 1 and 3, we also found decreased performance for Experiment 3 overall ($\chi^2(1) = 650, p < 2.2e-16^{***}$) and in each subcomparison group (see online Appendix D). Figure 4 summarizes the comparison between the three experiments for all conditions.

3.3.3 Discussion

These two control experiments let us isolate the specific contribution of context and show that it is informative beyond potential biases. In the Jabberwocky version (Ex-

²⁴Accuracy on attention checks and tense controls was very high (attention checks: 99.4%; tense controls: 95.8%), with no difference between groups. To compute accuracy on tense controls, we only included sentences that could not lead to an ambiguity (e.g., because of containing a temporal adverb) (10 out of 20 cases).

Table 5a Accuracy rates and binomial tests by condition, testing for significance vs. chance (Experiment 2) ($n = 116$, 20 observations per cell)

	Mean accuracy (se)		Exact binomial tests (two-sided)	
	possibility	necessity	possibility	necessity
ROOT-AFF-1	90.2% (0.030)	62.0% (0.062)	$p < .001^{***}$ 95% CI [0.88, 0.92]	$p < .001^{***}$ 95% CI [0.59, 0.65]
ROOT-AFF-2	71.8% (0.052)	73.0% (0.054)	$p < .001^{***}$ 95% CI [0.68, 0.74]	$p < .001^{***}$ 95% CI [0.70, 0.76]
ROOT-NEG	84.8% (0.036)	57.3% (0.061)	$p < .001^{***}$ 95% CI [0.82, 0.87]	$p = .00019^{***}$ 95% CI [0.54, 0.61]
EPI-AFF	88.6% (0.021)	64.6% (0.054)	$p < .001^{***}$ 95% CI [0.86, 0.90]	$p < .001^{***}$ 95% CI [0.61, 0.68]

Table 5b Accuracy rates and binomial tests by condition, testing significance vs. chance (Experiment 3) ($n = 129$, 20 observations per cell)

	Mean accuracy (se)		Exact binomial tests (two-sided)	
	possibility	necessity	possibility	necessity
ROOT-AFF-1	72.2% (0.031)	42.0% (0.029)	$p < .001^{***}$ 95% CI [0.69; 0.75]	$p = 1$ (NS) 95% CI [0.39; 0.45]
ROOT-AFF-2	63.4% (0.029)	45.3% (0.043)	$p < .001^{***}$ 95% CI [0.60; 0.67]	$p = 0.99$ (NS) 95% CI [0.42; 0.48]
ROOT-NEG	67.7% (0.027)	47.7% (0.041)	$p < .001^{***}$ 95% CI [0.64; 0.71]	$p = 0.88$ (NS) 95% CI [0.44; 0.52]
EPI-AFF	65.8% (0.019)	53.3% (0.032)	$p < .001^{***}$ 95% CI [0.63; 0.69]	$p = 0.038^*$ 95% CI [0.50; 0.56]

periment 3), participants' mean accuracy was 57.2% vs. 74.0% when they saw the sentence with content words (Experiment 2), and 79.9% when they saw the preceding dialogue (Experiment 1). This first confirms that syntax alone doesn't help: when we removed all semantic information, subjects were at chance.²⁵ Moreover, further analysis looking at the interaction between Force and Experiment (reported in online Appendix D, Table 16) suggests that the effect of having more contextual information is stronger for necessity modals than for possibility modals. If participants' high accuracy in Experiment 1 was due to a necessity bias, we would expect their performance to remain the same in Experiment 2: participants should guess necessity meanings, unless presented with direct evidence against it. Altogether, participants' high accuracy on possibility modals, even in the absence of context, suggests that if

²⁵The fact that participants perform better than chance on possibility contexts in the Jabberwocky experiment might reflect a general tendency to answer with possibility modals when given a forced choice task—perhaps because of their higher frequency, see fn. 24. It may also be driven by their higher accuracy in interrogative sentences, as discussed in Sect. 3.2.3 (mean accuracy for possibility in interrogatives: Exp1: 98.1%; Exp2: 96.2%; Exp3: 76.1%; in declaratives: Exp1: 75.1%; Exp2: 65.8%; Exp3: 59.5%).

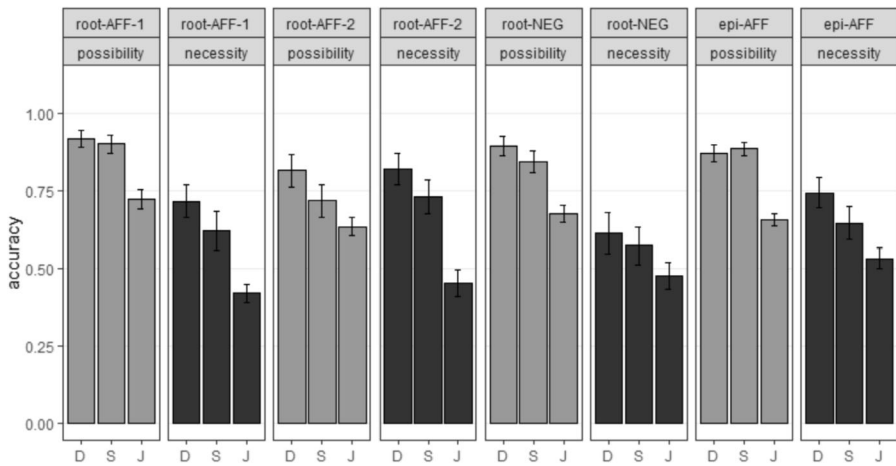


Fig. 4 Accuracy by condition comparing Experiment 1 (D: ‘dialogue’), Experiment 2 (S: ‘sentence-only’), and Experiment 3 (J: ‘Jabberwocky’)

they bring a force bias to the task, it is more likely to be a possibility, rather than a necessity, bias.

3.4 Experiment 4: Desirability

The results from Experiments 1-3 argue that the conversational context in which modals are used is informative about their force. But what is it about the context that is particularly informative? As discussed in Sect. 3.2.3, several factors could be at play. Our post-hoc analysis suggested that the cues may vary with flavor: for root modals, necessity modals seem associated with undesirable and effortful events; for epistemics, necessity modals seem to occur in contexts that highlight strong evidence that support the proposition expressed by the prejacent. We now turn to an experiment that tests the hypothesis that (un)desirability matters for root modals, as an initial proof of concept, and leave a more systematic probing of additional features of the context for future research.

We hypothesize that the desirability of the prejacent could be playing a crucial role in the acquisition of force for root modals. Desirability is a feature likely to be conceptually accessible to young children: the cognitive developmental literature suggests that children can reason about desires quite early on and understand that people can have incompatible desires (Wellman and Woolley 1990; Repacholi and Gopnik 1997; Rakoczy et al. 2007; Ruffman et al. 2018, a.o.). Moreover, preschool children have been shown to be sensitive to desirability for modal usage pragmatics, in particular compared to unmodalized expressions (Ozturk and Papafragou 2015). The first goal of Experiment 3 is to assess the availability of this cue in the input: do adults actually use necessity modals more frequently with undesirable events (e.g., ‘You *must/can* eat your brussels sprouts’) and possibility modals with desirable events (e.g., ‘You *can’t/must* have a cookie’)? Second, does this contribute to participants’ performance in Experiment 1, i.e., do adults actually rely on this cue to guess force?

Fig. 5 Experiment 4 stimuli: example trial (ROOT-AFF-1, *can*)

riding on Aran the horse

Does this sound fun?

no

yes

3.4.1 Methods

Procedure. Participants were asked to indicate whether various activities (e.g., ‘doing a puzzle’) sounded fun or not (see Fig. 5). They were told that the activities involved 2-year-old children and their mothers. The different activities corresponded to the prejacent²⁶ of the modals tested in Experiment 1:²⁷ for example, for ‘Can the dolly ride on Aran the horse?’, participants were asked whether ‘riding on Aran the horse’ sounded fun (‘yes’) or not (‘no’). We used the prejacent, rather than full modal sentences to avoid biasing toward positive responses for possibility modals and toward negative responses for necessity modals. Referential pronouns (e.g., *it*) were replaced whenever they could be recovered from the context (e.g., ‘Finding the green marker’ for ‘Can you find it?’). In each group, participants judged all 40 prejacent (42 trials: 20 possibility, 20 necessity; 2 initial practice items, which were removed from the analysis). To make sure participants kept paying attention, we had 10 attention checks (e.g., $1 + 3 = \underline{\quad}$). Instructions are given in online Appendix B. As our hypothesis concerns root modals, we ran the experiment only on ROOT-AFF-1 (*can* vs. *must*) and ROOT-AFF-2 (*can/able* vs. *have to*). **Rationale**—This experiment allows us to first assess the desirability of the different events in as objective a way as possible, to see if there is a relation between desirability (measured by the proportion of yes answers to ‘being fun’, a child-friendly way of assessing what is desirable) and force usage in the corpus. We can then probe whether adults used this cue to infer force in Experiment 1 by looking at the correlation between the desirability score in Experiment 4 and accuracy in Experiment 1. We expect a negative correlation for necessity modals (fewer ‘yes’ responses for accurate guesses of necessity uses) and a positive one for possibility modals (more ‘yes’ responses for accurate guesses of possibility uses).

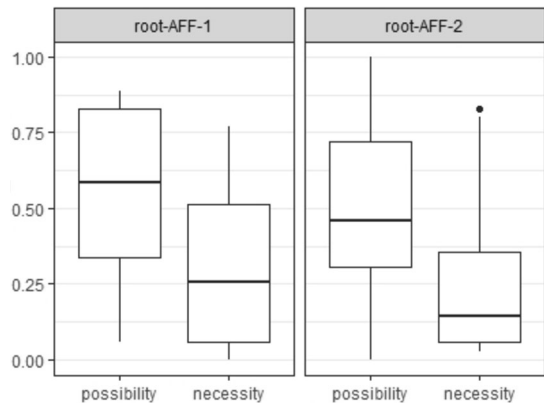
3.4.2 Results

Participants. We recruited 70 participants on Amazon Mechanical Turk (ROOT-AFF-1: 35, ROOT-AFF-2: 35; language: US English; 35 females, mean age: 40.4 years). Accuracy on attention checks was very high (99.6%), and we did not have to remove any participant from the analysis based on attention checks.

²⁶This is not true *stricto sensu*, as participants also lose the information about the subject (e.g., *I/you/Caroline...*).

²⁷An example of the experiment can be accessed at <https://farm.pcbex.net/r/lrxZaB/>.

Fig. 6 Distribution of ‘desirable’ answers for possibility and necessity contexts for each group



Analysis. Table 6 reports means of ‘yes’ answers (‘fun’) in all conditions. First, to test for the effect of Force (possibility vs. necessity), we fitted a generalized linear mixed effects model, testing Answer (dependent variable), with Force as a fixed effect and Subject and Item as random factors. We compared this model with a model without Force as fixed effect (glmer syntax for the maximal model: (Answer)~Force+(1|Subject)+(1|Item)); for the reduced model: ((Answer)~1+(1|Subject)+(1|Item)). ‘Fun’ answers were coded as ‘1’. We found a general effect of Force: participants judged prejacent extracted from possibility statements overall more ‘desirable’ than those extracted from necessity statements (Model comparison: $\chi^2(1) = 15.5$, $p = 8.2e-05^{***}$). We subsetted the data for each group, using the same formulae and comparison methods. The effect is significant in both groups (ROOT-AFF-1: $\chi^2(1) = 8.2$, $p = 0.0041^{**}$; ROOT-AFF-2: $\chi^2(1) = 6.2$, $p = 0.012^*$).²⁸ Figure 6 shows the distribution of ratings for possibility and necessity for the two groups. Then, we computed correlations between the desirability score (Experiment 4) and accuracy in Experiment 1 (see Fig. 7). For possibility, we found a weak positive correlation (Pearson’s $r = 0.12$) ($t(1398) = 4.42$, $p < 0.001^{**}$; 95%-CI: [0.065; 0.168]); for necessity, a weak negative correlation (Pearson’s $r = -0.073$) ($t(1398) = -2.74$, $p = 0.0063^{***}$; 95%-CI: [-0.125; -0.021]).

3.4.3 Discussion

Our results confirm our initial observations for Experiment 1 and show that there is a relation in children’s input between the desirability of the prejacent (evaluated by participants who were blind to the force of the modal originally used) and force. Adults use possibility modals more frequently with desirable events and necessity modals more frequently with undesirable events (mean desirability score for possibility modals [*can/able*]: 52.9%; for necessity modals [*must/have to*]: 28.6%). Furthermore, the lower accuracy for possibility modals with undesirable prejacent in

²⁸We sometimes had to step back to random-intercepts-only models when the model failed to converge with the full random-effects specification. We also checked that there was no significant difference between groups (ROOT-AFF-1 vs. ROOT-AFF-2) (Answer=1)~Group+(1|Subject)+(1|Item). Results of the model comparison were not significant ($\chi^2(1) = 0.22$, $p = 0.64$).

Table 6 Desirability scores and significance tests (binomial linear mixed effects models comparing possibility/necessity) for possibility and necessity modals

	Mean of desirable ('yes') answers (se)		Effect of Force
	possibility	necessity	
ROOT-AFF-1	56.0% (0.063)	31.4% (0.060)	$\chi^2(1) = 8.22, p = 0.0041^{**}$
ROOT-AFF-2	49.7% (0.067)	25.7% (0.057)	$\chi^2(1) = 6.2562, p = 0.012^*$
ALL	52.9% (0.045)	28.6% (0.041)	$\chi^2(1) = 15.5, p = 8.2e-05^{***}$
ALL	40.7%		

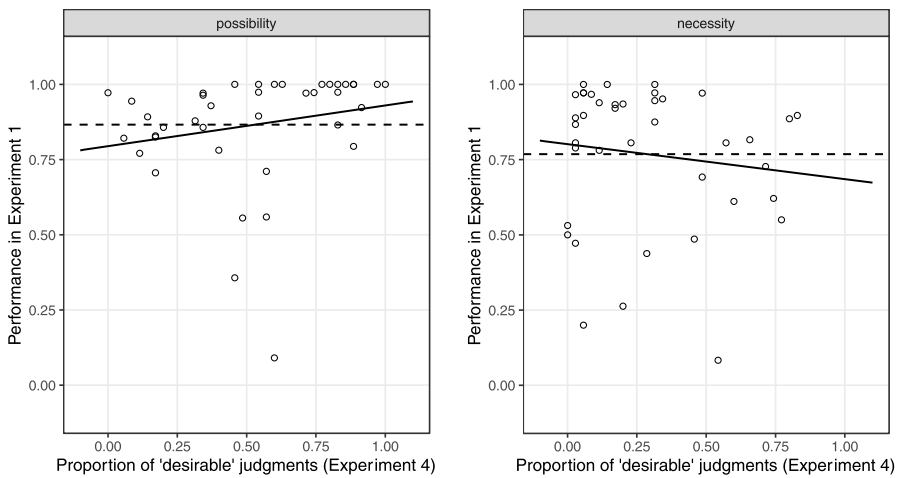


Fig. 7 Relation between accuracy in Experiment 1 (y-axis) and desirability score in Experiment 4 (x-axis) by force. Black lines correspond to Pearson's r. Dashed lines correspond to the mean accuracy in Experiment 1, for possibility and necessity contexts

Experiment 1 and for necessity modals with desirable prejacent suggests that adult participants in Experiment 1 likely made use of desirability in their force judgments. Together, this suggests that children could conceivably use this cue: it is available in the input, and the cognitive developmental literature suggests they are sensitive to desirability, though its association with obligations may require some life experience.

3.5 Summary: Children's modal input

Our corpus results show that children are exposed to many more possibility modals than necessity modals and that they hear the former relatively more often with negation than the latter. We have also seen that negation is unlikely to help learners figure out necessity modals and might in fact hinder their acquisition. It appears, however, to be potentially much more helpful for possibility modals. If learners can't rely on negation or other DE environments to solve the subset problem, do they need to rely on a necessity bias? Our experiments suggest that they may not need to, as the conver-

sational context in which modals are used is informative about both forces. If children are able to make use of these conversational cues, they need to rely on neither negation nor a necessity bias. One aspect of the context that could be particularly helpful for root modals is the desirability of the prejacent. Now that we have a clearer picture of children's modal input, and what, in principle, learners may be able to rely on or not, we turn to our study of children's productions.

4 Children's modal productions

To study children's modal production, we used the same methods as for the input study. We first present the results from our corpus analysis, comparing children's early productions to those of adults', and then present results from Experiment 5, which is based on the same paradigm as Experiment 1, but tests children's utterances.

4.1 Corpus study

4.1.1 Results

Like the adults, children produce more possibility modals than necessity modals, and the difference is even stronger (79.3% of children's modal productions vs. 72.5% of adults') (Table 7). *Can* is by far their most common modal (75.6% vs. 57.3% of adults'), and *have to* is their most frequent necessity modal (7.3% vs. 12.0% of adults'). Necessity modals are particularly rare with negation in their productions (only 5.1%), whereas negated possibility modals are very frequent: 51.0% (adults: necessity modals with negation: 10.1%; possibility modals with negation: 20.9%). Epistemic modals are overall very rare in child productions: they represent only 2.4% of children's modal utterances (114 cases, possibility: 93, necessity: 21) vs. 8.8% of all adults' modal utterances. Looking at the evolution of children's productions during the time period, summarized in Fig. 8a, we find that the relative proportion of necessity modals tends to increase with age: while they represent 12% of children's modal productions between 2 and 2;03, they represent 24.5% between 2.9 and 3 (Fig. 8b confirms that for adults, the relative proportion of possibility and necessity modals does not significantly change over time: we only find a slight increase of necessity modals).

4.1.2 Discussion

We found that children use (root) possibility modals frequently, both with and without negation, which we can take as initial evidence of productivity (Stromswold 1990). Children produced few necessity modals, and rarely with negation.²⁹ This difference might be explained by several factors: the difference in frequency in their input (if

²⁹ Similar distributional patterns (possibility modals being used more frequently than necessity modals and occurring frequently with negation) are found in Spanish and French (Jeretić 2018) and Dutch (van Dooren et al. 2022).

Table 7 Counts and percentages of modal uses by force, ordered by lemma frequency, with and without negation, for children (repetitions excluded: 17.0% of the data) ($X^2(1, N = 24830) = 92.6, p < 2.2e-16$)^a

	CHILD (<i>n</i> = 4844)		CHILD (<i>n</i> = 4800) ^b			
	all		no negation		negation	
POSSIBILITY	3841	79.3%	1861	49.0%	1937	51.0%
<i>can</i>	3663	75.6%	1739	48.0%	1881	51.9%
<i>might</i>	86	1.8%	78	97.5%	2	2.5%
<i>could</i>	80	1.6%	34	39.5%	52	60.4%
<i>able</i>	3	0.1%	1	33.3%	2	66.6%
<i>may</i>	9	0.2%	9	100%	0	0%
NECESSITY	1003	20.7%	950	94.8%	52	5.2%
<i>have to</i>	352	7.3%	345	98.0%	7	1.9%
<i>got to</i>	288	5.9%	283	98.3%	5	1.7%
<i>should</i>	22	0.5%	17	80.9%	4	19.0%
<i>need to</i>	217	4.5%	204	94.0%	13	5.9%
<i>must</i>	114	2.4%	94	82.5%	20	17.5%
<i>supposed</i>	9	0.2%	6	66.7%	3	33.3%
<i>ought to</i>	1	0.0%	1	100%	0	0%

^aNote that the chi-square assumption of independence of observations is violated by corpus samples, as the same speaker supplies multiple uses per cell. However, this test metric is commonly used in corpus linguistics for simple distributional comparisons and is not straightforwardly a violation, as we are comparing spontaneous utterances, not individuals (each spontaneous production is taken as a proxy for independence)

^bExcluding tag questions (0.9% of children's modal productions vs. 5.7% of adults' modal productions)

children grasp more frequent words first) and the differences in social status and topics of conversations between children and adults (for instance, children may be less prone to giving orders and thus less prone to using root necessity modals). Necessity modals tended to become more frequent over time. However, quantitative production data can only provide a partial picture of whether children use and understand modals correctly. To assess these productions in a more qualitative way, we ran Experiment 5 (identical in method to Experiment 1) on children's modals.

4.2 Experiment 5: Children's modal productions

The goal of this experiment was to investigate children's early modal productions to see whether they use modals in an adult-like way, by comparing their usage to adult usage (Experiment 1). Can adults guess the force of a modal used by a child, given the conversational context in which they use it?

4.2.1 Methods

Experiment 5 was identical to Experiment 1, except that we used children's utterances instead of adults' and made small changes in the instructions (see online Ap-

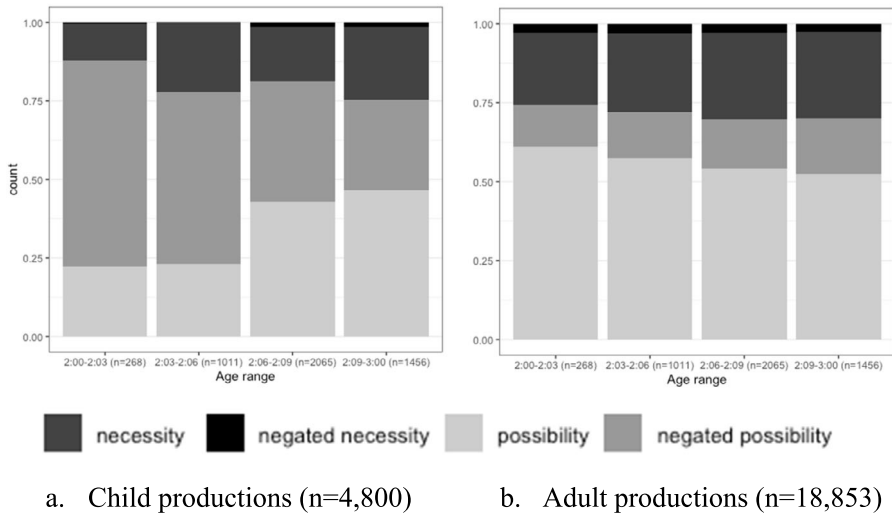


Fig. 8 Evolution of children's modal productions from 2 to 3 years of age by force and negation, binned in three-month periods

pendix B).³⁰ An example of the display is given in Fig. 9. We implemented the same conditions: ROOT-AFF-1, ROOT-AFF-2, ROOT-NEG, and EPI-AFF. Controls were also based on tense (past vs. future). **Extraction procedure**—Given the low frequency of negated necessity modals and epistemic necessity modals in child productions, we could test only 10 different contexts for ROOT-NEG necessity and 12 contexts for EPI-AFF necessity conditions.³¹ This did not make a difference for the participants, who always had 10 contexts to judge per condition (40 dialogues in the whole experiment: 20 trials: 10 possibility, 10 necessity; 20 controls: 10 past, 10 future). In all other conditions, 10 contexts were selected randomly out of a list of 20 contexts initially extracted from the corpus, in the same way as for the adult experiment. **Exclusion criteria**—Given the low frequency of negated necessity and epistemic modals, we did not remove cases where the modal already appeared in the preceding dialogue.³² We made sure to include examples in the training (*the/a*) and control items (past/future) where the right (or wrong) answer appeared in the preceding dialogue. Again, we removed Briticisms, but we did not correct children's ungrammatical utterances (e.g., *comed* for *came*), except in the case of *have to* when children omitted *to* (so that participants would not reject the answer because of its ungrammaticality). **Rationale**—We made the assumption that adults rely on their own competence to judge usage and that the dialogues preceding the modal sentence are equally in-

³⁰An example of the experiment can be accessed at <https://farm.pcbibex.net/r/mYOcOz/>.

³¹Because some of the negated *have to* in child productions were particularly unclear (e.g., 'I can't have to read it.'), we also used *not gotta* and *not need*. Details are provided in online Appendix B.

³²Out of 148 contexts, 36 of them had the modal appear in the preceding dialogue (24.3%) (uttered by the child: 11; by the mother or another adult: 20; by both: 5).

Fig. 9 Experiment 5 stimuli (child productions): example trial (*must*)

CHILD: train going round minute.
 MOTHER: in a minute?
 MOTHER: what... where's the train?
 CHILD: train... train must be there.
 MOTHER: it must be there?
 MOTHER: well.
 MOTHER: I can't see it.
 CHILD: _____ be at home.



formative for adults' and children's utterances.³³ If children use their modals in an adult-like way, we expect no difference in accuracy between Experiment 5 and Experiment 1. If they do not (e.g., they use *can* in a necessity situation when adults would use *must*, or they use *must* in a possibility situation when adults would use *can*), we expect a lower accuracy for children's utterances.

4.2.2 Results

Participants. We recruited 289 participants on Amazon Mechanical Turk (EPI-AFF: 74, ROOT-AFF-1: 72, ROOT-AFF-2: 73; ROOT-NEG: 72; language: US English; 173 females, mean age=40.2-years-old). We removed 18 participants (6.2%) who were less than 75% accurate on controls.³⁴ We thus present results for 273 participants (EPI-AFF: 68; ROOT-AFF-1: 70; ROOT-AFF-2: 70; ROOT-NEG: 65).

Analysis. Table 8 reports mean accuracy in each condition (summarized in Fig. 10). We first ran the same binomial tests as for Experiment 1. Participants performed better than chance in all conditions involving possibility, but not necessity: for ROOT-AFF-2 (*have to*) (mean accuracy: 42.6%) and ROOT-NEG necessity (*not have to*) (mean accuracy: 32.3%), they performed lower than chance (Table 8). **Force**—We used generalized linear mixed effects models testing Accuracy with Subject and Item as random factors, comparing a model with Force as fixed effect to a model without it (Maximal model: Accuracy~Force+(Force|Subject)+(1|Item)). The difference is significant in all conditions, with higher accuracy for possibility modals

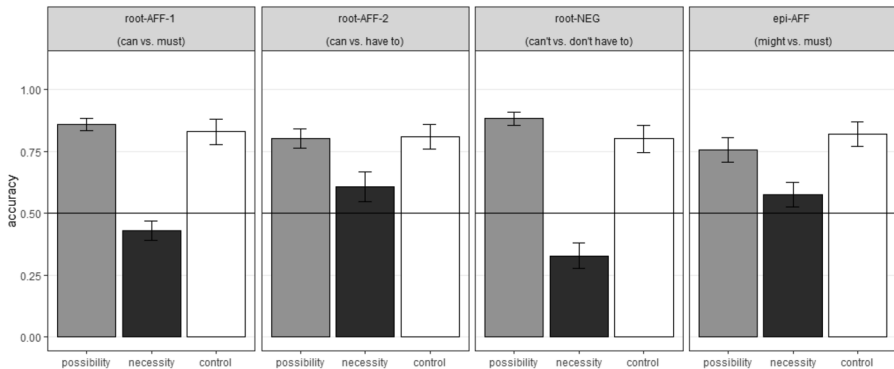
³³As a proxy, we checked that the mean length of dialogues was the same for adults and children (mean number of words for children's contexts: 39.6 words/dialogue; for adults' contexts: 38.9).

³⁴For the adult version, the proportion of errors on controls was very low (5.4%), with no difference between groups. For the child version, however, the initial proportion of errors on controls was quite high (21.6%): post-hoc analysis revealed that this came from five control contexts for which the accuracy was particularly low, and thus were not reliable controls. We decided to remove these five controls from our exclusion criteria, as they were particularly difficult and probably do not indicate that subjects were not doing the task correctly. After restricting to the 15 remaining controls, mean accuracy on controls was 90.0%, showing that participants were not answering randomly.

Table 8 Accuracy rates and binomial tests by condition, testing for significance vs. chance (Experiment 5: children's productions) ($n = 273$, 10 observations per cell)

	Mean accuracy ^a (se)		Exact binomial tests (two-sided)	
	possibility	necessity	Possibility	necessity
ROOT-AFF-1	85.1% (0.026)	42.6% (0.039)	$p < .001^{***}$ 95% CI [0.82, 0.88]	$p < .001^{***}$ 95% CI [0.39, 0.46]
ROOT-AFF-2	79.6% (0.041)	60.2% (0.060)	$p < .001^{***}$ 95% CI [0.77, 0.83]	$p < .001^{***}$ 95% CI [0.56, 0.63]
ROOT-NEG	88.2% (0.027)	32.3% (0.050)	$p < .001^{***}$ 95% CI [0.86, 0.91]	$p < .001^{***}$ 95% CI [0.29, 0.36]
EPI-AFF	75.6% (0.050)	56.8% (0.047)	$p < .001^{***}$ 95% CI [0.73, 0.80]	$p < .001^{***}$ 95% CI [0.53, 0.61]
Total	82.1% (0.019)	50.1% (0.028)		
ALL	67.4% (0.021)			

^aAccuracy corresponds to the mean accuracy across the 20 contexts initially extracted for each condition. On average, each context was seen by 34.7 participants (ranging between 24 and 47)

**Fig. 10** Accuracy by condition, Experiment 5, children's productions ($n = 273$)

(Model comparison: all: $\chi^2(1) = 20.49$, $p = 5.9e-6^{***}$; ROOT-AFF-1: $\chi^2(1) = 60.4$, $p = 7.7e-15^{***}$; ROOT-AFF-2: $\chi^2(1) = 7.37$, $p = 0.0066^{**}$; ROOT-NEG: $\chi^2(1) = 38.1$, $p = 6.6e-10^{***}$; EPI-AFF: $\chi^2(1) = 7.93$, $p = 0.0048^{**}$). **Negation**—We compared ROOT-AFF-2 and ROOT-NEG, as they included the same lemmas. We used the same method, comparing a model with Negation as fixed effect to a model without it (Maximal model: Accuracy~Negation+(1|Subject)+(1|Item)); for the interaction: Accuracy~Force*Negation+(1|Subject)+(1|Item)). We found a significant difference (*can* vs. *can't*: $\chi^2(1) = 3.65$, $p = 0.056^*$; *have to* vs. *not-have-to*: $\chi^2(1) = 6.74$, $p = 0.0093^{**}$; Interaction Force*Neg: Accuracy~Force*Group+(1|Subject)+(1|Item): $\chi^2(1) = 9.24$, $p = 0.0024^{**}$). **Flavor**—There was no effect of flavor ($\chi^2(1) = 0.14$, $p = 0.71$). **Age** (adult vs. child usage)—We then tested for the effect of Age (adult vs. child usage), first overall and then for each subcondition

Table 9 Results of the model testing effect of Age (adult usage vs. child usage)

	possibility	necessity
ROOT-AFF-1	$\chi^2(1) = 3.12, p = 0.078$ (NS)	$\chi^2(1) = 35.8, p = 2.1e-09^{***}$
ROOT-AFF-2	$\chi^2(1) = 5.80, p = 0.016^*$	$\chi^2(1) = 51.8, p = 6.3e-13^{***}$
ROOT-NEG	$\chi^2(1) = 2.78, p = 0.096$ (NS)	$\chi^2(1) = 21.1, p = 4.4e-06^{***}$
EPI-AFF	$\chi^2(1) = 3.76, p = 0.053$ (NS)	$\chi^2(1) = 0.22, p = 0.64$ (NS)
all	$\chi^2(1) = 15.9, p = 6.7e-05^{***}$	$\chi^2(1) = 175.7, p < .001^{***}$
ALL	$\chi^2(1) = 231.4, p < 2.2e-16^{***}$	

(Accuracy~Experiment+(1|Subject)+(1|Item)). We found a significant difference, with an overall lower accuracy for child usage ($\chi^2(1) = 260.5, p < .001^{***}$). Among possibility conditions, only ROOT-AFF-1 was significant; among necessity conditions, all comparisons were significant except EPI-AFF (Table 9). We found a strong interaction Force*Age: the difference in accuracy between possibility and necessity modals for child productions was larger than for adult productions ($\chi^2(1) = 46.2, p = 1.06e-11^{***}$).

4.3 Discussion

Even if participants were less accurate than when judging adults' modals, they were good at identifying possibility modals used by children, for both flavors (mean accuracy on all possibility modals: 82.1% vs. 87.5% when judging adult modals). Participant performance for necessity modals was much lower (only 50.1%, vs. 72.3% for adult modals), especially for negated uses (32.3%). Examples like (13) and (14), which led to particularly low accuracy, illustrate children's non-adult-like uses of necessity modals with and without negation (and confirm that the effect is likely not just due to participants expecting children to use possibility modals more often than necessity modals).

- (13) [. . .] CHILD: what shall I put first?
 CHILD: that.
 CHILD: what's that?
 MOTHER: pardon?
 CHILD: I **have to** see a cat. (Becky, 2;08.16) (HSP accuracy: 2.9%)
- (14) CHILD: . . . no eggs.
 MOTHER: I thought we had all of these eggs.
 CHILD: they not.
 CHILD: they go in the bag.
 CHILD: they going in there.
 CHILD: they go in there.
 MOTHER: oh you're putting them back in there now, are you?
 CHILD: you **don't have to** eat them. (Carl, 2;8.07) (HSP accuracy: 20.0%)

Together, these results suggest that children master possibility modals early, as they use them in an adult-like way. Their necessity modals, however, seem delayed: children do not use them in an adult-like way, suggesting that they either haven't mastered their underlying force yet or have difficulty deploying them in the right situations. But if the difficulty with necessity modals we observe here for 2-to-3-year-olds persists into the preschool years, it could explain both types of over-acceptance found in comprehension studies: children would accept necessity modals in possibility contexts because they haven't mastered their underlying force and accept possibility modals in contexts where adults prefer necessity modals because they lack a stronger alternative (i.e., they have not yet worked out scale-mate relations for English modals).

Importantly, we found no evidence in favor of a necessity bias. Children's highly adult-like uses of possibility modals might even suggest a bias toward possibility. Note that this lack of evidence doesn't necessarily entail that children don't rely on a necessity bias when acquiring modals.³⁵ It's conceivable that children use the bias to acquire necessity modals, but fail to use them in an adult-like way for independent reasons, as alluded to above. However, the lack of straightforward evidence for a necessity bias in our child results, together with its superfluity given our input results, suggests that a bias toward strong meanings is dispensable, even for modals.

5 General discussion and future directions

How do children figure out the force of their modals? In particular, what prevents them from falling prey to the subset problem modals give rise to and hypothesizing possibility meanings for necessity modals? To address these questions, we examined the modals that English-learning young children get exposed to and produce themselves. We found that children seem to master possibility modals early: by age 2, they are already using them productively, with and without negation, and in apparently adult-like ways. Children, however, seem to struggle with necessity modals. The few necessity modals they produce do not seem adult-like and appear in situations where adults prefer possibility modals. If this struggle with necessity modals persists into the preschool years, it could explain why prior studies show that children tend to accept them in possibility contexts (they're uncertain about their force), and also why they accept possibility modals in necessity contexts (they lack a stronger alternative).

Yet children eventually figure out necessity modals, and the question is how. From our input study, we see that given the way modals are used in speech to children, children cannot reliably make use of DE environments like negation, as Gualmini and Schwarz (2009) proposed as a general solution for subset problems. Negation may even be partly responsible for children's difficulties with necessity modals. First, its scopal behavior with modals is not uniform: some, but not all, necessity modals outscope negation. If children were to rely on negation to figure out the force of necessity modals, they could be misled into thinking that a modal like *must* expresses

³⁵As pointed out by an anonymous reviewer, our results may also be consistent with theories of acquisition that use a necessity bias that is sensitive to input frequencies, along the lines of Piantadosi et al.'s (2012b) Bayesian learner (see also Piantadosi et al. 2012a). We intend to address this in future work.

possibility (*must not* ~ *cannot*), if they assume that negation scopes over the modal. Second, we found that negated necessity modals are rare in speech to children, perhaps for functional reasons, as speakers can express non-necessity via scalar implicatures triggered by the simple use of a possibility modal. Finally, Experiment 1 shows that the context is the least clear about force for negated necessity modals. This seems to be due to their use in impossibility situations, as ways to soften requests or orders (e.g., ‘you *don’t have to* break those things’, used as a prohibition). Negation, however, might be quite useful for children to home in early on possibility meanings for possibility modals: it occurs frequently with possibility modals, and our experimental results suggest that negated possibility contexts are particularly informative about the force of the modal.

If learners can’t rely on DE environments to solve the subset problem for modals, might they then need a necessity bias? Such a bias is in principle not necessary, as our experiments show that the conversational context in which modals are used is informative about force. What exactly about the context gives away modal force? Our initial foray into contextual features suggests one factor that could be particularly helpful for deontic modals is the perceived (un)desirability of the prejacent (e.g., ‘you *have to* eat your peas’ vs. ‘you *can* have a cookie’). Desirability is a notion that should uncontroversially be available to young children, though its association with obligations may take some time. Experiment 4 confirms the potential usefulness of desirability and shows that root *necessity* modals tend to occur with undesirable prejacentes, and *possibility* modals with desirable prejacentes. Participants are better at guessing necessity modals when they occur with undesirable prejacentes. For epistemic modality, our post-hoc analysis suggests that contexts that explicitly highlight salient evidence in favor of the prejacent may bias interpretations toward necessity. Other aspects of the context could also prove useful, including situational cues (e.g., who the interlocutors are), cues from world knowledge (e.g., what is allowed or prohibited), or pragmatic cues (what the speaker is trying to achieve, in particular performing orders, permissions, or prohibitions). We plan to explore this further in future work.

Taken together, however, the results from our input study and our child study seem to lead to a conundrum: if the conversational context is informative for both forces (Experiments 2 and 3 suggest that context is particularly helpful for necessity modals), why should children struggle more with necessity modals? Several factors could be at play. First, these difficulties might be a matter of quantity, rather than quality, as our corpus results show that necessity modals are less common in the input.³⁶ Second, children may need to accrue some world knowledge to be able to make use of contextual cues (e.g., the association between obligations and undesirability). Children might also face conflicting cues. In particular, negative contexts seem useful for possibility modals, but misleading for some necessity modals. As discussed, if children assume the same scope relations for *can* and *must*, uses of *mustn’t* could suggest

³⁶We do not have enough data from each child to see whether children’s adult-like uses of necessity modals correlate with frequency in their input (i.e., whether children who hear more necessity modals use them more appropriately), but we intend to pursue this question in future work. We thank an anonymous reviewer for this great suggestion. See Dieuleveut (2021) for preliminary results and discussion.

to them that it expresses possibility. And interestingly, when we compare our two affirmative root conditions, we found a lower performance for children's *must* (42.6%) than for *have to* (60.2%), though this could reflect differences in their input frequency (*have to* represents 12% of all parents' modal utterances, *must* only 2.3%). The interaction with negation could also explain why possibility modals are mastered so early: our study of the input shows that negated possibility modals are frequent and used in particularly informative contexts (e.g., to talk about prohibitions or physical impossibilities). Last, children's struggles with necessity modals might not reflect a lack of knowledge of their underlying force, but rather difficulty deploying this knowledge in the right situations. Children may have trouble tracking what information is shared among interlocutors and thus grasping the intended domain of quantification (as has been argued for definite descriptions, where children seem to be overly permissive in using and accepting definites in contexts in which their uniqueness presupposition is not satisfied; Karmiloff-Smith 1979; Schaeffer and Matthewson 2005; Brockmann et al. 2018).

What we hope to have shown here is that the conversational context is a rich source of information for modal force, so that, in principle, learners can solve the subset problem without having to rely on potentially misleading negative contexts or on a controversial strong (necessity) bias. We have only begun to scratch the surface for which properties of the context are useful and why, but our results show that even the short, written snippets of conversation we provided to adult participants are informative, and this informativity is above and beyond the more limited context in the preadjacent alone (Exp. 1 vs. Exps. 2/3). In fact, even just the information in the preadjacent leads to above chance success in Experiment 2. Note that the context in real life is far richer than in Experiment 1 contexts, which lack prosodic and emotive information carried by speech, shared knowledge between mother and children, and visual information (among other information). Our studies, in combination with the fact that real life contexts are much richer, suggest that usage contexts may richly support modal force learning. That said, we don't know yet how representative these contextually useful cases are in children's actual experiences or if the additional factors (e.g., prosody) will be enriching, as seems intuitive, or obfuscating. Now that we are confident that information of some kind is available in the context and that desirability and mention of evidence are two promising types of information supporting modal meanings, we can pursue research asking whether children actually make use of these (and other) specific aspects of the context to acquire modal force.

Our child study shows no evidence for a necessity bias in children's early modal productions. In fact, children's early successes with possibility modals and failures with necessity modals could even suggest a bias toward possibility. Still, our results cannot rule out children actually having a necessity bias and grasping the force of necessity modals, but failing to use them in an adult-like way for independent reasons. In future work, we plan to test for such biases more directly through a novel modal word task, adapted from an adult study in Dieuleveut et al. (2021), to see what meanings children attribute to novel modals.

To address what aspects of the context are useful and whether children actually make use of them in their modal force acquisition, we plan on testing whether various features of the context in the input are good predictors of children's mastery of

necessity modals, as indexed by accuracy on the child HSP task. For instance, to see whether desirability actually matters in children's (root) modal acquisition, we could test whether frequent uses of necessity modals with undesirable prejacent in the input predict earlier mastery of necessity modals: will a child whose parents primarily use necessity modals with undesirable prejacent use necessity modals in an adult-like way sooner than a child whose parents use necessity modals more often with desirable prejacent? (See Dieuleveut 2021 for preliminary results.³⁷)

Before we conclude, we would like to briefly discuss potential implications of our findings for how children acquire modal force in languages beyond English, particularly in languages with 'variable force' modals. As discussed in Sect. 2.1, in a language like English where modals come in both forces, we can expect speakers to use possibility and necessity modals in fairly distinct situations and, notably, to avoid using possibility modals in necessity situations (modulo politeness considerations). And indeed, our input results show that speakers use possibility and necessity modals in distinct situations that are highly reflective of force. But in a language that lacks modal duals, speakers are more likely to use particular modals in both possibility and necessity situations. For variable force modals that are underlying possibility modals, like Nez Perce *o'qa*, it seems like negation would thus be crucial for learners to home in on its underlying force—just like it was for Deal (2011) to argue for a possibility analysis. For variable force modals that are underlying necessity modals, as in St'át'imcets or Washo, the challenge may be much greater. Not only might speakers use the same modals in possibility and necessity situations, but learners may not be able to rely on negation, given that it can't scope over modals in these languages. And yet speakers seem to have converged on necessity meanings for these modals, as evidenced by their preferred translations using English necessity modals (Rullmann et al. 2008; Bochnak 2015). Here fieldworkers can and do rely on such translations as evidence for the modals' underlying force, but this strategy is obviously inaccessible to the child. How do learners figure out their underlying force? This situation might at first blush argue for a necessity bias. However, it could also be that while these modals can in principle be used in possibility situations, in practice, variable force modals are mostly used in contexts where English speakers use necessity modals, in which case, their acquisition could involve the same reliance on contextual cues that we've proposed for the acquisition of English modals.

6 Conclusion

This study has laid the groundwork for understanding when and how children figure out the force of the modals in their language, and in particular, how they may solve the subset problem of modals, by looking at young children's natural productions and their input. Our child results suggest that by age 2, children have mastered possibility modals but struggle with necessity modals. Our input results suggest that, in principle, learners could learn force simply based on how modals are used, as the

³⁷For a similar investigation of what aspects of the input predict children's understanding of *think* and *know*, see Dudley (2017).

conversational context is informative about force. If children are able to pick up on these conversational cues, they could figure out modal force without having to rely on negation—which we’ve shown is particularly unreliable for necessity modals—or on a bias toward necessity—a bias for which we found no evidence in child productions. Our results are thus in line with recent discussions of other subset problems arguing that strong meaning biases may have no place in acquisition (Musolino et al. 2019).

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s11050-022-09196-4>.

Acknowledgements We thank Alexander Williams, Jeff Lidz, Tyler Knowlton, Yu’an Yang, Anissa Zaitsu, Dan Goodhue, Rémi Varloot, the Modality Group, the Language Acquisition Lab at the University of Maryland, Florian Schwarz, and two anonymous reviewers for helpful comments and suggestions, as well as our research assistants Monica Pizzo, Joon Lee, David Whitcomb, Madisen Fong, and Avni Gulrajani.

Funding This project is supported in part by NSF grant #BCS-1551628.

References

- Acredolo, Curt, and Karen Horobin. 1987. Development of relational reasoning and avoidance of premature closure. *Developmental Psychology* 23(1): 13–21. <https://doi.org/10.1037/0012-1649.23.1.13>.
- Austin, J. L. 1975. *How to do things with words*. Oxford: Oxford University Press.
- van der Auwera, Johan. 2001. On the typology of negative modals. In *Perspectives on negation and polarity items*, eds. Jack Hoeksema, Hotze Rullmann, Víctor Sánchez-Valencia, and Ton van der Wouden, 23–48. Amsterdam: John Benjamins. <https://doi.org/10.1075/la.40.03auw>.
- Baker, C. L. 1979. Syntactic theory and the projection problem. *Linguistic Inquiry* 10(4): 533–581. <https://www.jstor.org/stable/4178133>.
- Barner, David, and Asaf Bachrach. 2010. Inference and exact numerical representation in early language development. *Cognitive Psychology* 60(1): 40–62. <https://doi.org/10.1016/j.cogpsych.2009.06.002>.
- Barner, David, Neon Brooks, and Alan Bale. 2011. Accessing the unsaid: The role of scalar alternatives in children’s pragmatic inference. *Cognition* 118(1): 84–93. <https://doi.org/10.1016/j.cognition.2010.10.010>.
- Barr, Dale J., Roger Levy, Christoph Scheepers, and Harry J. Tily. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68(3): 255–278. <https://doi.org/10.1016/j.jml.2012.11.001>.
- Bates, Douglas, Martin Mächler, Ben Bolker, and Steve Walker. 2014. Fitting linear mixed-effects models using lme4. arXiv preprint arXiv:1406.5823.
- Berwick, Robert C. 1985. *The acquisition of syntactic knowledge*. Cambridge: MIT Press.
- Bindra, Dalbir, Kenneth A. Clarke, and Thomas R. Shultz. 1980. Understanding predictive relations of necessity and sufficiency in formally equivalent “causal” and “logical” problems. *JEP: General* 109(4): 422–443. <https://doi.org/10.1037/0096-3445.109.4.422>.
- Bochnak, M. Ryan. 2015. Unspecified modality in Washo. In *Proceedings of the eighteenth and nineteenth workshop on structure and constituency in the languages of the Americas*, eds. Natalie Weber and Sihwei Chen, 3–17. Vancouver: UBCWPL.
- Brockmann, Saskia, Sara McConnell, Valentine Hacquard, and Jeffrey Lidz. 2018. Children’s comprehension of pronouns and definites. In *Proceedings of Sinn und Bedeutung 22*, eds. Uli Sauerland and Stephanie Solt, 259–276. Berlin: ZAS.
- Brown, Penelope, and Stephen C. Levinson. 1987. *Politeness: Some universals in language usage*. Cambridge: Cambridge University Press.
- Byrnes, James P., and Michelle A. Duff. 1989. Young children’s comprehension of modal expressions. *Cognitive Development* 4(4): 369–387.
- Cariani, Fabrizio, and Paolo Santorio. 2017. Will done better: Selection semantics, future credence, and indeterminacy. *Mind* 127(505): 129–165. <https://doi.org/10.1093/mind/fzw004>.

- Chierchia, Gennaro, Stephen Crain, Maria Teresa Guasti, Andrea Gualmini, and Luisa Meroni. 2001. The acquisition of disjunction: Evidence for a grammatical view of scalar implicatures. In *Proceedings of the 25th Boston University conference on language development*, eds. A. H.-J. Do, L. Domínguez, and A. Johansen, 157–168. Somerville: Cascadilla Press.
- Cinque, Guglielmo. 1999. *Adverbs and functional heads: A cross-linguistic perspective*. Oxford: Oxford University Press.
- Clark, Eve V. 1987. The principle of contrast: A constraint on language acquisition. In *Mechanisms of language acquisition*, ed. B. MacWhinney, 1–33. Hillsdale: Lawrence Erlbaum Associates.
- Coates, Jennifer. 1988. The acquisition of the meanings of modality in children aged eight and twelve. *Journal of Child Language* 15(2): 425–434. <https://doi.org/10.1017/S0305000900012435>.
- Cournane, Ailís. 2015. Modal development: Input-divergent L1 acquisition in the direction of diachronic reanalysis. PhD dissertation, University of Toronto.
- Cournane, Ailís. 2021. Revisiting the epistemic gap: Evidence for a grammatical source. *Language Acquisition* 28(3): 215–240. <https://doi.org/10.1080/10489223.2020.1860054>.
- Crain, S. 2012. *The emergence of meaning*. Cambridge: Cambridge University Press.
- Crain, Stephen, and Rosalind Thornton. 1998. *Investigations in universal grammar*. Cambridge: MIT Press.
- Crain, Stephen, Weijia Ni, and Laura Conway. 1994. Learning, parsing and modularity. In *Perspectives on sentence processing*, eds. Charles Clifton Jr., Lyn Frazier, Keith Rayner, and Charles Clifton, 443–467. Hove, UK: Psychology Press.
- Deal, Amy Rose. 2011. Modals without scales. *Language* 87(3): 559–585.
- De Haan, F. 1997. *The interaction of modality and negation: A typological study*. Milton Park: Taylor & Francis.
- De Haan, Ferdinand. 2011. Disambiguating modals: Constructions and must. Manuscript, Oracle.
- Dell, F. 1981. On the learnability of optional phonological rules. *Linguistic Inquiry* 12(1): 31–37. <https://www.jstor.org/stable/4178203>.
- Dieuleveut, Anouk. 2021. Finding modal force. PhD dissertation, University of Maryland.
- Dieuleveut, Anouk, Annemarie van Dooren, Ailís Cournane, and Valentine Hacquard. 2019a. Learning modal force: Evidence from children’s production and input. In *Proceedings of the 2019 Amsterdam Colloquium*, 111–122.
- Dieuleveut, Anouk, Annemarie van Dooren, Ailís Cournane, and Valentine Hacquard. 2019b. Acquiring the force of modals: Sig you guess what sig means? In *Proceedings of the 43rd annual to the Boston University Conference on Language Development (BUCLD43)*, eds. Megan M. Brown and Brady Dailey, 189–202.
- Dieuleveut, Anouk, Ailís Cournane, and Valentine Hacquard. 2021. Finding the force: A novel word learning experiment with modals. *Experiments in Linguistic Meaning* 1: 136–146. <https://doi.org/10.3765/elm.1.4883>.
- van Dooren, Annemarie, Anouk Dieuleveut, Ailís Cournane, and Valentine Hacquard. 2017. Learning what ‘must’ and ‘can’ must and can mean. In *Proceedings of the 21st Amsterdam colloquium*, eds. Alexandre Cremers, Thom van Gessel, and Floris Roelofsen, 225–234.
- van Dooren, Annemarie, Anouk Dieuleveut, Ailís Cournane, and Valentine Hacquard. 2022. Figuring out root and epistemic uses of modals: The role of the input. *Journal of Semantics*.
- Drubig, Hans Bernard. 2001. On the syntactic form of epistemic modality. Manuscript, University of Tübingen.
- Dudley, Rachel E. 2017. The role of input in discovering presuppositions triggers: Figuring out what everybody already knew. PhD dissertation, University of Maryland.
- von Stechow, Kai, and Sabine Iatridou. 2008. How to say *ought* in foreign: The composition of weak necessity modals. In *Time and modality*, eds. Jacqueline Guéron and Jacqueline Lecarme, 115–141. Dordrecht: Springer. https://doi.org/10.1007/978-1-4020-8354-9_6.
- Gillette, Jane, Henry Gleitman, Lila Gleitman, and Anne Lederer. 1999. Human simulations of vocabulary learning. *Cognition* 73(2): 135–176. [https://doi.org/10.1016/s0010-0277\(99\)00036-0](https://doi.org/10.1016/s0010-0277(99)00036-0).
- Grice, H. P. 1975. Logic and conversation. In *Speech acts*, ed. J. P. Kimball. Vol. 3 of *Syntax and Semantics*, 41–58. Leiden: Brill. https://doi.org/10.1163/9789004368811_003.
- Gualmini, Andrea, and Vincenzo Moscati. 2009. The early steps of modal and negation interactions: Evidence from child Italian. In *Romance languages and linguistic theory*, eds. Enoch O. Aboh, Elisabeth van der Linden, Josep Quer, and Petra Sleeman, 131–144. Amsterdam: John Benjamins. <https://doi.org/10.1075/rllt.1.07gua>.
- Gualmini, Andrea, and Bernhard Schwarz. 2009. Solving learnability problems in the acquisition of semantics. *Journal of Semantics* 26(2): 185–215. <https://doi.org/10.1093/jos/ffp002>.

- Hackl, Martin. 1998. On the semantics of “ability attributions”. Manuscript, MIT.
- Hacquard, Valentine. 2011. Modality. In *Semantics: An international handbook of natural language meaning*, eds. C. Maienborn, K. von Heusinger, and P. Portner, 1484–1515. Berlin: Mouton de Gruyter.
- Hirst, William, and Joyce Weil. 1982. Acquisition of epistemic and deontic meaning of modals. *Journal of Child Language* 9(3): 659–666. <https://doi.org/10.1017/S0305000900004967>.
- Hoffmann, T. R. 1966. Past tense replacement and the English modal auxiliary system. Harvard Computation Laboratory report NSF 17.
- Horn, Laurence R. 1972. On the semantic properties of logical operators. PhD dissertation, UCLA.
- Iatridou, Sabine, and Hedde Zeijlstra. 2013. Negation, polarity, and deontic modals. *Linguistic Inquiry* 44(4): 529–568. https://doi.org/10.1162/ling_a_00138.
- Jeretić, Paloma. 2018. Evidence for children’s dispreference for weakness: A corpus study. Manuscript, New York University.
- Karmiloff-Smith, Annette. 1979. Micro-and macrodevelopmental changes in language acquisition and other representational systems. *Cognitive Science* 3(2): 91–117. https://doi.org/10.1207/s15516709cog0302_1.
- Koring, Loes, Luisa Meroni, and Vincenzo Moscati. 2018. Strong and weak readings in the domain of worlds: A negative polar modal and children’s scope assignment. *Journal of Psycholinguistic Research* 47(6): 1193–1217. <https://hdl.handle.net/1721.1/131876>.
- Kratzer, Angelika. 1981. Partition and revision: The semantics of counterfactuals. *Journal of Philosophical Logic* 10(2): 201–216. <https://doi.org/10.1007/BF00248849>.
- Kratzer, Angelika. 1991. Modality. In *Semantics: An international handbook of contemporary research*, eds. C. Maienborn, K. von Heusinger, and P. Portner, 639–650. Berlin: Mouton de Gruyter.
- Kratzer, Angelika. 2012. *Modals and conditionals: New and revised perspectives*. Oxford: Oxford University Press.
- Kuczaj, Stan A., and Michael P. Maratsos. 1975. What children can say before they will. *Merrill-Palmer Quarterly* 21(2): 89–111. <http://www.jstor.org/stable/23083972>.
- Landau, Barbara, and Lila Gleitman. 1985. *Language and experience: Evidence from the blind child*. Vol. 8 of *Cognitive science series*. Cambridge: Harvard University Press.
- MacWhinney, Brian. 2000. *The CHILDES project: The database*, Vol. 2. Hove, UK: Psychology Press.
- Mandelkern, Matthew Aaron, Ginger Schultheis, and David Boylan. 2015. I believe I can ϕ . In *Proceedings of the 20th Amsterdam Colloquium*, eds. Thomas Brochhagen, Floris Roelofsen, and Nadine Theiler, 256–265.
- Manzini, M. Rita, and Kenneth Wexler. 1987. Parameters, binding theory, and learnability. *Linguistic Inquiry* 18(3): 413–444. <https://www.jstor.org/stable/4178549>.
- Matthewson, Lisa. 2013. Gitksan modals. *International Journal of American Linguistics* 79(3): 349–394. <https://doi.org/10.1086/670751>.
- Moscati, Vincenzo, and Stephen Crain. 2014. When negation and epistemic modality combine: The role of information strength in child language. *Language Learning and Development* 10(4): 345–380. <https://doi.org/10.1080/15475441.2014.880640>.
- Moscati, Vincenzo, and Andrea Gualmini. 2007. More facts that isomorphism cannot explain. In *Proceedings of Semantics and Linguistic Theory* 17, 202–209.
- Musolino, Julien. 2006. On the semantics of the Subset Principle. *Language Learning and Development* 2(3): 195–218. https://doi.org/10.1207/s15473341lld0203_3.
- Musolino, Julien, Kelsey Laity d’Agostino, and Steve Piantadosi. 2019. Why we should abandon the semantic subset principle. *Language Learning and Development* 15(1): 32–46. <https://doi.org/10.1080/15475441.2018.1499517>.
- Noveck, Ira A. 2001. When children are more logical than adults: Experimental investigations of scalar implicature. *Cognition* 78(2): 165–188. [https://doi.org/10.1016/s0010-0277\(00\)00114-1](https://doi.org/10.1016/s0010-0277(00)00114-1).
- Noveck, Ira A., Simon Ho, and Maria Sera. 1996. Children’s understanding of epistemic modals. *Journal of Child Language* 23(3): 621–643. <https://doi.org/10.1017/S0305000900008977>.
- Orita, Naho, Rebecca McKeown, Naomi H. Feldman, Jeffrey Lidz, and Jordan Boyd-Graber. 2013. Discovering pronoun categories using discourse information. In *Proceedings of the annual meeting of the Cognitive Science Society*. 6 pp. <http://lntool.github.io/NSF-projects/CCF-1018625/papers/Pronouns.pdf>.
- Ozturk, Ozge, and Anna Papafragou. 2015. The acquisition of epistemic modality: From semantic meaning to pragmatic interpretation. *Language Learning and Development* 11(3): 191–214. <https://doi.org/10.1080/15475441.2014.905169>.
- Papafragou, Anna. 1998. The acquisition of modality: Implications for theories of semantic representation. *Mind & Language* 13(3): 370–399. <https://doi.org/10.1111/1468-0017.00082>.

- Peterson, Tyler. 2010. Epistemic modality and evidentiality in Gitksan at the semantics-pragmatics interface. PhD dissertation, University of British Columbia.
- Piantadosi, Steven T. 2011. Learning and the language of thought. PhD dissertation, MIT.
- Piantadosi, Steven T., Joshua B. Tenenbaum, and Noah D. Goodman. 2012a. Modeling the acquisition of quantifier semantics: A case study in function word learnability. Manuscript, University of Rochester.
- Piantadosi, Steven T., Joshua B. Tenenbaum, and Noah D. Goodman. 2012b. Bootstrapping in a language of thought: A formal model of numerical concept learning. *Cognition* 123(2): 199–217. <https://doi.org/10.1016/j.cognition.2011.11.005>.
- Piéraut-Le Bonniec, Gilberte. 1980. *The development of modal reasoning genesis of necessity and possibility notions*. Cambridge: Academic Press.
- Pinker, Steven. 1979. Formal models of language learning. *Cognition* 7(3): 217–283. [https://doi.org/10.1016/0010-0277\(79\)90001-5](https://doi.org/10.1016/0010-0277(79)90001-5).
- Rakoczy, Hannes, Felix Warneken, and Michael Tomasello. 2007. “This way!”, “No! That way!”—3-year-olds know that two people can have mutually incompatible desires. *Cognitive Development* 22(1): 47–68. <https://doi.org/10.1016/j.cogdev.2006.08.002>.
- Rasin, Ezer, and Athulya Aravind. 2020. The nature of the semantic stimulus: The acquisition of every as a case study. *Natural Language Semantics* 29: 339–375. <https://doi.org/10.1007/s11050-020-09168-6>.
- R Core Team. 2013. R: A language and environment for statistical computing. <https://www.r-project.org/>.
- Repacholi, Betty M., and Alison Gopnik. 1997. Early reasoning about desires: Evidence from 14- and 18-month-olds. *Developmental Psychology* 33(1): 12–21. <https://doi.org/10.1037//0012-1649.33.1.12>.
- Robinson, Elizabeth J., Martin G. Rowley, Sarah R. Beck, Dan J. Carroll, and Ian A. Apperly. 2006. Children’s sensitivity to their own relative ignorance: Handling of possibilities under epistemic and physical uncertainty. *Child Development* 77(6): 1642–1655. <https://doi.org/10.1111/j.1467-8624.2006.00964.x>.
- Ruffman, T., J. Aitken, A. Wilson, A. Puri, and M. Taumoepeau. 2018. A re-examination of the broccoli task: Implications for children’s understanding of subjective desire. *Cognitive Development* 46: 79–85.
- Rullmann, Hotze, Lisa Matthewson, and Henry Davis. 2008. Modals as distributive indefinites. *Natural Language Semantics* 16(4): 317–357. <https://www.jstor.org/stable/23748121>.
- Schaeffer, Jeannette, and Lisa Matthewson. 2005. Grammar and pragmatics in the acquisition of article systems. *Natural Language & Linguistic Theory* 23(1): 53–101. <https://doi.org/10.1007/s11049-004-5540-1>.
- Searle, John R. 1969. *Speech acts: An essay in the philosophy of language*. Cambridge: Cambridge University Press.
- Skordos, Dimitrios, and Anna Papafragou. 2016. Children’s derivation of scalar implicatures: Alternatives and relevance. *Cognition* 153: 6–18. <https://doi.org/10.1016/j.cognition.2016.04.006>.
- Snedeker, Jesse. 2000. Cross-situational observation and the semantic bootstrapping hypothesis. In *Proceedings of the thirtieth annual child language research forum*, ed. Eve V. Clark. Stanford: Center for the Study of Language and Information. 7 pp.
- Snedeker, Jesse, Lila Gleitman, and Michael Brent. 1999. The successes and failures of word-to-world mapping. In *Proceedings of the twenty-third Boston University Conference on Language Development*, eds. A. Greenhill, M. Hughes, H. Littlefield, and H. Walsh. Somerville: Cascadilla Press. 12 pp.
- Stalnaker, R. 1978. Assertion. *Syntax and Semantics* 9: 315–332.
- Stromswold, Karin J. 1990. Learnability and the acquisition of auxiliaries. PhD dissertation, MIT.
- Theakston, Anna L., Elena V. Lieven, Julian M. Pine, and Caroline F. Rowland. 2001. The role of performance limitations in the acquisition of verb-argument structure: An alternative account. *Journal of Child Language* 28(1): 127–152.
- Thornton, Rosalind, and Graciela Tesan. 2013. Sentential negation in early child English. *Journal of Linguistics* 49(2): 367–411. <https://doi.org/10.1017/S0022226712000382>.
- Wellman, Henry M., and Jaqueline D. Woolley. 1990. From simple desires to ordinary beliefs: The early development of everyday psychology. *Cognition* 35(3): 245–275. [https://doi.org/10.1016/0010-0277\(90\)90024-e](https://doi.org/10.1016/0010-0277(90)90024-e).
- White, Aaron S., Philip Resnik, Valentine Hacquard, and Jeffrey Lidz. 2016. The contextual modulation of semantic information. Manuscript, LingBuzz. https://lingbuzz.net/lingbuzz/003188/current.pdf?_s=IsOs-TEnLbcihMd.
- Xu, Fei, and Joshua B. Tenenbaum. 2007. Word learning as Bayesian inference. *Psychological Review* 114(2): 245–272. <https://doi.org/10.1037/0033-295X.114.2.245>.
- Yanovich, Igor. 2016. Old English **motan*, variable-force modality, and the presupposition of inevitable actualization. *Language* 92(3): 489–521.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Anouk Dieuleveut¹  · **Annemarie van Dooren**² · **Ailís Cournane**³ · **Valentine Hacquard**⁴

✉ A. Dieuleveut
anouk.dieuleveut@gmail.com

A. van Dooren
annemarievdooren@gmail.com

A. Cournane
cournane@nyu.edu

V. Hacquard
hacquard@umd.edu

¹ Université Paris Cité, CNRS, LLF, Paris, France

² University of Groningen, Groningen, The Netherlands

³ New York University, New York, USA

⁴ University of Maryland, College Park, MD, USA