

Revealing pragmatic processes through a one-word answer: When the French reply Si

Ira Noveck, Nicolas Petit, Ye Tian, Giuseppina Turco

▶ **To cite this version:**

Ira Noveck, Nicolas Petit, Ye Tian, Giuseppina Turco. Revealing pragmatic processes through a one-word answer: When the French reply Si. Journal of Memory and Language, 2021, 120, pp.104245. 10.1016/j.jml.2021.104245 . hal-03494087

HAL Id: hal-03494087
https://hal.science/hal-03494087

Submitted on 20 Dec 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Revision 4 submitted to JML 23 March 2021

Revealing pragmatic processes through a one-word answer: When the French reply *Si*

Ira Noveck[1], Nicolas Petit[2],

Ye Tian[1], Giuseppina Turco[1]

[1]Université de Paris, LLF, CNRS
Paris, France

[2] Centre Hospitalier Le Vinatier
Lyon, France

Highlights

- Surveys the way negatively-polarized questions prompt response particles across languages.
- Proposes how contrapositive answers can be revealing of pragmatic contributions.
- Shows how hallmark pragmatic effects can be extended to conventional expressions.
- Includes a remarkable finding in which 4-year-olds are faster than 6-year-olds.

Abstract

Like many languages, European French has a contrapositive response option (*Si*) to reject the negative content of a question and to express accord with the questioner's implicit affirmative. Consider the question "Barack does not eat meat?" (in French) where the response *Si* indicates that he does. Inspired by Gricean analyses, we view *Si* as an expression that includes a pragmatic component. Based on extant studies that typically show that a pragmatic inference is cognitively costly, we predicted that the articulation of *Si* ought to appear costly compared to felicitous *Oui* and *Non* answers. We created an original task that enjoins a participant to remove a box's cover (while searching for a candy) before hearing a puppet's question. In the critical Negative-Si (NS) condition, the participant finds the candy in, say, a white box (when two boxes are under consideration) and the interlocutor-puppet's negative question is *It is not in the white box?* Besides rates of accurate responses, our main dependent variable was Response Reaction Times (RRT's), viz. the time to naturally *voice* an answer (*Si* in this case). Controls were the Affirmative-Oui (AO), Affirmative-Non (AN), and Negative-Non (NN) conditions. Importantly, the puppet began each trial with one of three kinds of prior belief, a) by declaring that the candy is surely in, or; b) surely not in, the to-be-presented box or; c) by saying "I don't know where it is." These were included to determine whether answerers consider the questioner's prior epistemic state when responding. Experiment 1 compared 6-year-olds to adults and found that i) proficient uses of *Si* are costly with respect to the other three conditions and that; ii) answers in the wake of a "I don't know where it is" prompt slowdowns when compared to the other two declarations. Both findings are consistent with our pre-registered predictions. Four-year-olds, investigated in Experiment 2, pattern almost identically with the 6-year-olds, with one major exception. Their *fastest* response occurs when answering *Si*, leading to a unique developmental effect. Our account for this finding is that four-year-olds rely on a minimally semantic representation of *Si*, which encodes disagreement between the negative content of the question and the facts. We propose that there are pragmatic processes intrinsic to *Si* – which ultimately signals *agreement* with the questioner's implicit affirmative -- and that mastering these requires greater maturity.

Keywords

Answering systems, negative questions, contrapositive answers, Gricean communication, pragmatic development, Theory of Mind

1. Introduction

Imagine a young couple, Michelle and Barack, on their first date as they get ready to order a meal. Michelle notices that Barack is considering a vegan salad at which point she asks him (1):

(1) Do you not eat meat?

By convention in English, a carnivorous Barack would probably say "Yes" (or "Yes, I do") while a vegetarian or vegan Barack would probably say "No" (or "No, I don't"). Even though the *accompanying statement* (the "I do" or "I don't") for each might clarify the *responsive* (the *Yes* or *No*), these anglophone responses prompt ambiguity because (as we spell out shortly) one does not know what the answer is referring to, the negative content in the question or its implicit affirmative. Interestingly, several languages have answering options that transcend these ambiguity concerns through a third offering, i.e., a response particle that goes beyond *Yes* or *No*. For example, if the above exchange were to take place in European French (*Vous ne mangez pas de viande ?*) and indeed Barack does eat meat, the natural response would be neither *Oui* (Yes) nor *Non* (No), but *Si*. It is this *contrapositive* response (Choi, 1991) that is our paper's focus.

The remainder of this Introduction is divided into three sections. In the first section, we will discuss in greater depth the European French response *Si* by explaining when it arises and how this sort of disambiguating expression is found sporadically across languages. This section will also present the findings from prior work that addresses answering systems, contrapositive responses and other related phenomena. The second more theoretical section begins by describing two semantic accounts that cover the contrapositive response. We then present our approach which argues that *Si* is an expression that contains a pragmatic procedural component that indicates that the answerer is, not only disagreeing with the negative content of the question but is also, agreeing with the inquirer's implicit affirmative. In the last section, we will introduce our original paradigm which asks polar questions about the presented contents of a box. Using our experimental pragmatic approach, we will lay out specific predictions, which draw on findings from extensive work with conversational implicatures and other extant literature documenting how pragmatic processing typically entails informational gain with a cognitive cost. We will make the prediction that mature uses of *Si* are more demanding than standard answers and that one should find evidence indicating that mastering a *Si* response among younger children calls on growing pragmatic abilities. This last section also describes our two Experiments. In Experiment 1, we compare adults to 6-year-olds, with the initial expectation being that 6-year-olds will have a relatively undeveloped mastery of such responses. In light of data revealing that 6-year-olds are largely similar to adults, albeit slower overall, Experiment 2 ran the same study with

4-year-olds. Data from this experiment indeed reveal that our youngest participants are unlike the adults and six-year-olds when they are in position to answer *Si*.

1.1. Empirical background

All languages appear to have straightforward expressions for answering affirmatively-worded questions, such as *Do you eat meat?* Some examples are *Yes* and *No* in English, *Oui* and *Non* in French, as well as *ing* and *ani* in Korean. As we describe below, variations across languages arise when one considers responses to negatively-polarized questions.

We begin by underlining how the affirmative reply *Yes* to negatively-polarized questions like the one in (1) is ambiguous. A *Yes* could be either confirming the question (*Yes* confirms that *Barack does not eat meat*) or it could be referring to the question's (or questioner's) implicit affirmative proposition (*Barack does indeed eat meat*). A negative reply *No* to (1) is triply ambiguous and thus potentially even more confusing. That is, *No* could lead the questioner to draw an affirmative through a double negation (*it is not the case that Barack does not eat meat*) or as a negative answer to the implied affirmative. Another possible interpretation of a *No* to (1) is that it is echoing the negation in the question (to say, *Barack indeed does not eat meat*). One can see how a second advantage of having the European French contrapositive option *Si* in an answering system is that it makes the negative response to a negative question clearer as well. A *Non* to (1) in European French readily means that Barack does not eat meat.

The ambiguity raised by negative questions can be seen in other, more structural, ways. Following Pope (1973), scholars typically distinguish between two systems with respect to possible answers to negative questions: one is referred to as a *truth-based* (or as an *Agree/Disagree*) system and the other as a *polarity-based* (or as a *Positive/Negative* [P/N]) system (see Akiyama, 1979; Choi, 1991; Claus et al., 2017; Jones, 1999; Li et al., 2016).[1] In a truth-based system, the responder accepts the truth value of the question (almost as if it were a statement), at which point a positive reply indicates that the statement is true and a negative reply indicates that it is false. If (1) were posed to a vegetarian Barack in Korean, he would reply with the equivalent of a "Yes" (*ïng*), indicating that he is *agreeing with* the negative content of the question (*yes, what you literally said – I don't eat meat -- is the case*). A negative reply to (1), *ani*, would indicate that he is disagreeing with it (*what you literally said – I don't eat meat -- is not the case*). To put it another way, this system addresses the question taken literally practically so that an affirmative or a negative answer would indicate that the answerer agrees or disagrees with it, respectively. Truth-based systems are rather common. Other languages that use

[1] Li et al. (2016) argue that this distinction is not hard and fast.

the truth-based system include Japanese, Navajo, Georgian, Polish, and Armenian, to name a few (see Asatiani, 1999; Jones, 1999).

In a polarity-based system (like English), a negative question indicates that there was "an original positive belief which can be conveyed by a positive sentence" (Jones, 1999). As far as Michelle's question in (1) is concerned, it is as if she had formerly thought that "You [Barack] eat meat" before expressing a refutation of that initial proposition through a negative question. Once the affirmative is on the table, so to speak, a *Yes* or *No* response from Barack appears to address that. A *Yes* in the Polarity system is addressing the implicit affirmative (*you eat meat*) and a *No* is rejecting it. In this way, the answer addresses the questioner's (initial) affirmative. Arguably, it is the negative question, like Michelle's in (1), that makes intentions between questioners and answerers readily available. As Akiyama (1979, p. 487) proposes in (2), answers to negative questions in the Polarity system reveal something about the answerer's intentions, too:

(2) English speakers answer *yes* or *no*, depending on their underlying intention about the matter. [In responding to *Aren't you going*], they answer *Yes, I am going*, if they intend to go, and *No, I am not going*, if they don't.

Typologists further point out that the responsive (the *yes* or *no* part of answers) and the accompanying statements (the *I am* or *I am not* part) agree with each other in Polarity systems (see Jones, 1999). Other languages that use the Polarity system include French, German, Hebrew, Swedish, Italian among many others.

Returning to European French, the answer *Si* addresses two features of the negative question in one stroke. It is disagreeing with the negative content of the question *and* it is answering affirmatively to the inferable belief that was arguably representative of the questioner's prior intention. Other languages that have such contrapositives include German (*doch*), Dutch (*jawel*), Swedish (*jo*), Icelandic (*ju*), and Hungarian (*dehogynem*).[2] Interestingly, the response option *Si* does not exist in Quebecois French (see Noveck, 2001a for a summary of languages that have contrapositives following up on a *Linguist list* query). Another reported historical-linguistic fact is that *yes* used to be a contrapositive in Old English, which reserved *yea* for affirmative queries, such as "Are you hungry?" (Akiyama, 1979).

[2] Choi's (1991) seminal paper describes how Korean, which uses a truth-based system, also provides a contrapositive option (*mace*). After an exchange between Jong-Bok Kim (Kyung Hee University) and the first author, however, it appears that this might be a controversial position. Kim writes that "[Choi was probably referring to] *mac-a* [which] means *you are right* or *what you say is correct*. This expression is thus not a response particle though it might be used as an agreeing predicate not only to a positive but also to a negative polar question."

The psycholinguistic and developmental literatures have not persistently focused on these interesting cases. A handful of researchers, however, have empirically investigated children's competence in mastering different answering systems (along with contrapositive responses) over the years. Akiyama (1979) showed, through an investigation of 3- to 6-year-old Japanese- and English-speaking monolingual and bilingual children, that the English polarity-based system appears easier to master than the Japanese truth-based system and that the difference appears to be due specifically to the Japanese children's difficulty in answering negative questions. In a similar direction, corpus work shows that French children use *Si* by the time they are three years of age (Choi, 1991) while what the author considered to be an equivalent option, *mace*, does not appear to be established among Korean children at that age. All told, one finds that young children have an early facility with contrapositive responses in certain languages, but much work remains to be done.

Given that German speakers use *doch* for a contrapositive answer, it is relevant to consider how this term is used and understood more generally as a discourse particle in developmental studies. Dimroth et al. (2018) elicited corrective utterances from 4- to 6-year-olds as well as adults and showed that the younger participants were more likely than the adults to produce discourse particles such as *doch*, through an elicitation task. For example, a young participant would say *On my picture the girl did DOCH tear up the banknote* (to mean *the girl did indeed tear up the banknote*) whereas an adult would forego the *doch* and make the same point intonationally, as in *On my picture the girl DID tear up the banknote*. Despite a predilection to use *doch* among younger children, *comprehension* tasks with this discourse particle show that children around 5 years of age are still not adultlike. In another intriguing study, Schmerse et al. (2013) set up a paradigm consisting of three steps in which a puppet aims to find a toy that could be in one of two boxes (Box A or Box B). The first step occurs before the puppet even begins its search, when it makes a declaration while standing next to one box. For example, the puppet says "it is certainly not in this box" while standing next to Box A (meaning that the toy is certainly not in Box A). In one critical condition, the second step consists in the participant witnessing the puppet inspect Box A, without sharing what it found inside. The third step consists in hearing the puppet use *doch* to correct her prior belief with a phrase that resembles *the toy is DOCH in the box*, which can be glossed in English to mean *the toy is INDEED in the box* (that is, Box A). It turns out that the 5-year-olds were adultlike at interpreting *doch* in such a condition (they chose Box A to find the toy) as well as in a second condition in which the opening declaration was positive, as in "It is certainly in this box." In contrast, the children were non-adultlike (and they responded at chance levels) in a third condition when the scenario was slightly more complex (i.e. the scenario was identical to the first one described above except that the puppet inspected Box B in the second

step). Schmerse et al. (2013) argue (pp. 128-129) that their data confirm that *doch* generates inferences to listeners about (modifications to) the speaker's belief state while they also show that younger children are less adept than adults at accessing such mental-state inferences. Such findings are unsurprising in light of much developmental data, e.g. with respect to children's growing mastery of using definite and indefinite articles (Maratsos, 1974; Power & Dal Martello, 1986).

1.2. Theoretical Approaches

1.2.1 Semantic classifications

There are two semantic accounts that aim to classify response particles in answering systems, including contrapositives. Consider first Farkas and Bruce (2010), who present a semantic system that distinguishes between two sorts of polarity. One is absolute polarity, which refers to the response particle itself (in their notation, *Yes* would be considered "+" and *No* "-"). The other is relative polarity, which refers to the polarity of the statement with respect to the utterance's context; in their notation, "agree" refers to cases where a statement is consistent with a fact in context and "reverse" refers to cases in which the facts do not cohere with the statement (e.g. Barack eats meat when that is indeed the case would be noted as "agree"; Barack eats meat when he is in fact a vegetarian would be noted as "reverse"). A *Yes* and a *No* in light of a positive question correspond, respectively, with the relative polarities of "agree" and "reverse." In their system, a contrapositive response such as *Si* is (viewed as unique and is) considered a "+" case (meaning that its absolute polarity is positive) but its relative polarity is one that reverses. This can be appreciated when the contrapositive *Si* is used in response to (1); while it is standardly classified as a "+", its relative polarity reverses (because the negative content in the question is inconsistent with reality).

Krifka (2013), who aims to account for the *Ja*, *Nein* and (the contrapositive) *Doch* in German, assumes that response particles pick up a referent introduced by a statement or question. In the case of a positive statement or question (e.g. "Barack eats meat" or "Barack eats meat?") the *antecedent* is positive; for such cases, a *Ja* affirms and a *Nein* rejects it. This much is straightforward. In the case of a negative statement (e.g. "Barack does not eat meat"), the antecedent could be that *Barack does not eat meat* or that *Barack does eat meat*; according to his account, which of these two is more salient will be determined by context. He argues that the contrapositive *Doch*, which of course arises in the wake of the negative statement (as a response to *Barack does not eat meat* or its question form), addresses the affirmative antecedent while serving as a means for a "rejection" (meaning, *Doch* ultimately serves to *refute* the sentence *Barack does not eat meat*).

These two semantic accounts overlap a great deal in that they classify response particles with respect to the polarity of an incoming sentence (or question) and in the way they focus on how the answer affirms or rejects reality. While the two accounts emphasize slightly different aspects of response particles, both provide minimal *descriptions* of *Si,* i.e., they describe under what conditions it is called for. While these are cogent accounts, they do not address at least three features of *Si* responses that are more psychological in nature. First, neither considers just how salient the implicit affirmative must be in the wake of a negative question so that it justifies the *Si* response; in Krifka's account, an affirmative or negative antecedent appear to have equal status after a negative question. Such equanimity is inconsistent with experimental findings that robustly show how affirmative antecedents are salient in negation processing among participants (see Tian & Breheny, 2016). Second, the epistemic status of a negative question's affirmative antecedent is also not considered. That is, when Michelle asks (1), it is arguably to indicate that she would have assumed Barack to be a carnivore but she is not so sure about it (once she saw him perusing the vegan salads). Thirdly, and more generally, these accounts do not address the intention reading aspect of the contrapositive (the sort highlighted by Akiyama's quote in [2]). Being semantic accounts, they are naturally less concerned with the psychological or the intentional dispositions of interlocutors.

1.2.2 Grice-inspired accounts

In order to appreciate *Si* as a form of intention-reading, we turn to classic pragmatic analyses starting with Paul Grice (1989) who viewed utterance interpretation, not as the mere decoding of linguistic signals but, as part of an effort to understand a speaker's intentions. His seminal model developed the notion of *implicature*, which can be viewed as an extralinguistic inference that emerges as a listener aims to determine the speaker's intention. He distinguished between two kinds of implicature. *Conversational implicature* refers to inferences that are not attached to particular content in the sentence. For example, in (3), Peter's reply implicitly indicates to Mary that he would rather not go out.

(3) Mary: Do you want to go to the cinema?
Peter: I have a lot of work to do.

Peter's implicit "No" emerges through inferencing from the utterance in reply to Mary's question plus presumed context. Later, we will describe the experimental findings related to the most investigated case of a conversational implicature, *scalar inference* or *scalar implicature* (consider how *Some*

of those paintings are forgeries invites a pragmatic reading, *Some but not all of those paintings are forgeries*) and how these drive our empirical predictions.

To better categorize conversational implicatures (as extra-linguistic pragmatic inferences that are not attached to particular words), Grice presented a contrasting case -- *conventional implicature*. This category refers collectively to *expressions* that contain information that is pragmatic in nature while not affecting truth conditions. To appreciate these, compare *but* with *and*. While both are conjunctions and share the same truth conditions, *but* is associated with further pragmatic information, viz. a contrast between conjuncts. To appreciate this, compare the two statements in (4):

(4) (a) George Clooney is famous but he is nice.
(b) George Clooney is famous and he is nice.

If a speaker were to say (4a), it reveals that she normally expects that being famous excludes one from being nice; the *but* tells the listener that that expectation is not met here. Note that the conjunction *and* in (4b) does not provide such a contrast. Grice referred to cases like *but* as *conventional* because its pragmatic import is carried by an expression that is encoded in the language.

We propose that the contrapositive response *Si* works like a conventionalized implicature in that it intelligibly provides the (negatively posing) questioner with a more precise and certain answer. The response *Si* indicates that the negative content in the question is to be rejected while it also shows accord with the implicit affirmative antecedent in the questioner's negative query. Much like *but* is a pragmatically rich version of *and*, *Si* is a pragmatically rich version of an affirmative answer. The upshot is that the response *Si* is not only a fascinating expression, it allows one to experimentally investigate a case of conventional implicature.

Grice's is not the only theory of pragmatic inference-generating words. In work inspired by Relevance Theory (Sperber & Wilson, 1986), Blakemore (1987, 2002) distinguished between words encoding a concept (conceptual words), such as *boy* and *runs*, and words encoding an instruction regarding how to process a concept (procedural words). The best examples of words having procedural meaning are discourse connectives, such as *because, so* and *but*. According to Blakemore, the procedures linked to discourse connectives constrain the kind of cognitive effects that the utterance can draw (also see Wilson & Sperber, 1993). To make this clear, consider the following two segments in (5):

(5) a. John can open Tom's safe. b. He knows the combination.

Without a connective, the segment in (5b) can be viewed as evidence for (5a) or it can be considered a consequence of (5a). Note how the connective *after all* would lean one towards the former reading, i.e., where (5b) is evidence for (5a), and how the connective *so* would lean one towards the latter, i.e., where (5b) is a consequence of (5a). This is how connectives constrain. Their presence eliminates ambiguities by indicating with greater specificity the speaker's intended meaning. The answer *Si* similarly constrains because it guides the hearer (the questioner in this case) to more specifically understand the speaker's (the answerer's) intended meaning.

1.3. The current paradigm

Practically speaking, *Si* is an attractive expression to study because it can be used spontaneously by participants as part of a task, making for two convenient dependent variables: a) its rate of appearance, when used appropriately, and; b) the time taken before it is articulated. We began with the hypothesis that cases involving (mature uses of) the response *Si* require more intensive effort than the other standard responses in French so situations that call for it are likely to lead to missed opportunities for using it and relatively longer pre-answering times. Our reasoning is that a prerequisite of articulating *Si* (at least among mature speakers) relies on inferential pragmatic work that occurs before answering. Importantly, we include children because we expected that they would be less likely than adults to resolve situations that call for *Si*. We began our study by comparing 6-year-olds to adults (Experiment 1) and, upon discovering the children's near adultlike facility with *Si*, we pursued our investigation with 4-year-olds (Experiment 2).

1.3.1 Responding naturally

To create a situation in which a spontaneous response (a *Oui, Non* or *Si*) is called for, we set up a scenario in which a puppet asks a question about a box that the participant has just inspected for the presence (or absence) of a candy. The question is either affirmative (e.g. *Il est dans la boîte blanche?* - "It is in the white box?") or negative (e.g. *Il n'est pas dans la boîte blanche?* – "It is not in the white box?").[3] Here, we point out some key features of the task.

Given evidence (the presence or absence of the candy in the inspected box), the participant is poised to answer four kinds of experimental items, two following an affirmative question and two following a negative one. Affirmative questions about the presented box (e.g. *Il est dans la boîte blanche?*) can lead to the response *Oui*, in the event that the candy had been found to be inside that box (we will refer to this henceforth as an "Affirmative-Oui" condition, or AO), or these questions

[3] We use the color white throughout in our examples to simplify the task's description, but there were, of course, a variety of (six) differently colored boxes.

can lead to *Non* if the candy had not been found in the presented box (this is the "Affirmative-Non", or AN, condition). Negative questions about the presented box (e.g. *Il n'est pas dans la boîte blanche?*) can lead to *Si*, in the event that the box indeed contained the candy (this is the "Negative-Si", or NS, condition) or they can lead to *Non*, in the event that the box was empty (this is the "Negative-Non", or NN, condition). We add here that the paradigm allows for many filler items. For example, according to the instructions, an inspected empty box ought to be taken as an indication that the candy is in the other (uninspected) box; so, the paradigm allows for questions about the alternative box as well. One of the paradigm's goals is to make the puppet's questions unpredictable, thus compelling participants to remain alert throughout. We will present the paradigm in detail later.

1.3.2 The puppet's declaration of its prior belief states

Given that we view the response *Si* as a pragmatically rich response that addresses the questioner's epistemic state, we also investigate how an answerer's performance is affected by declarations -- from the eventual questioner -- before evidence is revealed. We pursue this line in light of work that shows how awareness of the speaker's epistemic state prior to a stimulus affects pragmatic inference-making (Breheny et al., 2013; see also Tian et al., 2010). For example, Breheny et al. (2013) recorded participants' eye-gazes in a visual search task under two conditions. In one, a confederate (who would soon be the speaker) and the participant both observe the same sequence of events (e.g., a video showing a woman who put one fork in one Box [A], a second fork in a second Box [B] and then a spoon in the second Box [B]). When participants then hear the speaker's test statement -- "The woman put a fork in Box A and a fork and a spoon into Box B" – their eye-gazes indicate that they anticipated hearing about the spoon. In the second condition, the same kind of sequence is presented during the video presentation, except that the confederate-speaker's view is plainly blocked prior to the placement of the spoon. When participants hear the same test statement as the one above, this time their eye-gazes revealed that the listener-participant did not anticipate hearing about the spoon. This indicates that, in a "fully Gricean" model, a listener considers the speaker's intentions, i.e. her epistemic state, while processing utterances.

In light of such findings (and the technique introduced by Schmerse et al., 2013), we added a second feature to this task, which occurred at the beginning of each trial. Namely, the puppet shares its initial belief about the presence of the candy in one of the boxes by making one of three declarations before a box would be presented, viz. *It is surely in the white box, It is surely not in the white box,* or *I don't know where it is*, which we called *positive*, *negative* and *neutral declarations*, respectively. Our hypothesis was that the engagement of *Si* would be facilitated (e.g. would be faster and more accurate) if the participant-answerer already had information in hand about the questioner's

prior – positive or negative – beliefs. Another way to put it is that the participant-answerer has further inferences to make about the speaker's epistemic state in the *neutral* condition at the moment of the question and should take longer to respond in this condition. From the outset, we did not make specific predictions concerning the negative and positive declarations, other than to say that these *polarized* declarations provide a focus on a box from the beginning of the trial. Polarized declarations were pressed into service to be unlike the trials that begin with a neutral declaration in which the puppet is not committed to knowledge about the target box's content. Essentially our expectation was that a neutral declaration puts the participant in a position to not know what will happen next (to not know what box ought to emerge or what box might or might not have the candy), whereas a polarized declaration prepares the participant to have certain expectations that could be fulfilled or not. We also included this feature of the task for a second, more practical reason, which is that it made the task engaging for both children and adults. Given the highly repetitive nature of the task, adding the initial belief step in each trial was beneficial because it provided variety.

1.3.3 Predictions

We made two classes of predictions with respect to the two main features of the paradigm, both of which rely on prior findings in the experimental pragmatics literature. The first concerns the response *Si* in relation to the other responses. Given that we propose that the contrapositive response *Si* addresses the negative question's intended affirmative as well as its negative content, we anticipate that the comprehension procedure preceding the production of *Si* responses necessarily requires increased cognitive processing on the part of the interlocutor-participant (compared to the other responses). This leads to the first prediction (as pre-registered via OSF, https://osf.io/6vjx8/), which is that *Si*, when used appropriately, will require more time before it is deployed compared to the other response options (i.e. *Oui* and *Non*).

This prediction is inspired by a large corpus of studies showing that the comprehension associated with enriched pragmatic readings of utterances (or test sentences) typically comes with extra cognitive costs compared to linguistically-based readings of the same utterances (which are processed without pragmatic enrichments). This has been the hallmark of the most studied experimental pragmatic object in the literature, scalar inference or scalar implicature, through a range of techniques. To make this clear with one example, consider an item from Breheny et al.'s (2006) self-paced reading task in which participants would read vignettes presented in segments. Critically, these vignettes included a disjunctive phrase (e.g. *meat or fish*) that could be presented in two different contexts. In one, a participant would receive background information that encourages an upper bound reading to the disjunction (a participant reads about an offer for a free hot meal which on that day can be meat

or fish); this was expected to lead to an enriched pragmatic reading (*meat or fish but not both*). In the other, the background is non-committal which allows the disjunction to keep a semantic (lower-bounded) reading (a participant reads about a dietitian who tells children that protein can be found in meat or fish). A disjunctive segment having a pragmatic reading indeed takes longer to process than one having a semantic reading. In general, participants' processing times of sentences that engage pragmatic inferences typically take longer than those same sentences that do not (e.g. Bott and Noveck, 2004; Breheny et al., 2006; Tomlinson et al., 2013, Huang & Snedeker, 2018; including among embedded cases [Chemla et al, 2017]). Producing a task's response (whether it be to provide a truth value or just a button-press in a self-paced reading task) takes longer when one can safely assume that an isolable pragmatic procedure had been engaged beforehand. Overall, these robust findings indicate that pragmatic processing is effortful as it renders the incoming linguistically encoded sentence more informative.

A corollary to this first prediction is that we should find growing difficulty with respect to correctly employing a *Si* response as we investigate performance among increasingly younger participants. This prediction is based on robust findings from a very wide range of studies on linguistic pragmatic development (that were first accrued prior to the adult processing findings described above) showing that children tend to become more pragmatically competent with age (Noveck, 2001b; Noveck et al., 2001; Pouscoulous et al., 2007; Glenwright & Pexman, 2010; Stiller et al., 2015; for the most recent extensive summary of such work, see Katsos et al., 2016; Noveck, 2018) and that this is due, at least partly, to children's growing Theory of Mind abilities, i.e. to their growing ability to attribute thoughts to others (e.g. Leslie et al., 2004; Bosco & Gabbatore, 2017). While we anticipated from the outset that younger participants will behave in a non-adult manner, it was unclear to us how this would manifest itself (so we did not pre-register a *specific* developmental prediction).

The second general pre-registered prediction is based on the prior work concerning the way contextual information is integrated. Assuming that pragmatic language processing is incremental and predictive, as Breheny et al. (2013) claim, any prior information – positive or negative -- that puts the focus on, or raises expectations about, the box in the question ought to facilitate the participant's answer. As far as negative questions are concerned, the linguistic stimulus potentially provides information about a positive state of affairs. If there is an affirmative hypothesis (or even a negative one) from the outset, in the form of the puppet's prior belief, the inferable positive content from the negative question ought to be more readily available than in cases where there is no a priori declaration. As indicated earlier, we did not anticipate a difference between the two polarized cases. We did anticipate a difference between the neutral case, on the one hand, and the two polarized cases taken together, on the other.

2. Experiment 1

We begin by describing the paradigm in greater detail and by specifically focusing on how we measure reaction times, which is the length of time it takes for a participant to respond to a puppet's prerecorded question from the earliest answerable moment, viz. from the time the color of the box in the question is identified. We began by anticipating that 6-year-olds' articulations of *Si* would be more error prone and slower than the adults'.

2.1. Methods

2.1.1. Participants

Twenty-seven monolingual French 6-year-olds and 41 French adults participated in the study. The children's mean age (range) was 6;0 (5;6 – 6;5) and the adults' was 22;3 (18;1 – 29;2). None were reported to have hearing or speech disorders nor were any detectable. The children were recruited from a local private school. Adults were recruited through the first author's research center when he was based in Lyon.

2.1.2. Stimuli

The experiment was presented as a game-like situation in which a participant was told that a candy (*un bonbon*) was hidden in one of two boxes and that, after inspection of one of the boxes, it was the participant's task to respond to a question. The main interlocutor in the experiment was a raggedy doll who appeared on a computer screen as if sitting in the middle of a room. Each trial began with two differently colored boxes in front of the puppet on the laptop's screen -- e.g. a white box in the middle/left side of the screen and a red box in the middle/right side. Six box colors were used: blue, red, orange, green, brown and white (*bleue, rouge, orange, verte, marron,* and *blanche*).

The puppet's question arose after a) the participant had heard the puppet's declared epistemic state about one box's relation to the candy (e.g. *It is surely in the white box*) and b) the participant had inspected one of the boxes. All told, the doll made two utterances per trial (as described in greater detail shortly), the second of which was an affirmative or negative question (e.g. *It is/It is not in the white box?*)

There were four kinds of critical trials: a) Affirmative questions (e.g. *Il est dans la boîte blanche?* "It is in the white box?") whose expected answer was *Oui* (AO trials), in light of the fact that the participant had affirmed the presence of the candy in the mentioned box; b) Affirmative questions whose expected answer was *Non* (AN trials), in light of the fact that the participant had affirmed the absence of the candy in the box mentioned in the question; c) Negative questions (e.g. *Il n'est pas dans la boîte blanche?* "It is not in the white box?"), whose expected answer was *Si* (NS trials), in

light of the fact that the participant had affirmed that the candy was indeed in the box mentioned in the question, and; d) Negative questions whose expected answer was *Non* (NN trials), in light of the fact that the participant had affirmed the absence of the candy in the box mentioned in the question.

With three kinds of declarations (positive, negative, and neutral) and four conditions (AO, AN, NS, NN), and two instances for each, one arrives at 24 experimental trials. These experimental items were interspersed with 16 filler trials, which came in two varieties. The first sort of filler consisted of four items (thus 10% of the total) in which the puppet started off with a polarized declaration about one box before the other to-be-examined box slid out (incidentally, these cases led to answers that corresponded to each of the four types of answers/conditions). These were included so that participants should not routinely expect the declaration to foreshadow which of the two boxes will emerge from the screen. The remaining twelve fillers were trials in which the declaration did concern the emerged-and-examined box (making them similar to the first two steps of experimental trials); however, the trial's question focused on the box remaining on the screen. These twelve fillers were distributed across the three declaration conditions. These were included so that the participant's reactions should not rely on the discovery in the box but should occasionally rely on making deductive inferences from their observations (if the examined box was empty it implies that the candy is in the unexamined box, the one remaining on the screen; likewise, if the examined box contained the candy it implies that the box on the screen is empty). It can be seen that the 24 experimental trials represent a majority (60%) of the session but that the inclusion of the other 16 fillers compel participants to remain attentive throughout the trial. See Figure 1 for a portrayal of the paradigm and Table 1 for some exemplary trials.

- - - - - - - - - - - - - - - -

Insert Figure 1 and Table 1 here

- - - - - - - - - - - - - - - -

The question was presented in its declarative form (*Il est dans la boîte X?*), which was preferred over other types of French question forms such as those with subject inversion (ex. *Est-il dans la boîte X?*) and those with the initial interrogative particle *est-ce que* (ex. *Est-ce qu'il est dans la boîte X?*). This declarative form without subject inversion is typically used for information and confirmation-seeking questions in French, with the aim of eliciting an answer (in this case, a *oui, non* or *si* answer) and asking for confirmation respectively (for more details, see Delais-Roussarie et al. 2015).

All declarations and questions of the experiment were recorded beforehand (for a relevant discussion concerning such materials, see Kerbrat-Orecchioni, 1991). All sentences were recorded

(44.1kHz, 16 Bit) onto a laptop, using a headset microphone AKG C 410 and a Roland Quad Capture audio interface, by a (35-year-old) French monolingual female informant. Recording took place in the sound-proof cabin of the *Laboratoire de Phonétique et Phonologie* (LPP) of the Université de Paris Sorbonne-Nouvelle. The informant was told to read the set of sentences as fluently and naturally as possible. The set of neutral and polarized declarative sentences were realized with a falling contour (Hi L*-L% in F_ToBI annotation, Delais et al. 2015), typical of French statements. Positive and negative questions were realized with an intonation contour typical of French information-seeking questions (see Figures 2[a-b] below), i.e. with an overall rising contour (prenuclear configuration: aL H* followed by the nuclear configuration: H*H%, Delais-Roussarie et al. 2015). Each of the puppet's two sets of utterances -- declarations and questions -- was verified to have a consistent melodic contour throughout by *Praat* software (Boersma & Weenik, 2018).

- - - - - - - - - - - - - - - - -

Insert Figure 2(a-b) here

- - - - - - - - - - - - - - - - -

2.1.3 Procedure

The experimenter and the participant sat face-to-face with a laptop computer having a rotatable screen between them (see Fig. 3 for a summary). The laptop's screen was spun so that the participant faced the screen while the experimenter viewed the back of the screen along with the computer's keyboard. This was designed so that the experimenter could control the onset of the puppet's on-screen statements via Microsoft's Powerpoint Presentation and could carefully choreograph the release of each box (as described in Step 3 below).

- - - - - - - - - - - - - - - - -

Insert Figure 3 here

- - - - - - - - - - - - - - - - -

Behind the screen, the experimenter kept two sets of the six differently colored boxes: one set with a candy in each box and one set whose boxes remained empty. A trial was comprised of the following seven steps: 1) As two differently-colored boxes sat in front of the puppet on the computer screen, she made a declaration (the puppet moved back and forth slightly on the screen as it made a prerecorded statement) about the location of the candy (e.g. *It is surely in the white box*). 2) One of the boxes (usually but not necessarily the mentioned box) moved towards its side of the screen before disappearing "off-stage." 3) The experimenter (who kept a descriptive list of trials and anticipated

what sort of box should mimic the moving box) physically completed the moving box's motion from the screen so that it appeared in real life; i.e. the box on the screen appeared to slide out from the screen and on to the table via a ramp. 4) The participant picked up the box, removed the box's lid, and inspected its contents (which could either contain the candy or be empty). 5) The participant replaced the lid and put the inspected box back on the table. 6) The participant heard the puppet's question (e.g. *It is in the white box? It is not in the white box?*). 7) To conclude each trial, the participant responded to the question. The experimenter took back the box from the table before preparing for the next trial (See Fig. 3 for an illustration of two kinds of trial).

There were two lists of item presentation. One was pseudo-randomized so that a list observed a set of criteria (to be discussed immediately below) and the other was that list inverted. The criteria were as follows: A) Items from each condition were evenly distributed over the session. For example, the NS item that was presented twice with a neutral declaration appeared once in the first half and once in the second half; likewise, for the other items and their accompanying declarations. B) The filler items were also distributed across conditions and declaration types. C) Items that call for a *Si* response never appeared back-to-back. D) An item calling for a *Si* response was not the first to be presented nor was it among the training items. E) Likewise, an item calling for a neutral declaration was not the first presented nor was it among the training items. See the Appendix for exemplary items and our pre-registered experiment for a complete list of experimental and filler items. Each session took about 10 minutes including a brief break midway.

Participants were requested to reply naturally to the puppet's questions. Participants' responses were recorded with the embedded microphone of the laptop, which allowed the experimenter to concentrate on the manipulations behind the scenes. These recordings were later used to report the answers provided by participants, i.e., to determine rates of normative responses and the speed at which the answers were provided. The critical reaction time, what we call the Response Reaction Time (the RRT), is the latency between the earliest detection of the box color in the question and the participant's initial voicing of an answer. The color term, which in French occurs as the last word in the question, is the earliest moment at which one can answer the question (until then both boxes are in play leaving participants uncertain about the content of the puppet's statement). The precise starting point was attributed to the end of the first phoneme of the color word (*orange*: o*range, *bleue*: b*leue, *verte*: v*erte, *blanche*: b*lanche, *marron*: m*arron, and *rouge*: r*ouge). The starting point was determined primarily by identifying the start of the visual signal as detected by *Audacity* software (https://www.audacityteam.org); however, choosing the precise starting point also required listening to the auditory signal itself (See Fig. 4).

- - - - - - - - - - - - - - - -

Insert Figure 4 here

- - - - - - - - - - - - - - - -

2.1.4. Data selection

Analyses concerned only the 24 experimental items (filler items were excluded). Of these 24, it was discovered – and at the moment of proper analysis -- that one experimental item intended to be an AO trial (item 38 in our file "trials" at our registration site) was presented as an AN trial for roughly half of the participants. We determined that removing the item entirely from our analyses was the most cautious route to take. It follows that there were 23 experimental items per participant.

Given our dependent measure, we considered only those responses that began with *Oui, Non* and *Si*. These included a small percentage of experimental responses (10%) across both age groups that came appended with "complements" (such as *Si, il est dans la boite* or *Non, il n'est pas dans la boite*).[4] Otherwise, three 6-year-olds were removed from the analyses because they consistently did not use a response particle (these participants would answer in other ways, e.g., *it's in the red box*, or they would comment before and during the question). Thirty-eight responses (30 responses from the 6-year-olds and 8 from adults) that do not begin with *Oui*, *Non* or *Si* (e.g., responses that include *je ne sais pas* - "I don't know") were also removed from the analysis, leaving a total of 522 responses from 6-year-olds and 935 from adults.

We analyzed the experimental items with respect to 2 age groups (6-year-olds vs. adults), 2 box contents (empty, full) and 2 question types (negative, positive). This description puts us in a position to understand the experimental conditions. For example, a positive question that concerns an examined full box is one that ought to generate a *Oui* response (the AO condition) and a negative question that concerns an examined full box is one that ought to generate a *Si* (the NS condition). The analysis also includes an assessment of participants' behaviors in the wake of the puppet's opening declaration, with the polarized declarations (whether they be positive or negative) being compared to the neutral declaration. We anticipated in our pre-registration that the polarized cases, unlike the neutral cases, would facilitate comprehension or speed of response because these place a focus on a box from the beginning of the trial. In light of a suggestion from a Reviewer, we will also carry out secondary analyses with the declaration variable being treated as having three factors (i.e. rather than viewing them as *neutral* versus *polarized*, we will investigate the three declarations, *neutral, negative* and *positive* or, to be more specific, *neutral vs negative* and *neutral vs positive*).

[4] Analyses that exclude responses with complements (for all analyses in the paper) revealed that they did not affect our outcomes so we include all cases that begin with *Oui, Non* and *Si* (with or without complements).

2.1.5. Statistical analyses

Our dependent measures were ACCURACY and RESPONSE REACTION TIMES (RRTs). The accuracy data were tested statistically using generalized linear mixed effects models with a logit link function in a Bayesian framework[5] (Jaeger, 2008). The variable ACCURACY was coded as 1 (accurate) and 0 (not accurate). For the analysis of the RRTs, the data were fitted with a linear mixed effects model (in a Bayesian framework) that assumes a shifted-log-normal distribution of the data (Nicenboim et al. 2018). This model takes into account two important properties of the distribution of RRTs in psycholinguistic experiments: 1) the data are strongly right-skewed, as one would expect for standard reaction time data; and; 2) the data are constrained to be positive numbers (negative numbers are not possible). Importantly, the RRT analysis was run only on correct responses given by the participants. We removed those data points that were at least 2 standard deviations above and below a specific group's mean for each condition. Overall, for Experiment 1, 1.09% of the data were removed which amounted to another 16 data points (3 for the 6-year-olds, 13 for the adults). This left 519 data points (from the 6-year-olds) and 922 data points (from the adults) for the statistical analysis.

We begin by describing our initial planned analysis with respect to accuracy and RRT. Both Bayesian models were fitted, using the *brms* package (Bürkner, 2018) in R (R Core Team, 2019). For both models, we began our estimations by considering these four main fixed effects - AGE GROUP (adults vs. 6-year-olds), QUESTION (negative vs. positive), BOX CONTENT (empty vs. full), and DECLARATION (neutral vs. polarized) - plus all possible interactions. In the random effects part of the models for SUBJECT, we estimated the main effects for QUESTION (negative vs. positive), BOX CONTENT (empty vs. full), DECLARATION (neutral vs. polarized) and their interactions, as well as the correlations between the random intercepts and slopes. In the random effects part for ITEM, we estimated the intercept, the slope of the variable AGE GROUP (adults vs. 6-year-olds) and the correlation between the intercept and the slope (Baayen et al. 2008; Barr et al. 2013).[6] The priors that were specified for all the model parameters were weakly informative (Nicenboim & Vasishth, 2016).

The level of factors AGE GROUP, QUESTION, BOX CONTENT, and DECLARATION were coded with sum contrasts: 1 for adults and -1 for 6-year-olds (Experiment 1), 1 for 4-year-olds and -1 for 6-year-olds (Experiment 2); 1 for negative question and -1 for positive question; 1 for empty box and -1 for full box; 1 for neutral declaration and -1 for polarized declaration. In the posterior plots, since the

[5] For a more detailed review of the advantages offered by a *Bayesian* analysis over *frequentist* methods, see Wagenmakers (2007), Nicenboim and Vasishth (2016), Sorensen, Hohenstein, & Vasishth (2016), among others.

[6] This was done because, for the random factor SUBJECT, the fixed factor AGE GROUP is, of course, *between*-subject whereas ITEM, which is a within-subject factor, is determined by a systematic variation of the factors QUESTION, CONTENT and DECLARATION.

variables are coded (-1 vs. 1), the "0" is taken as the reference point, which means that there is no difference between the two conditions (for a secondary analysis, in which the neutral and two polarized declarations are treated individually, the predictor DECLARATION was coded with sliding differences: -2/3 for negative vs. 1/3 for neutral; +2/3 for positive vs. -1/3 for neutral). We are interested in the probability that the parameter of interest is greater or smaller than zero.

For each of these parameters, the Bayesian model will provide the estimated mean ($\hat{\beta}$) and the range of the posterior distribution. The 95% credible intervals will define the range within which it can be certain (with a probability of 0.95) that it includes the true value of the parameter.

We defined a weakly informative prior on each of the model coefficients for the fixed and random effects, using a normal distribution with mean 0 and standard deviation 1. We specified 4 Markov chains, each chain containing 6000 iterations. For each of the four chains, we discarded the first half (the so-called warm-up phase) and only considered the second half (in which the chains get stable), leaving us with 12000 iterations in total. In order to evaluate the model, we first looked at the stability of the chain in the post-warmup phase (Lunn et al., 2012). We then checked for indications of convergence by looking at whether the *Rhat* statistics was equal to 1 for all of the model parameters (Gelman et al. 2014).

2.2. Results

2.2.1. Accuracy

In terms of ACCURACY, adults and children performed at ceiling. The adults' accuracy rate surpassed 98% correct and the 6-year-olds surpassed 95%. These high rates of accuracy indicate that the task was generally not difficult for all participants. The rare errors were not revealing of patterns (e.g. when *Si* was expected, errors included cases of *Non* and *Oui* in roughly equal measure).

The Bayesian model calculates a posterior distribution for each of the model parameters. The posterior plot (Fig. 5) together with the output of the model, as given in Table 2 (as well as in Tables 3-5), indicate how much the model quantifies the probability that the effect is greater or smaller than zero. To interpret the posterior distribution plot in Fig. 5, the black dot represents the mean of the distribution, the green bar is the range of distribution, the orange line represents the 95% credible intervals (meaning that 95% of the distribution lies within these bars). For the main effects, the direction of the effect is given by the sign (because we coded these parameters as 1/-1, cf. section 2.1.5). We considered as effects with strong evidence only parameter coefficients whose posterior distribution had at least 95% probability of being greater or smaller than zero.

- -

Insert Tables 2- 5 and Figure 5 here

- -

These outcomes led us to focus on the parameters that yielded results indicating high probabilities that they go beyond the null. Such a high probability was found for the parameter QUESTION, for which the mean accuracy was higher for positive questions (98.9%) than for negative questions (95.3%), ($\hat{\beta}$ =-0.78, CrI = [-1.53, -0.06], P($\hat{\beta}$) > 0 = 0.02; see Fig. 6). None of the other parameters yielded results indicating exceptional probabilities.

- - - - - - - - - - - - - - - - -

Insert Figure 6 here

- - - - - - - - - - - - - - - - -

2.2.2.Response Reaction times (RRTs)

As the coefficients in Fig. 7 and in Table 3 show, the model reveals that the zero lies outside the 95% mark, as credible intervals of their posterior distribution, for the parameters AGE GROUP, QUESTION, and DECLARATION and for the interaction QUESTION BY CONTENT.

- - - - - - - - - - - - - - - - -

Insert Figures 7-11b here

- - - - - - - - - - - - - - - - -

Guided by these parameters, we highlight those that provide reliable effects. For the parameter AGE GROUP, the mean RRT was faster for adults than for 6-year-olds ($\hat{\beta}$ =-0.08, CrI = [-0.12, -0.04], P($\hat{\beta}$) > 0 = 0; see Fig. 8).[7] The adults were roughly 158 ms faster than the 6-year-olds overall (759.1 ms *vs* 916.6 ms, respectively). For the parameter QUESTION (Fig. 9), the mean RRT was faster for positive questions than for negative questions ($\hat{\beta}$ =0.07, CrI = [0.04, 0.11], P($\hat{\beta}$) > 0 = 1). For the parameter DECLARATION (Fig. 10), the mean RRT was faster in the wake of polarized beliefs than for neutral beliefs ($\hat{\beta}$ =0.06, CrI = [0.03, 0.09], P($\hat{\beta}$) > 0 = 1).

The interaction QUESTION BY CONTENT (Fig. 11a) indicates, that in the wake of opening a box that is full, negative questions prompted the slowest RRT's whereas positive questions the fastest; meanwhile, empty boxes prompted RRT's that were between these two extremes. This interaction,

[7] Here, and elsewhere, probabilities that are equal to 0 or 1 are referring to cases where rounding leads to approximately 0 or to approximately 1.

which also gives us information about the timing of each particle response, has 100% probability of being smaller than zero ($\hat{\beta}$ =-0.04, CrI = [-0.08, -0.01], P($\hat{\beta}$) > 0 = 0). The slowest condition – receiving a negative question in the wake of discovering a box that is full -- is precisely the NS condition, the conditions under which the accurate response is *Si.* As summarized above, the fastest response is provided in the full content–positive question AO condition (when the response is *Oui*), the two remaining conditions, i.e. the AN and NN conditions, fall between these two extremes (see Fig. 11b for a presentation based on response).

2.2.3 Secondary analysis of Experiment 1: Accuracy

In light of a reviewer's comment, we carry out a second analysis in which DECLARATION is subdivided into three separate conditions, reflecting the three sorts of declarations (neutral, negative and positive).[8] The Bayesian model calculates a posterior distribution for each of the model parameters, which now presents the DECLARATION condition as neutral vs. positive and neutral vs. negative. Like in the primary analyses, the posterior plot (Fig. 12) together with the output of the model, as given in Table 4 (as well as in Tables 2, 3 and 5), indicate how much the model quantifies the probability that the effect is greater or smaller than zero. These outcomes led us to focus on the parameters that yielded results indicating high probabilities that they go beyond the null. Such a high probability remained for the parameter QUESTION, the mean accuracy was higher for positive questions than for negative questions ($\hat{\beta}$ =-0.78, CrI = [-1.52, -0.05], P($\hat{\beta}$) > 0 = 0.02; see Fig. 6).

The secondary analysis did produce an interaction between polarity particles (QUESTION*CONTENT) and the neutral-negative DECLARATION ($\hat{\beta}$ =1.04, CrI = [-0.19, 2.22], P($\hat{\beta}$) > 0 = 0.96; see Fig. 13) that was not revealed by our original (primary) analysis. As Figure 13 shows, the NS condition (responding *Si*) in this experiment prompted the lowest rate of accurate responses when it followed a neutral declaration (90%), distinguishing it from the high rate of *Si* production in the negative declaration condition. This interaction indicates that producing *Si* under neutral declaration conditions is relatively challenging. Despite this effect, this accuracy rate is still very high; accuracy was at least

[8] In light of our concern for the speaker's (the puppet's) epistemic state, a second – and arguably more psychologically refined -- way to subdivide the DECLARATION condition is to consider whether the puppet's polarized declaration was *right* or *wrong* before receiving the trial's test question. In this analysis, the 'puppet-is-right' condition combines the AO and NS (full) cases following the positive declaration with the AN and NN (empty) cases following the negative declaration; the 'puppet-is-wrong' condition combines the AO and NS (full) cases following the negative declaration with the AN and NN (empty) cases following the positive declaration (see Figure 1). This recombination has the added advantage of balancing the number of negative occurrences (the uses of "not" and the discovery of empty boxes) across the two conditions before the question arises. Of course, neutral cases are unaffected by this reconfiguration. With this re-composition of the DECLARATION condition for the secondary analyses (where *puppet-is-right* replaces "positive" and *puppet-is-wrong* replaces "negative"), the model's outcomes are practically identical for this Experiment as they are in the original (Secondary) analysis. The upshot is that one can consider a positive declaration a proxy for puppet-is-right and a negative declaration a proxy for puppet-is-wrong in these secondary analyses.

95% for all other conditions. None of the other parameters yielded results indicating exceptional probabilities.

- - - - - - - - - - - - - - - - - - -

Insert Figures 12-13 here

- - - - - - - - - - - - - - - - - - -

2.2.4 Secondary analysis of Experiment 1: RRT's

As the coefficients in Fig. 14 below and in Table 5 show, the model with three declarations still reveals that the zero lies outside the 95% mark, as credible intervals of their posterior distribution, for the parameters AGE GROUP ($\hat{\beta}$ =-0.07, CrI = [-0.11, -0.03], P($\hat{\beta}$)> 0 = 0), QUESTION ($\hat{\beta}$ =0.08, CrI = [0.04, 0.11], P($\hat{\beta}$) > 0 = 1), for the interaction QUESTION BY CONTENT ($\hat{\beta}$) -0.05, CrI = [-0.08, -0.01], P($\hat{\beta}$) > 0 = 0.01) and also for the variable DECLARATION (neutral vs. negative: $\hat{\beta}$ =0.13, CrI = [0.05, 0.2], P($\hat{\beta}$)> 0 = 1; neutral vs. positive: $\hat{\beta}$ =-0.1, CrI = [-0.18, -0.02], P($\hat{\beta}$) > 0 = 0.01).

- - - - - - - - - - - - - - - - -

Insert Figures 14-15 here

- - - - - - - - - - - - - - - - -

For the parameter DECLARATION (Fig. 15), the mean RRT was slowest in the wake of neutral beliefs and slower than each of the two other declarations (positive: mean = 883.5 ms, SD = 231.8; neutral: mean = 945.6, SD = 267.5; negative: mean = 835.4, SD = 248.5). These analyses are thoroughly consistent with the findings of the primary analyses while adding greater specificity.

2.2.5 Tertiary analysis of Experiment 1: A focus on the RRT of Si

A second unplanned analysis concerns the speed with which *Si* responses are given as a function of the declaration condition. This analysis was suggested by a reviewer in order to buttress our claim that the relatively slow *Si* response is due to the unique pragmatic inferencing it engenders. The reviewer reasoned that if *Si* is associated with epistemic inferencing, as we claim, then one ought to find that the RRT's of *Si* production ought to vary as a function of the declaration condition. A robust finding in this direction would have the added benefit of assuaging doubts that *Si* is slow for reasons other than pragmatic inferencing. In other words, if one can show that the RRT of *Si* varies as a function of declaration it would indicate that responders are answering as quickly as possible, across

varying epistemic conditions, and not because *Si* is just generically unusual in some uninteresting way.

We therefore investigated the RRT of *Si* with respect to (three levels of) declaration, even though it was not justified by a DECLARATION-related interaction in any of the earlier models. We performed the analyses on 231 *Si* responses for the adults and 129 *Si* responses for the 6-year-olds. Two linear mixed effects models were run separately, one for each age group (see Figure 16). The models fitted logged transformed RRTs of the *Si* responses as a function of DECLARATION (negative, neutral, positive). Subject was treated as a random factor. By-subject random slopes for the effect of declaration were also included. For the six-year-olds' *Si* cases, the model showed that compared to neutral cases (mean=1046 ms; SD=310.2), negative cases (mean=1029 ms; SD=390.6; β =7.08, SE=3.22, *t*-value=2.19, $p<.05$) and positive cases were significantly faster (mean=953.3 ms; SD=264.8; β =7.93, SE=3.13, *t*-value=2.52, $p<.05$), whereas there was no difference between positive and negative (p=.3). This is consistent with findings in the Secondary analysis. For the adult *Si* cases, the model showed a different effect of DECLARATION: negative declarations (mean=868.5 ms, SD=739.7) were significantly slower than neutral declarations (mean= 845.5 ms, SD=209; β =-7.08, SE=3.22, *t*-value=-2.19, $p<.05$) and significantly faster than positive ones (mean=927.8 ms, SD=489.2; β =-7.93, SE=3.13, *t*-value=-2.53, $p<.05$), whereas no difference was found between positive declarations and neutral ones (p=.7).

- - - - - - - - - - - - - - -

Insert Figure 16 here

- - - - - - - - - - - - - - -

2.2.6 Quaternary analysis of Experiment 1: The RRT's in the wake of a negative question

Here, we carry out a more precise statistical test to determine the extent to which the *Si* response in the NS condition is slower than the *Non* response in the NN condition. This third unplanned analysis — suggested by a reviewer who carefully inspected Figure 11b — was carried out to verify statistically that *Si* in the NS condition is the slower of the two. This exploration is relevant for two reasons. One is that discerning the relative RRT speed for this pair, both of which occur in the wake of a negative question, carries theoretical implications since the *Non* in the NN condition refers to the question's negative content and the *Si* in the NS condition refers to the question's affirmative antecedent. If the speeds for these two responses is equivalent, as a semantic analysis would suggest, then one ought to expect a null result. On the other hand, if the *Si* relies on a pragmatic component, as we claim, there should be indications that the *Si* takes longer. The other reason is more general, which is that we predicted that the response *Si* in the NS condition would take longest. While we found an

interaction supporting that claim (see Figure 11a), it would be more convincing if we could provide a specific statistical test comparing the *Si* response to its closest control.

To test for differences between the RRT's of the *Si* response in the NS condition and the *Non* response in the NN condition, we perform a t-test in a Bayesian framework using the package BEST (Kruschke & Meredith 2020) The t-test shows that for both groups taken together (six year olds and adults), that there is a 99% probability that the RRT of the *Si* response in the NS condition is indeed slower than the *Non* in the NN condition is (mean of the posterior for the difference [and the 95% credible intervals of the posterior]: 81.7 [9.9, 157.2]).[9]

2.3. Discussion of Experiment 1

This experiment was designed to investigate natural *Si* responses in the context of an exchange in order to determine whether situations that call for these responses are costly in terms of accuracy and RRT's with respect to its controls. One prediction was that we would find a relative slowdown when *Si* responses are anticipated in the NS condition with respect to the other three responses in the AO, AN, and NN conditions. Another prediction was that there would be a slowdown for trials that began with a neutral declaration relative to the polarized ones. Both sorts of slowdown would indicate that the answerer brings into consideration the questioner's epistemic state. In the NS condition, we propose that the answerer needs to indicate, not only that the negative content in the question is wrong but, that the questioner's implicit affirmative is correct. In the case of a slower response in the wake of a neutral (as opposed to a polarized) declaration, the answerer is disadvantaged temporally by not knowing the questioner's epistemic state prior to the question. Both of these findings were confirmed as per our pre-registered predictions.

The participants' (both the six-year-olds' and adults') answers reveal that the lead-up time to *Si* responses in the Negation-Si (NS) condition do appear to require more cognitive effort than to responses in the other three experimental conditions. This is most evident from the data summarized in Figures 11a and 11b, which show that -- across participants – the RRT is longest before the *Si* response. It is further buttressed by the t-test in the last (quaternary) analysis, which compared the RRT's of the *Si* responses in the NS condition to *Non* responses in the NN condition to statistically test the idea that the latency of the two are not equivalent. The outcome indicated that the RRT of the *Si* response is indeed the slower of the two, thus statistically justifying our claim that it is slowest overall. Another reason this result is important is that it shows that accessing and addressing the

[9] The mean difference as reported by the Bayesian t-test here is slightly more conservative (when compared to the difference between the NS and NN conditions shown in Figure 11b) because the mean values in the t-test are based on a Bayesian estimation (which draws random numbers from a prior to create a posterior).

affirmative antecedent of the negative question is, arguably, at least part of the reason for the slower *Si* response in the NS condition when compared to the *Non* response in the NN condition, whose answer relies on referring to the question's expressed negative content.

Another finding that is consistent with our predictions concerns the way prior beliefs, in the form of polarized declarations, appear to facilitate the speed of response. The data in Fig. 10 show how polarized declarations (a positive or negative declaration about the puppet's expectation) facilitate responses with respect to neutral ones. These findings are maintained when using a model that breaks down the DECLARATION condition into its three kinds. While we had remained modest about the role of "polarized" declarations, the unplanned secondary analysis confirmed our predictions with greater precision. When they were separated into three levels, we continued to find that neutral declarations generally were slower than each of the two polarized ones.

The tertiary analysis investigated *Si* responses separately to determine whether they are affected by epistemic priors. It turned out that they are, indicating that the *Si* response is not just generically slow for some odd reason. Rather, it appears that participants are answering as quickly as possible and that epistemic concerns appear to matter. This finding was unearthed to address a reviewer's concern (it was not planned nor was it justified by prior analyses indicating an interaction) so we do not pursue it further here other than to say that it appears that *Si* itself is affected by the DECLARATION condition.

Although we did not specify our developmental predictions overall, there are two ways one could expect the children to differ from the adults. One is for the children to make substantially more errors in the NS condition compared to adults and the other is that they would take longer to respond than the adults when correctly using the contrapositive. Whereas the 6-year-olds make slightly more errors than the adults (5% errors versus 2%, respectively) and while they are generally slower than the adults overall (by an order of 160 msec), their responses very largely pattern with the adults. Age does not interact with other factors in Experiment 1, with one exception, i.e. the secondary analysis revealed that 6-year-olds were least accurate when the NS condition was combined with a neutral declaration. While this one result is consistent with our predictions, major developmental differences were not in evidence here. Both groups are largely comparable in that the 6-year-olds manifest adult-like behavior with respect to *Si*. This is why, in an effort to uncover developmental effects, we turn to younger children in Experiment 2.

3. Experiment 2

The findings from Experiment 1 determined that the 6-year-olds, while very slightly more error-prone -- and consistently slower -- than the adults overall, were sufficiently advanced in their mastery

of the response system including with respect to the *Si* response. While maintaining our hypothesis that less experienced interlocutors have greater difficulty with the pragmatic procedure intrinsic to *Si*, we sought out a group of younger children, viz. 4-year-olds, in order to determine whether we can uncover evidence of non-adultlike production of *Si* among them. We used the 6-year-olds from Experiment 1 as a comparison. Our expectations remained the same, which were that responses from younger French speakers will reveal difficulties or challenges with *Si* when compared to the older children.

3.1. Methods

3.1.1. Participants

A group of 21 French-speaking 4-year-olds, whose mean age (range) was 4;8 (4;3 – 5;0), was compared with the group of 24 6-year-olds from Experiment 1. They were all tested individually in a quiet room. None were reported to have hearing or speech disorders nor were any difficulties detectable. The children were recruited from a local private school.

3.1.2. Materials and procedure

The materials and procedure were identical to those of Experiment 1.

3.1.3. Data selection and statistical analyses

As in Experiment 1, responses other than those that began with *Oui*, *Non* and *Si* (e.g. *je ne sais pas* - "I don't know") were removed from the analysis. This amounts to 80 removed responses from the 4-year-olds, leaving them with 403 responses (we reiterate that 30 were initially removed from the 6-year-old data for this reason, leaving a total of 544 responses). We then removed data points that fell 2 standard deviations above or below that specific group's mean per condition. Aside from the 3 data points removed from the 6-year-olds' data (as reported for Experiment 1), 3 data points were removed from among the 4-year-olds for Experiment 2, representing 0.65% of the data. Overall, the 6-year-olds yielded 535 data points for the analysis and the 4-year-olds 400.[10] Statistical analyses were carried out using the same general procedure as in Experiment 1.

3.2. Results

3.2.1. Accuracy

[10] The AO item (#38) that was excluded from the analyses of Experiment 1 due to a technical mishap was run as intended in Experiment 2; nevertheless, this item was excluded from the analysis here as well in order to be compatible with the 6-year-olds' data set.

As seen in Fig. 17 (and Table 6) the coefficients show that the zero lies outside the 95% credible intervals of their posterior distribution for the parameters QUESTION and CONTENT. These results show that there are no general age differences in terms of ACCURACY. Like in Experiment 1, there is higher accuracy in responding to positive questions as opposed to negative ones ($\hat{\beta}$ =-0.74, CrI = [-1.44, -0.04], P($\hat{\beta}$) > 0 = 0.02). This is graphically displayed in Fig. 18. We also find that empty boxes produce more accurate responses than full ones for all the children as a group ($\hat{\beta}$ =0.87, CrI = [0.12, 1.65], P($\hat{\beta}$) > 0 = 0.99). Fig. 19 shows how questions concerning empty boxes promote correct responses at a rate of about 97% and that full boxes lead to accuracy rates of roughly 94%. This tendency was present in the prior experiment but could not be confirmed at a high probability.

- - - - - - - - - - - - - - - -

Insert Figures 17-19 and Table 6 here

- - - - - - - - - - - - - - - -

3.2.2. RRTs

As the coefficients in Fig. 20 (and Table 7) show, the zero lies outside the 95% credible intervals of their posterior distribution for the parameters QUESTION and DECLARATION as well as for the interaction AGE GROUP*QUESTION*CONTENT.

- - - - - - - - - - - - - - - -

Insert Figure 20 and Table 7 here

- - - - - - - - - - - - - - - -

Guided by the credible intervals for the above parameters, we more closely examine these effects. We begin with those that are consistent with the findings from Experiment 1. For the parameter QUESTION (Fig. 21), the mean Response RRTs was faster for positive questions than for negative questions ($\hat{\beta}$ =0.04, CrI = [0, 0.08], P($\hat{\beta}$) > 0 = 0.98). For the parameter DECLARATION (Fig. 22), the mean RRTs was faster for polarized declarations than for neutral declarations ($\hat{\beta}$ =0.05, CrI = [0.01, 0.08], P($\hat{\beta}$) > 0 = 0.99). Unlike in Experiment 1, there is no reportable difference based on the parameter AGE GROUP (which compares 4-year-olds to 6-year-olds here, $\hat{\beta}$ =-0.04, CrI = [-0.1, 0.03], P($\hat{\beta}$) > 0 = 0.13) nor for the interaction QUESTION*CONTENT ($\hat{\beta}$ =-0.01, CrI = [-0.05, 0.03], P($\hat{\beta}$) > 0 = 0.27).

- - - - - - - - - - - - - - - -

Insert Figures 21-22 here

- - - - - - - - - - - - - - - -

Here we turn to a particularly revealing effect that is unique to the 4-year-old to 6-year-old comparison: The AGE GROUP*QUESTION*CONTENT interaction ($\hat{\beta}$ =0.05, CrI = [0.01, 0.09], P($\hat{\beta}$) > 0 = 0.99). As shown in Fig. 23a, RRT's are fastest when the 4-year-olds are answering a negative question about a full box. This refers to the response *Si*. Whereas the 4-year-olds' and 6-year-olds' RRT's appear comparable for the cases where they answer (a) a positive question about a full box (*Oui*), (b) a positive question about an empty box (*Non*), and (c) a negative question about an empty box (*Non*), the 4-year-olds are noticeably *faster* than the 6-year-olds when answering a negative question about a full box, i.e. in answering *Si*. Fig. 23b presents the data in terms of the four kinds of responses across conditions; one can see that the *Si* response is particularly fast among the 4-year-olds when compared to the 6-year-olds. For the 6-year-olds *Si* takes longest to produce of the four conditions (1014 ms) while for the 4-year-olds *Si* is the fastest of the four (780 ms).

- - - - - - - - - - - - - - - -

Insert Figures 23a-23b here

- - - - - - - - - - - - - - - -

3.2.3. Secondary analysis of Experiment 2: Accuracy

As in Experiment 1, the Bayesian model calculates a posterior distribution for each of the model parameters, which now presents the DECLARATION condition as neutral vs. positive and neutral vs. negative (see Table 8). The high probability remained for the parameter QUESTION, the mean accuracy was higher for positive questions than for negative questions ($\hat{\beta}$ = -0.82, CrI = [-1.49, -0.16], P($\hat{\beta}$) > 0 = 0.01; see Fig. 24), for the parameter content ($\hat{\beta}$ =0.59, CrI = [-0.14, -1.32], P($\hat{\beta}$) > 0 = 0.95) and, like for Experiment 1's Secondary analysis, the model reveals an interaction that was not evident in our primary analysis, i.e. between polarity particles (QUESTION*CONTENT) and neutral-negative DECLARATION ($\hat{\beta}$ =1.10, CrI = [-0.08, 2.27], P($\hat{\beta}$) > 0 = 0.97; see Fig. 25). Like in Experiment 1's Secondary Analysis, the NS condition (responding *Si*) prompted the lowest rate of accurate responses (84%) when it followed a neutral declaration (distinguishing it from rates of accuracy for the negative declaration condition). This interaction indicates that the *Si* response under neutral declaration conditions is relatively challenging. Accuracy was higher for all other conditions. The remaining parameters did not yield results indicating extreme probabilities.

- - - - - - - - - - - - - - - -

Insert Figures 24-25 and Table 8 here

- - - - - - - - - - - - - - - - -

3.2.4 Secondary analysis of Experiment 2: RRT's

The model with three declarations largely confirms the findings in the primary analyses (see Table 9). As in the primary analysis, it reveals that the zero lies outside the 95% mark for the parameter QUESTION ($\hat{\beta}$ =0.04, CrI = [0.01, 0.07], P($\hat{\beta}$) > 0 = 0.99). It reveals that the zero lies outside the 95% mark for the declaration Neutral vs. Negative ($\hat{\beta}$ =0.13, CrI = [0.05, 0.21], P($\hat{\beta}$) > 0 = 1), while it is on the edge for the declaration Neutral vs. Positive ($\hat{\beta}$ = -0.07, CrI = [-0.15, 0.01], P($\hat{\beta}$) > 0 = 0.05). See Figures 26 and 27. The distinction between the Neutral and Positive declarations was clearer in Experiment 1. Finally, zero lies outside the 95% mark for the interaction AGE GROUP*QUESTION*CONTENT ($\hat{\beta}$ =0.05, CrI = [0.01, 0.09], P($\hat{\beta}$) > 0 = 0.99), as it did in the primary analysis.

- - - - - - - - - - - - - - - - -

Insert Figures 26-27 and Table 9 here

- - - - - - - - - - - - - - - - -

3.2.5. In the wake of the Negative Question: Comparing the RRT's of Si and Non

As we did for the analyses of Experiment 1, where we compared the RRT preceding *Si* articulation to that of *Non* in the NN condition (the Quaternary analysis after Experiment 1), it becomes relevant to do the same here. We thus perform the same t-test as we did above — in a Bayesian framework — for the four-year-olds. In this case, we find that there is a 99% probability that the *Si* response in the NS condition has lower (hence faster) RRTs than the *Non* response in the NN condition (mean of the posterior for the difference between the RRT's for *Si* versus *Non* [and the 95% credible intervals of the posterior]: -151.4 [-281.6, -22.6]). In line with the findings that show that the RRT of the *Si* response is particularly *fast* for the four-year-olds, it is also faster than the *Non* response in the NN condition.

3.3 Discussion of Experiment 2

This study with 4-year-olds provides further insight into the mastering of the response *Si*. We begin by pointing out the commonalities across the two groups of children. With respect to accuracy, the data from the 4-year-olds are largely comparable with those from the 6-year-olds (and adults) for whom positive questions prompt greater accuracy than negative ones. Likewise, when the two age groups in Experiment 2 are combined, the neutral declaration condition prompts the lowest accuracy

rates with respect to *Si*. With respect to RRT's, effects are largely similar across the two experiments. Mean RRTs were faster after positive questions than after negative questions and mean RRTs were faster after a polarized declaration as opposed to a neutral one (at 100% probability). To a very great extent, the 4-year-olds are comparable to the 6-year-olds.

There is one important exception however. Fascinatingly, the 4-year-olds respond remarkably fast in the condition calling for *Si* when compared to the 6-year-olds. While the 4-year-olds appear to provide the *Si* response with regularity and with relatively low error rates, they do so very quickly compared to the 6-year-olds. As we reported above, NS conditions provided the fastest response of the four conditions among the 4-year-olds whereas it was the source of the slowest responses for the 6-year-olds (and for the adults). By investigating this lower age group, we have come closer to understanding what distinguishes the less sophisticated answerers from their older cohorts when they use *Si*. As we discuss in greater detail in the General Discussion, we take these results to indicate that the 4-year-olds have a semantic appreciation of *Si* but they do not engage with it pragmatically in the way the six-year-olds do.

4. General Discussion

This study has focused on a contrapositive answer to negative questions, viz. *Si* in French. We proposed that this is a pragmatically rich response that is ideal for investigating what Grice would have considered a conventional implicature or what Blakemore would refer to as a word having a procedural meaning. The response *Si* allows the answerer to disentangle a negative question by indicating that its negative content is inconsistent with reality while showing accord with the negative question's implicit affirmative.

In order to properly investigate this contrapositive response, we invented a paradigm whose affirmative questions ultimately allow for natural answers such as *Oui* or *Non* and whose negative questions allow for *Si* or *Non*, based on uncovered evidence. The paradigm allowed us to examine two features that address the speaker-questioner's (i.e. the puppet's) epistemic state. The first concerns the speaker's (the puppet's) epistemic state at the outset of each trial, which could be polarized (where the puppet declares certainty about the candy's presence or absence in the to-be-examined box) or neutral (where the puppet declares not knowing where the candy is). In both Experiments 1 and 2, RRT's were faster in the wake of the puppet's polarized declarations compared to neutral ones overall. This is a first piece of evidence indicating that participants are sensitive to their interlocutor having a prior epistemic state (compared to a condition which starts off with declared ignorance).

As we indicated when we registered the study, we assumed that the positivity versus negativity of the initial claim was irrelevant. We anticipated that a declaration about (the presence or absence of

the candy in) the presented box (the one in the puppet's ultimate question) would anticipate a focus on that box in the task's test question and our data support that prediction. In contrast, the neutral declaration does not provide any cues for the upcoming question. Our fine-grained (secondary) analyses confirmed that RRT's in the wake of a neutral declaration were slower than those that followed each of the two polarized ones. We introduced epistemic states for the puppet to provide the background for the investigation of the second feature, the response *Si*, which is – more obviously -- our object of interest. We argue that mature uses of *Si* rely on pragmatic processes to fine-tune, i.e. to modify, the questioner's epistemic state so that it better aligns with the answerer's.

We reported two novel findings by analyzing the RRT's that precede *Si* in the context of all four conditions. The first is that among the most experienced speakers -- the adults and 6-year-olds -- the latency prior to the *Si* response for NS items was longest when compared to the responses in the other three (AO, AN, and NN) cases. A follow-up statistical test further showed that we can claim, with high confidence, that *Si* responses in the NS condition are slower than the *Non* responses in the NN conditions. To appreciate the extent to which these results are intriguing, compare these findings with those in classic categorization tasks in which participants respond true or false to statements such as *A star is above a plus* in light of a picture showing, say, a star below a plus sign (e.g. see Clark and Chase, 1972; also see Just & Carpenter, 1971) or that present questions such as *Is a robin a bird?* based on world knowledge (Akiyama et al., 1979). The four conditions in our paradigm (AO, AN, NS, and NN) are highly similar to the four used in these classic tasks, which are typically labelled, respectively, as True-Affirmative (TA), as in *Is a robin a bird?*, False-Affirmative (FA), as in *Is a robin a fish*?, False-Negative (FN), as in *Isn't a robin a bird?*, and True-Negative (TN), as in *Isn't a robin a fish?*. When provided statements (e.g. *A robin is a bird*) and the options True versus False, these classic tasks typically show how reaction times increase progressively as one moves from TA to FA to FN and to TN; the idea being that an increasing number of negations (in the answer or the question) leads to longer reaction times. (Note how the correct response *False* to a False-Negative assertion such as *A robin is not a bird* does not provide as rich an answer as a contrapositive.) That FN cases yield faster response times than TN cases has also been shown when the stimuli were questions and the response buttons were *Yes* or *No* and where *Yes* was the anticipated response.[11] In our

[11] In related work, Akiyama (1979, 1984, Akiyama et al., 1979; Akiyama & Guillory, 1983) aimed to make a distinction between *answering systems* (responding *Yes* or *No* to a question) and *verification systems* (answering *Right* or *Wrong* to a statement). For example, Akiyama and Guillory (1983) reported that English-speaking children as young as four had no difficulty answering negative questions, such as *Don't you have a mouth?* (by answering *Yes*) but that children as old as five had difficulty responding to equivalent items presented as statements, such as *You don't have a mouth* (while offering the response options *Right/Wrong*). This indicates that young English speakers can use the Polarity option *Yes* without difficulty to answer False Negative questions, but that verification, which relies more on properly processing the sentence's negative content (in order to say *Wrong* to a statement, one needs to note how the sentence itself is false), does raise some difficulties.

experiments here, the NS condition, which is analogous to the consistently second slowest (FN) condition in classic categorization tasks (and which is typically faster than the TN case in English), produces the longest natural RRT's among the most mature French responders (the adults and 6-year-olds) here. That there is a statistical difference with respect to RRT's between answering *Si* in the NS condition and answering *Non* in the NN condition among six-year-olds and adults was confirmed in Experiment 1's quaternary analysis.

The second novel finding is that the 4-year-olds do not exhibit prolonged latencies prior to responding *Si*. In fact, their *Si* responses are the fastest for their age group. Experiment 2 provides the developmental literature with a unique finding in which younger children are faster than their older cohorts (6-year-olds) in the same condition on the same task. While this finding appears exceptional, we now show that it actually makes sense in the context of other developmental pragmatic patterns in the literature.

Multiple studies show that younger children tend to interpret utterances with semantic readings before making pragmatic strides with age. Typically, this is described through truth-evaluations that evolve from semantic to pragmatic as children grow older. For example, three classic studies on "or" (Braine and Rumain, 1981; Paris, 1973; Sternberg, 1979) show how younger children tend to interpret the disjunction inclusively (which corresponds to its semantic reading) in truth-table-like tasks before becoming inclined -- with age -- to draw out exclusive readings, which reflect pragmatic inference. As described in the Introduction, this developmental pattern has been applied more generally to account for scalar implicatures. While studies in that literature (Noveck, 2001a; Papafragou & Musolino, 2003; Pouscoulous et al., 2007) have long assumed that adult-like responses provide a *signal* of pragmatic maturity, a more fully Gricean account ought to rely on evidence that shows how pragmatic enrichment serves to address a speaker's epistemic state (for work in this direction, see Breheny et al., 2013).

As the Introduction indicated, experimental pragmatic investigations have also revealed that readings of sentences that evidently include a scalar inference are typically more time consuming than those that do not include them among adults (see Bott and Noveck, 2004; Breheny et al., 2006; Huang & Snedeker, 2018). It can be seen that the current study is exploiting similar pragmatic effects (concerning development and time course) to account for additional processing here. The current work is in line with such claims and it is the first to isolate slowing latencies (during a critical brief window) of development. We propose that the four-year-olds' *Si* in the current study relies on a minimal semantic representation, one indicating merely rejection, without relying on pragmatic processes and without necessarily accessing the negative question's implicit affirmative. This explains their fast RRT's. In contrast, we propose that more sophisticated users of *Si* (among the six-year-olds and

adults) do rely on accompanying pragmatic processes to address the negative questioner's implicit affirmative and to express agreement with it. According to our account, these are intention-reading pragmatic steps (which, based on the RRT's, appear obligatory for the older participants) that slow the answering procedure down.

This account explains why a less mature response in the NS condition would not be expected to be something other than *Si*, e.g., *Oui,* and why Choi's (1991) corpus analysis finds *Si* responses emerging at ages as young as two-and-three-quarters years of age. Learning that *Si* is the appropriate answer — when disagreeing with the negative content of the question — is arguably more fundamental than engaging with the *pragmatic* steps associated with *Si,* which, again, involve accessing the negative questioner's affirmative and showing commitment by agreeing with it.

A subtle alternative to our proposal would be that *Si* answers evolve, with age, to simply show agreement with the negative question's implicit affirmative. This alternative is drawn from Krifka (2013) who assumes that the contrapositive refers unbiasedly but singularly to the positive antecedent of a negative question. We do not think this alternative is correct because adults in English (a language using a Polarity-based system without a contrapositive) answer *Yes* to what Akiyama et al. (1979) referred to as False Negative cases (which are comparable to our NS items) and at speeds that are typically faster than those for their True Negative cases (which are comparable to our NN items). Of course, we report just the opposite in a language using a Polarity-based system with a contrapositive. As should be clear, we assume that the relatively slow speed of the RRT's prior to the *Si* response in Experiment 1 indicate that French speakers address two features of the NS trial. Future research could help further examine our proposal which assumes that *Si*'s two aspects -- one s*emantic* which expresses disharmony (between the negative content in the question and reality) and one *pragmatic* which expresses accord with the question's implicit affirmative -- co-occur among adults.

Here we turn to an alternative account of our data that arose through the Action Editor's reconceptualization of our analyses, which goes as follows. Arguably, the exceptional appearance of the *Si* response -- as depicted by the crossover interaction in Figure 11a -- could be viewed as the summed costs of a) receiving a negative question and b) rejecting that question's negative content. In other words, one could recode the data so that they characterize participants' answers with respect to the question's polarity and the response's relative polarity (as either rejecting or accepting the content in the question), instead of labelling responses with respect to the paradigm's independent variables (question polarity and box content) as we did. Our reaction to this proposed alternative is twofold. One is that a recoding would not change Experiment 1's central finding, which is that a *Si* response is associated with the longest RRT's of the four conditions. As we indicated above, this is in itself an original finding because classic studies from the cognitive literature would lead one to predict that

RRT's in our NN condition to be slowest. The work from Akiyama et al. (1979) is, again, particularly illustrative because it reports how responses to their False Negative items, whether they are expressed as assertions that call for *False* responses or as questions that call for *Yes* answers, are typically faster than responses to the True Negative items. This leads to our second reaction, which is that our analyses allow one to appreciate how a conventionalized answer in European French, *Si*, essentially combines the two features investigated independently in Akiyama et al.'s study -- rejecting the negative content in the question (which drives *False* responses to assertions) and showing accord with the questioner's implicit affirmative (which justifies *Yes* answers to the question form). The reconceptualization compels analyses to focus on just the former, i.e., the answer particle's relative polarity (whether the response accepts or rejects), which amounts to making theoretical commitments about *Si*'s meaning.[12]

We now turn to one last question: Could the relatively slow *Si* responses among the six-year-olds and adults be attributable to something other than the pragmatic steps that we argue are intrinsic to it? To postulate an alternative possibility, as suggested by a reviewer, perhaps *Si* is simply less accessible than European French's other two response particles? Our answer to this query is threefold. First, in an effort to consider the above alternative, we explored the possibility that perhaps the results among 6-year-olds and adults concerning *Si* are due to its being a relatively rare response. We thus searched the Lexique2 French lexical database (New et al, 2004; New et al, 2007) to determine the frequency of the response particles *Oui, Non* and *Si*. It turns out that *Si* is highly frequent in French discourse thus lending doubt to this possibility. To be more specific, we sought the frequencies of response particles from a movie subtitles corpus, which was chosen because a) this corpus comes closest to reflecting oral rather than written speech and; b) this database distinguishes between the two main uses of *Si* in French (the other being equivalent to the conditional, much like *If* in English). In this corpus, all three response particles are among the top 0,5% of the most frequent adverbs (the actual frequencies -- in occurrences per million -- for *Oui, Non* and *Si* are 3207, 4040 and 2107, respectively). Across all grammatical categories, these three words were among the 0,1% most frequent referenced lemmas. So, while *Si* is the least frequent of the three French response articles it is nonetheless a very highly-used expression. We consider it doubtful that *Si* is considerably less accessible than the other two. Second, the 4-year-olds' exceptionally fast *Si* puts paid to the idea that pronouncing *Si*, in itself, slows down responses universally. Assuming that four-year-olds' responses

[12] Note, too, that our analyses are maximally felicitous, i.e., the suggested recoding would not be generally applicable to the NS condition across languages while our analyses would. For example, if the task were carried out in other languages -- such as English or Japanese -- that have no equivalent to *Si* the recoding would not translate. An English *Yes* in our NS condition would show accord with the negative question's implicit affirmative without indicating rejection (or *reversing* to use Farkas and Bruce's terminology) and a Japanese *iie* would indicate rejection without referring to the negative question's implicit affirmative.

reflect a minimally semantic representation, as we argue above, one can see that this is sufficient for accessing and producing the answer *Si*. Finally, as our analyses have indicated (see our tertiary analyses of Experiment 1), it appears that the mature responses of *Si* – like the other response particles – are affected by the epistemic status of the puppet. This indicates that *Si* is being produced as fast as conditions permit.

To sum up, the paper's two main findings – that the adult (or adult-like) employment of *Si* in our task is more effort demanding than its controls and that early competent uses of *Si* can be used without its pragmatic component – resonate with those classically reported with other pragmatic effects, including scalar implicatures. Experimentally speaking, we have shown how *Si*'s pragmatic contribution can be isolated. Theoretically speaking, these data force one to consider — contra Bach (1999), for example — that adult uses of words expressing procedural meaning, as in the case of *Si,* intrinsically depend on pragmatic components.

Acknowledgements

This work was presented for the first time at *XPrag.it* in Pavia in 2018. The authors wish to express their gratitude to Naomi Yamaguchi for being the voice of the puppet (recorded at the LPP in Paris), to Kathleen Birot, Camille Procuna, and Laure Decelle for their roles as experimenters and coders (in Lyon), to Yair Haendler (under the auspices of the LabEx EFL in Paris) for helping us develop the statistical analyses, to Deirdre Wilson and Dan Sperber for feedback on specific issues, and to Adrien Staub and three anonymous reviewers whose detailed comments on prior manuscripts greatly improved the paper.

Author notes

Our open source materials, including predictions and data related to this work, can be found at https://osf.io/6vjx8/. Further information about the experiment or our analyses can be obtained by contacting I.N., N.P., or G.T.

References

Akiyama, M. M. (1979). Yes-no answering systems in young children. *Cognitive Psychology*, *11*(4), 485-504.

Akiyama, M. M. (1984). Are language-acquisition strategies universal? *Developmental Psychology*, *20*(2), 219.

Akiyama, M. M., Brewer, W. F., & Shoben, E. J. (1979). The yes-no question answering system and statement verification. *Journal of Memory and Language*, *18*(3), 365.

Akiyama, M. M., & Guillory, A. W. (1983). The ontogeny of the verification system. *Journal of Verbal Learning and Verbal Behavior*, *22*(3), 333-340.

Asatiani, R. (1999). The semantics and typology of yes/no particles: A cross linguistic study. Open Society Institute.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of memory and language*, *59*(4), 390-412.

Bach, K. (1999). The myth of conventional implicature. *Linguistics and philosophy*, 327-366.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of memory and language*, *68*(3), 255-278.

Blakemore, D. (2002). *Relevance and Linguistic Meaning: The Semantics and Pragmatics of Discourse Markers*. CUP: New York.

Blakemore, D. (1987). *Semantic constraints on relevance*. Oxford: Blackwell.

Boersma, P., Weenink, D. (2018) *Praat: doing phonetics by computer* [Computer program], Version 6.0.37, retrieved 3 February 2018 from http://www.praat.org/.

Bosco, F. M., & Gabbatore, I. (2017). Sincere, deceitful, and ironic communicative acts and the role of the theory of mind in childhood. *Frontiers in psychology*, *8*, 21.

Bott, L., & Noveck, I. A. (2004). Some utterances are underinformative: The onset and time course of scalar inferences. *Journal of memory and language*, *51*(3), 437-457.

Braine, M. & Rumain, B. (1981). Children's comprehension of "or": Evidence for a sequence of competencies. *Journal of Experimental Child Psychology, 31*, 46-70.

Breheny, R., Katsos, N., & Williams, J. (2006). Are generalised scalar implicatures generated by default? An on-line investigation into the role of context in generating pragmatic inferences. Cognition, 100(3), 434-463.

Breheny, R., Ferguson, H. J., & Katsos, N. (2013). Taking the epistemic step: Toward a model of on-line access to conversational implicatures. *Cognition*, *126*(3), 423-440.

Bürkner, P-C. (2018). Advanced Bayesian Multilevel Modeling with the R Package brms. The R Journal, 10(1), 395-411. doi:10.32614/RJ-2018-017

Choi, S. (1991). Children's answers to yes-no questions: A developmental study in English, French, and Korean. *Developmental psychology*, *27*(3), 407.

Clark, H. H., & Chase, W. G. (1972). On the process of comparing sentences against pictures. *Cognitive psychology*, *3*(3), 472-517.

Claus, B., Meijer, A. M., Repp, S., & Krifka, M. (2017). Puzzling response particles: An experimental study on the German answering system. *Semantics and Pragmatics*, *10*.

Delais-Roussarie, E., Post, B., Avanzi, M., Buthke, C., Di Cristo, A., Feldhausen, I., ... & Sichel-Bazin, R. (2015). Intonational phonology of French: Developing a ToBI system for French. In: Frota, S. and Prieto, P. eds. *Intonational variation in Romance*. Oxford: OUP.

Dimroth, C. Schimke S., Turco, G. (2018) Focusing Functional Elements: Affirmative Particles and Verum Focus in First Language Acquisition of German. *Language Acquisition*, 25(3), 268-283.

Farkas, D.F. & Bruce, K. (2010). On reacting to assertions and polar questions. *Journal of Semantics, 27*, 81–118.

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2014). Bayesian data analysis (Third, Chap. 11) Boca Raton, FL: Taylor & Francis.

Glenwright, M., & Pexman, P. M. (2010). Development of children's ability to distinguish sarcasm and verbal irony. *Journal of Child Language*, *37*(2), 429-451.
Grice, H. P. (1989). *Studies in the Way of Words*. Harvard University Press.
Huang, Y. T., & Snedeker, J. (2018). Some inferences still take time: Prosody, predictability, and the speed of scalar implicatures. *Cognitive psychology*, *102*, 105-126.
Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of memory and language*, *59*(4), 434-446.
Jones, B.M. (1999). The Welsh answering system. Mouton de Gruyter: Berlin/New York.
Just, M. A., & Carpenter, P. A. (1971). Comprehension of negation with quantification. *Journal of Verbal Learning and Verbal Behavior*, *10*(3), 244-253.
Katsos, N., Cummins, C., Ezeizabarrena, M. J., Gavarró, A., Kraljević, J. K., Hrzica, G., ... & Noveck, I. A. (2016). Cross-linguistic patterns in the acquisition of quantifiers. *Proceedings of the National Academy of Sciences*, *113*(33), 9244-9249.
Kerbrat-Orecchioni, C. (1991). *La question*. Presses universitaires de Lyon.
Krifka, M. (2013). Response particles as propositional anaphors. In *Semantics and linguistic theory* (Vol. 23, pp. 1-18).
Kruschke, J. K. & Meredith, M. (2020). *BEST: Bayesian Estimation Supersedes the t-Test*. R package version 0.5.2. https://CRAN.R-project.org/package=BEST
Leslie, A. M., Friedman, O., & German, T. P. (2004). Core mechanisms in 'theory of mind'. *Trends in Cognitive Sciences*, *8*(12), 528-533.
Li, F., González-Fuente, S., Prieto, P., & Espinal, M. T. (2016). Is Mandarin Chinese a truth-based language? Rejecting responses to negative assertions and questions. *Frontiers in Psychology*, *7*, 1967.
Lunn, D., Jackson, C., Best, N., Spiegelhalter, D., & Thomas, A. (2012). *The BUGS book: A practical introduction to Bayesian analysis*. Chapman and Hall/CRC.
Maratsos, M. P. (1974). Preschool children's use of definite and indefinite articles. *Child Development*, 446-455.
New, B., Pallier, C., Brysbaert, M., & Ferrand, L. (2004). Lexique 2: A new French lexical database. *Behavior Research Methods, Instruments, & Computers, 36(3)*, 516-524.
New, B., Brysbaert, M., Veronis, J., & Pallier, C. (2007). The use of film subtitles to estimate word frequencies. *Applied Psycholinguistics, 28(4),* 661-677.
Nicenboim, B. & Vasishth, S. (2016). Statistial methods for linguistic research: Foundational ideas – Part II. *Language and Linguistics Compass 10*, 591-613.
Nicenboim, B., Vasishth, S., Engelmann, F. and Suckow, K. (2018), Exploratory and Confirmatory Analyses in Sentence Processing: A Case Study of Number Interference in German. Cogn Sci, 42: 1075-1100. doi:10.1111/cogs.12589
Noveck, I. A. (2001a). Linguistlist Query on French *"Si"* from *https://linguistlist.org/issues/12/12-259.html*, ed. Lydia Grebenyova.
Noveck, I. A. (2001b). When children are more logical than adults: Experimental investigations of scalar implicature. *Cognition*, *78*(2), 165-188.
Noveck, I. (2018). *Experimental Pragmatics: The Making of a Cognitive Science*. Cambridge University Press.
Noveck, I. A., Bianco, M., & Castry, A. (2001). The costs and benefits of metaphor. *Metaphor and Symbol*, *16*(1-2), 109-121.
Papafragou, A., & Musolino, J. (2003). Scalar implicatures: experiments at the semantics–pragmatics interface. *Cognition*, *86*(3), 253-282.
Paris, S. (1973). Comprehension of language connectives and propositional logical relationships. *Journal of Experimental Child Psychology, 16*, 278-291.
Pope, E. (1973). Question-answering systems. In *Proceedings of Chicago Linguistics Society* (pp. 482-92).

Pouscoulous, N., Noveck, I. A., Politzer, G., & Bastide, A. (2007). A developmental investigation of processing costs in implicature production. *Language acquisition*, *14*(4), 347-375.
Power, R. J., & Dal Martello, M. F. (1986). The use of the definite and indefinite articles by Italian preschool children. *Journal of Child Language*, *13*(1), 145-154.
R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/
Schmerse, D., Lieven, E., & Tomasello, M. (2013). Discourse particles and belief reasoning: The case of German *doch*. *Journal of Semantics*, *31*(1), 115-133.
Sperber, D., & Wilson, D. (1986). *Relevance: Communication and cognition* (Vol. 142). Cambridge, MA: Harvard University Press.
Sternberg, R. J. (1979). Developmental patterns in the encoding and combination of logical connectives. *Journal of Experimental Child Psychology, 28,* 469-498.
Stiller, A. J., Goodman, N. D., & Frank, M. C. (2015). Ad-hoc implicature in preschool children. *Language Learning and Development*, *11*(2), 176-190.
Sorensen, T., Hohenstein, S., & Vasishth, S. (2016). Bayesian linear mixed models using stan: A tutorial for psychologists, linguists, and cognitive scientists. Quantitative Methods for Psychology, 12(3):175-200 DOI 10.20982/tqmp.12.3.p175
Tian, Y., & Breheny, R. (2016). Dynamic pragmatic view of negation processing. In *Negation and polarity: Experimental perspectives* (pp. 21-43). Springer, Cham.
Tian, Y., Breheny, R., & Ferguson, H. (2010). Why we simulate negated information: a dynamic pragmatic account. *The Quarterly Journal of Experimental Psychology,* 63, 2305-2312.
Tomlinson Jr, J. M., Bailey, T. M., & Bott, L. (2013). Possibly all of that and then some: Scalar implicatures are understood in two steps. *Journal of memory and language*, *69*(1), 18-35.
Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of p-values. Psychonomic Bulletin & Review, 14, 779–804.
Wilson, D., & Sperber, D. (1993). Linguistic form and relevance. *Lingua*, *90*(1), 1-25.

Tables

Table 1. One example item per condition. Examples 1-4 are instances of each of the four experimental conditions (translated into English). Examples 5 and 6 are two kinds of fillers. The examples also illustrate each of the three kinds of puppet declaration. See text for further information.

Example number	Condition Name	Boxes Presented	Puppet Declaration	Box examined/ Status	Puppet Question	Expected Answer
1.	AO	Brown/ Blue	I don't know where it is.	Blue/ Full	It is in the blue box?	Oui
2.	AN	Orange/ Blue	It is surely in the orange box.	Orange/ Empty	It is in the orange box?	Non
3.	NS	White/ Red	It is surely not in the white box.	White/ Full	It is not in the white box?	Si
4.	NN	Green/ Red	It is surely in the green box.	Green/ Empty	It is not in the green box?	Non
5.	$Filler_1$	White/ Orange	It is surely in the orange box.	White/ Full	It is not in the white box?	Si
6.	$Filler_2$	White/ Green	I don't know where it is.	White/ Empty	It is in the green box?	Oui

Table 2. Summary of Accuracy performance from Experiment 1 based on the posterior distribution of the fixed effects model parameters. For each parameter, the table illustrates the estimated mean of the posterior, the 95% credible intervals and the probability that the posterior is smaller than zero.

Effect	*Estimated mean* $(\hat{\beta})$	*95% credible intervals*	$P(\hat{\beta}) < 0$
Age group	0.53	[-0.17, 1.21]	0.07
Question	**-0.78**	**[-1.53, -0.06]**	**0.98**
Content	0.35	[-0.39, 1.10]	0.17
Declaration	-0.15	[-0.83, 0.57]	0.67
Age group: Question	0.47	[-0.15, 1.14]	0.07
Age group: Content	-0.45	[-1.12, 0.19]	0.92
Question: Content	-0.10	[-0.82, 0.63]	0.61
Age group: Declaration	-0.11	[-0.73, 0.50]	0.64
Question: Declaration	-0.42	[-1.13, 0.27]	0.88
Content: Declaration	0.14	[-0.56, 0.85]	0.34
Age group: Question: Content	0.47	[-0.13, 1.12]	0.06
Age group: Question: Declaration	-0.12	[-0.73, 0.48]	0.65
Age group: Content: Declaration	0.10	[-0.54, 0.72]	0.37
Question: Content: Declaration	0.48	[-0.21, 1.16]	0.08
Age group: Question: Content: Declaration	0.25	[-0.35, 0.87]	0.21

Table 3. Summary of Reaction Response times (RRTs) from Experiment 1 based on the posterior distribution of the fixed effects model parameters. For each parameter, the table illustrates the estimated mean of the posterior, the 95% credible intervals and the probability that the posterior is smaller than zero.

Effect	*Estimated mean* $(\hat{\beta})$	*95% credible intervals*	$P(\hat{\beta}) < 0$
Age group	**-0.08**	**[-0.12, -0.04]**	**1.00**
Question	**0.07**	**[0.04, 0.11]**	**0.00**
Content	0.00	[-0.03, 0.03]	0.53
Declaration	**0.06**	**[0.03, 0.09]**	**0.00**
Age group: Question	0.02	[-0.01, 0.05]	0.07
Age group: Content	0.01	[-0.01, 0.04]	0.12
Question: Content	**-0.04**	**[-0.08, -0.01]**	**1.00**
Age group: Declaration	-0.01	[-0.04, 0.01]	0.89
Question: Declaration	-0.01	[-0.04, 0.02]	0.79
Content: Declaration	0.01	[-0.02, 0.04]	0.16
Age group: Question: Content	0.01	[-0.01, 0.03]	0.17
Age group: Question: Declaration	0.01	[-0.01, 0.03]	0.11
Age group: Content: Declaration	0.01	[-0.01, 0.03]	0.24
Question: Content: Declaration	0.01	[-0.02, 0.04]	0.26
Age group: Question: Content: Declaration	0.00	[-0.02, 0.02]	0.43

Table 4. Secondary analysis: Summary of Accuracy from Experiment 1 based on the posterior distribution of the fixed effects model parameters. For each parameter, the table illustrates the estimated mean of the posterior, the 95% credible intervals and the probability that the posterior is smaller than zero.

Effect	*Estimated mean* $(\hat{\beta})$	*95% credible intervals*	$P(\hat{\beta}) < 0$
Age group	0.43	[-0.26 – 1.11]	0.11
Question	**-0.78**	**[-1.52 – 0.05]**	**0.98**
Content	0.21	[-0.54 – 0.94]	0.28
(Declaration) Neutral: Negative	-0.43	[-1.61 – 0.76]	0.77
(Decl.) Positive: Neutral	0.77	[-0.51 – 2.03]	0.11
Age group: Question	0.43	[-0.21 – 1.08]	0.09
Age group: Content	-0.33	[-1.02 – 0.32]	0.84
Question: Content	-0.20	[-0.92 – 0.53]	0.71
Age group: (Decl.) Neutral: Negative	0.12	[-0.95 – 1.20]	0.41
Age group: (Decl.) Positive: Neutral	0.28	[-0,86 – 1.43]	0.32
Question: (Decl.) Neutral: Negative	-0.94	[-2.12 – 0.27]	0.94
Question: (Decl.) Positive: Neutral	-0.06	[-1.33 – 1.16]	0.54
Content: (Decl.) Neutral: Negative	0.49	[-0.68 – 1.65]	0.20
Content: (Decl.) Positive: Neutral	-0.43	[-1.65 – 0.81]	0.76
Age group: Question: Content	0.44	[-0.18 – 1.08]	0.08
Age group: Question: (Decl.) Neutral: Negative	0.08	[-0.99 – 1.13]	0.44
Age group: Question: (Decl.) Positive: Neutral	0.24	[-0.92 – 1.41]	0.34
Age group: Content: (Decl.) Neutral: Negative	-0.03	[-1.13 – 1.04]	0.53
Age group: Content: (Decl.) Positive: Neutral	-0.08	[-1.25 – 1.08]	0.55
Question: Content: (Decl.) Neutral: Negative	**1.04**	**[-0.19 – 2.22]**	**0.04**
Question: Content: (Decl.) Positive: Neutral	-0.31	[-1.57 – 0.98]	0.69
Age group: Question: Content:(Decl.) Neutral: Neg.	-0.09	[-1.18 – 0.98]	0.56
Age group: Question: Content: (Decl.) Pos.: Neutral	-0.66	[-1.82 – 0.51]	0.86

Table 5. Secondary analysis: Summary of Reaction Response times (RRTs) from Experiment 1 based on the posterior distribution of the fixed effects model parameters. For each parameter, the table illustrates the estimated mean of the posterior, the 95% credible intervals and the probability that the posterior is smaller than zero.

Effect	*Estimated mean* $(\hat{\beta})$	*95% credible intervals*	$P(\hat{\beta}) < 0$
Age group	**-0.07**	**[-0.11 – -0.03]**	**1.00**
Question	**0.08**	**[0.04 – 0.11]**	**0.00**
Content	-0.01	[-0.04 – 0.03]	0.65
(Declaration) Neutral: Negative	**0.13**	**[0.05 – 0.2]**	**0.00**
(Decl.) Positive: Neutral	**-0.10**	**[-0.18 – -0.02]**	**0.99**
Age group: Question	0.02	[-0.01 – 0.04]	0.11
Age group: Content	0.01	[-0.01 – 0.03]	0.16
Question: Content	**-0.05**	**[-0.08 – -0.01]**	**0.99**
Age group: (Decl.) Neutral: Negative	-0.02	[-0.07 – 0.03]	0.79
Age group: (Decl.) Positive: Neutral	0.03	[-0.02 – 0.08]	0.10
Question: (Decl.) Neutral: Negative	-0.02	[-0.10 – 0.06]	0.72
Question: (Decl.) Positive: Neutral	0.02	[-0.06 – 0.10]	0.28
Content: (Decl.) Neutral: Negative	0.03	[-0.04 – 0.11]	0.19
Content: (Decl.) Positive: Neutral	-0.03	[-0.11 – 0.06]	0.74
Age group: Question: Content	0.01	[-0.01 – 0.03]	0.17
Age group: Question: (Decl.) Neutral: Negative	0.04	[-0.01 – 0.09]	0.07
Age group: Question: (Decl.) Positive: Neutral	-0.02	[-0.07 – 0.03]	0.75
Age group: Content: (Decl.) Neutral: Negative	0.00	[-0.05 – 0.05]	0.47
Age group: Content: (Decl.) Positive: Neutral	-0.03	[-0.08 – 0.03]	0.86
Question: Content: (Decl.) Neutral: Negative	0.02	[-0.06 – 0.09]	0.32
Question: Content: (Decl.) Positive: Neutral	-0.01	[-0.10 – 0.07]	0.65
Age group: Question: Content:(Decl.) Neutral: Neg.	-0.01	[-0.06 – 0.04]	0.61
Age group: Question: Content:(Decl.) Pos.: Neutral	-0.01	[-0.07 – 0.04]	0.71

Table 6. Summary of Accuracy from Experiment 2 based on the posterior distribution of the fixed effects model parameters. For each parameter, the table illustrates the estimated mean of the posterior, the 95% credible intervals and the probability that the posterior is smaller than zero.

Effect	*Estimated mean* $(\hat{\beta})$	*95% credible intervals*	$P(\hat{\beta}) < 0$
Age group	0.10	[-0.58, 0.80]	0.39
Question	**-0.74**	**[-1.44, -0.04]**	**0.98**
Content	**0.87**	**[0.12, 1.65]**	**0.01**
Declaration	0.18	[-0.51, 0.91]	0.30
Age group: Question	0.42	[-0.22, 1.09]	0.10
Age group: Content	0.22	[-0.48, 0.93]	0.28
Question: Content	-0.18	[-0.89, 0.53]	0.69
Age group: Declaration	0.24	[-0.40, 0.89]	0.24
Question: Declaration	-0.28	[-0.97, 0.41]	0.79
Content: Declaration	0.36	[-0.33, 1.08]	0.15
Age group: Question: Content	0.36	[-0.29, 1.03]	0.14
Age group: Question: Declaration	0.02	[-0.63, 0.67]	0.48
Age group: Content: Declaration	0.32	[-0.30, 0.98]	0.16
Question: Content: Declaration	0.37	[-0.32, 1.04]	0.14
Age group: Question: Content: Declaration	0.14	[-0.49, 0.79]	0.32

Table 7. Summary of Reaction Response times (RRTs) from Experiment 2 based on the posterior distribution of the fixed effects model parameters. For each parameter, the table illustrates the estimated mean of the posterior, the 95% credible intervals and the probability that the posterior is smaller than zero.

Effect	*Estimated mean* $(\hat{\beta})$	*95% credible intervals*	$P(\hat{\beta}) < 0$
Age group	-0.04	[-0.10, 0.03]	0.87
Question	**0.04**	**[0.00, 0.08]**	**0.02**
Content	0.02	[-0.02, 0.05]	0.18
Declaration	**0.05**	**[0.01, 0.08]**	**0.01**
Age Group: Question	-0.01	[-0.05, 0.02]	0.77
Age Group: Content	0.03	[-0.01, 0.07]	0.06
Question: Content	-0.01	[-0.05, 0.03]	0.73
Age group: Declaration	-0.02	[-0.06, 0.02]	0.89
Question: Declaration	-0.01	[-0.04, 0.03]	0.63
Content: Declaration	0.01	[-0.03, 0.04]	0.36
Age group: Question: Content	**0.05**	**[0.01, 0.09]**	**0.01**
Age group: Question: Declaration	0.02	[-0.02, 0.06]	0.16
Age group: Content: Declaration	0.00	[-0.04, 0.04]	0.51
Question: Content: Declaration	-0.01	[-0.05, 0.02]	0.79
Age group: Question: Content: Declaration	-0.02	[-0.06, 0.02]	0.87

Table 8. Secondary analysis: Summary of Accuracy from Experiment 2 based on the posterior distribution of the fixed effects model parameters. For each parameter, the table illustrates the estimated mean of the posterior, the 95% credible intervals and the probability that the posterior is smaller than zero.

Effect	*Estimated mean* $(\hat{\beta})$	*95% credible intervals*	$P(\hat{\beta}) < 0$
Age group	-0.01	[-0.69 – 0.66]	0.51
Question	**-0.82**	**[-1.49 – 0.16]**	**0.99**
Content	**0.59**	**[-0.14 – 1.32]**	**0.05**
(Declaration) Neutral: Negative	-0.33	[-1.49 – 0.83]	0.71
(Decl.) Positive: Neutral	0.39	[-0.82 – 1.63]	0.26
Age group: Question	0.31	[-0.31 – 0.93]	0.16
Age group: Content	0.16	[-0.53 – 0.87]	0.32
Question: Content	-0.28	[-0.96 – 0.39]	0.80
Age group: (Decl.) Neutral: Negative	0.24	[-0.89 – 1.34]	0.33
Age group: (Decl.) Positive: Neutral	-0.18	[-1.36 – 0.98]	0.61
Question: (Decl.) Neutral: Negative	-0.72	[-1.87 – 0.45]	0.89
Question: (Decl.) Positive: Neutral	-0.06	[-1.24 – 1.11]	0.53
Content: (Decl.) Neutral: Negative	0.35	[-0.81 – 1.48]	0.27
Content: (Decl.) Positive: Neutral	-0.95	[-2.16 – 0.25]	0.94
Age group: Question: Content	0.33	[-0.28 – 0.97]	0.15
Age group: Question: (Decl.) Neutral: Negative	0.29	[-0.81 – 1.42]	0.31
Age group: Question: (Decl.) Positive: Neutral	0.15	[-1.03 – 1.32]	0.40
Age group: Content: (Decl.) Neutral: Negative	-0.14	[-1.25 – 0.97]	0.60
Age group: Content: (Decl.) Positive: Neutral	-0.63	[-1.81 – 0.53]	0.86
Question: Content: (Decl.) Neutral: Negative	**1.10**	**[-0.08 – 2.27]**	**0.03**
Question: Content: (Decl.) Positive: Neutral	0.07	[-1.10 – 1.26]	0.45
Age group: Question: Content:(Decl.) Neutral: Neg.	0.01	[-1.09 – 1.10]	0.50
Age group: Question: Content: (Decl.) Pos.: Neutral	-0.21	[-1.34 – 0.93]	0.64

Table 9. Secondary analysis: Summary of Reaction Response times (RRTs) from Experiment 2 based on the posterior distribution of the fixed effects model parameters. For each parameter, the table illustrates the estimated mean of the posterior, the 95% credible intervals and the probability that the posterior is smaller than zero.

Effect	*Estimated mean* $(\hat{\beta})$	*95% credible intervals*	$P(\hat{\beta}) < 0$
Age group	-0.03	[-0.09 – 0.04]	0.81
Question	**0.04**	**[0.01 – 0.07]**	**0.01**
Content	0.02	[-0.02 – 0.05]	0.19
(Declaration) Neutral: Negative	**0.13**	**[0.05 – 0.21]**	**0.00**
(Decl.) Positive: Neutral	**-0.07**	**[-0.15 – 0.01]**	**0.95**
Age group: Question	-0.02	[-0.06 – 0.02]	0.84
Age group: Content	0.03	[-0.01 – 0.07]	0.07
Question: Content	-0.01	[-0.04 – 0.03]	0.66
Age group: (Decl.) Neutral: Negative	-0.02	[-0.12 – 0.07]	0.66
Age group: (Decl.) Positive: Neutral	0.07	[-0.03 – 0.17]	0.07
Question: (Decl.) Neutral: Negative	0.00	[-0.08 – 0.08]	0.49
Question: (Decl.) Positive: Neutral	0.02	[-0.06 – 0.10]	0.29
Content: (Decl.) Neutral: Negative	0.04	[-0.04 – 0.12]	0.13
Content: (Decl.) Positive: Neutral	0.02	[-0.06 – 0.10]	0.30
Age group: Question: Content	**0.05**	**[0.01 – 0.09]**	**0.01**
Age group: Question: (Decl.) Neutral: Negative	0.06	[-0.03 – 0.15]	0.10
Age group: Question: (Decl.) Positive: Neutral	-0.02	[-0.12 – 0.08]	0.64
Age group: Content: (Decl.) Neutral: Negative	0.01	[-0.08 – 0.11]	0.39
Age group: Content: (Decl.) Positive: Neutral	0.02	[-0.08 – 0.12]	0.36
Question: Content: (Decl.) Neutral: Negative	-0.02	[-0.09 – 0.06]	0.67
Question: Content: (Decl.) Positive: Neutral	0.04	[-0.04 – 0.12]	0.15
Age group: Question: Content:(Decl.) Neutral: Neg	-0.04	[-0.13 – 0.05]	0.81
Age group: Question: Content: (Decl.) Pos: Neut'l	0.04	[-0.06 – 0.14]	0.19

Figures

Fig. 1. A graphic describing the origins of the 24 Experimental items (that lead to the questions in the blue section) following the three Declaration conditions; these are cases in which a) a (Negative or Positive) declaration was made about a box or else the puppet said "I don't know where it is", b) a box emerged to be inspected before c) receiving a query about its contents. Also described are the origins of the filler items, which come in two varieties: 1) Four filler items refer to cases where the declaration concerns one box in the first phase of the trial (in orange) and yet the other box emerges to be inspected. 2) Twelve filler items refer to cases where the declaration is about the box that emerges (like in the experimental trials); however, the question concerns the box that remains on the screen.

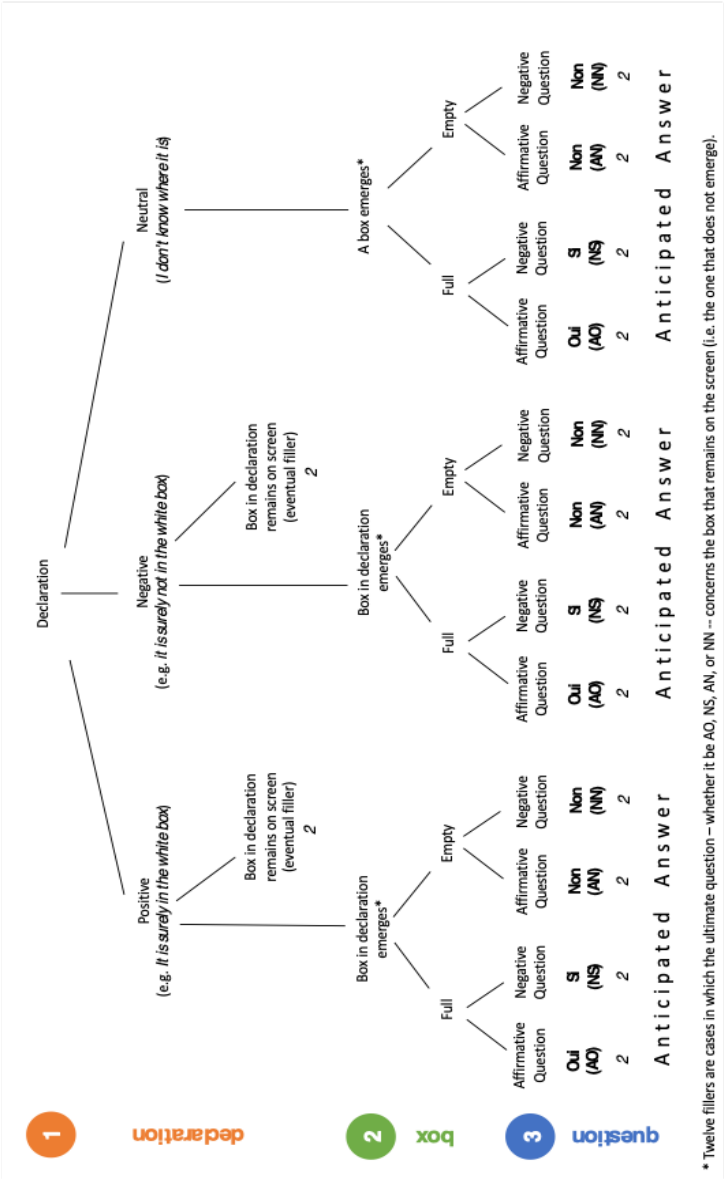
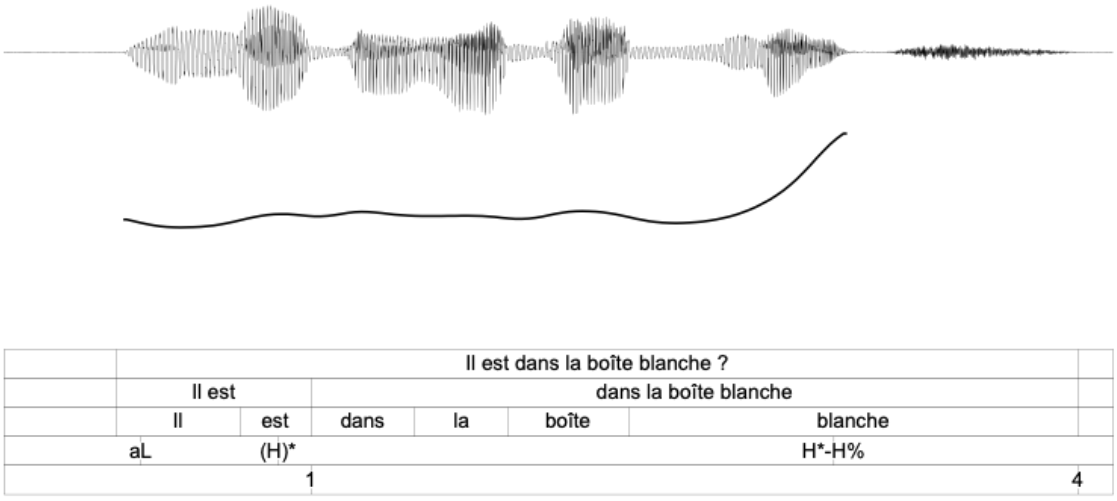


Fig. 2(a-b). Two examples of French polarity questions (panel a: positive question; panel b: negative question). The first three tiers contain the orthographic transcription of the recorded question segmented at the utterance and at a phrase/ word level. The last two tiers contain information about the tone and the phrasing of the question using a *ToBI*-style annotation (Delais et al. 2015).

(a)



(b)

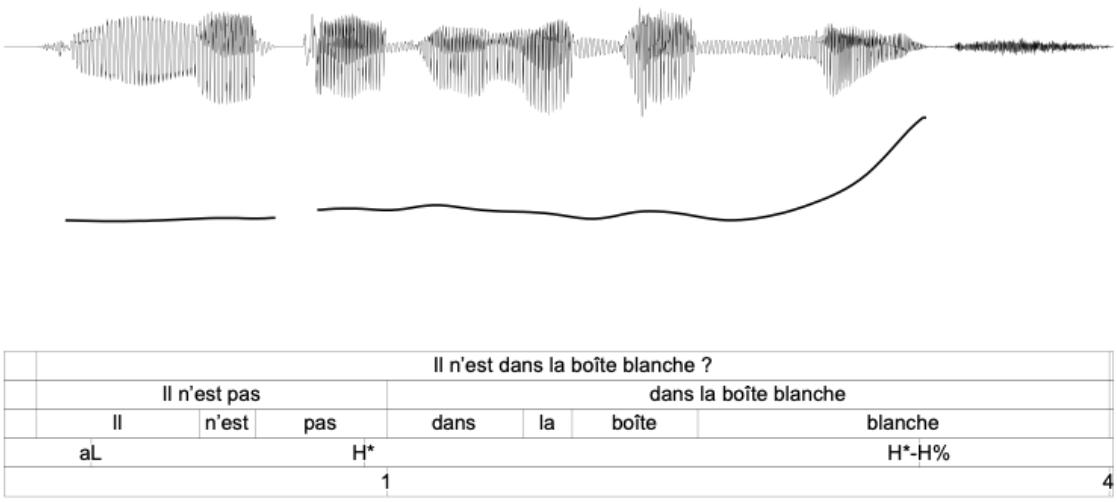


Fig. 3. An illustration of two trials with the same declaration, which begins (as translated into English) with "It is surely in the white box" (Panel 1) before that box slides out (Panels 2-3). Critically, the participant receives and inspects the contents of the box and re-lids it (Panels 4-5) before the test question is posed and answered (see Panels 6-7). In the first example, the participant is ultimately asked "It is in the white box?" which is expected to elicit the response *Oui*. In the second example (comprising Panels 1-5 and 6'-7'), the participant is asked "It is not in the white box?" to which the participant is expected to produce the response *Si*.

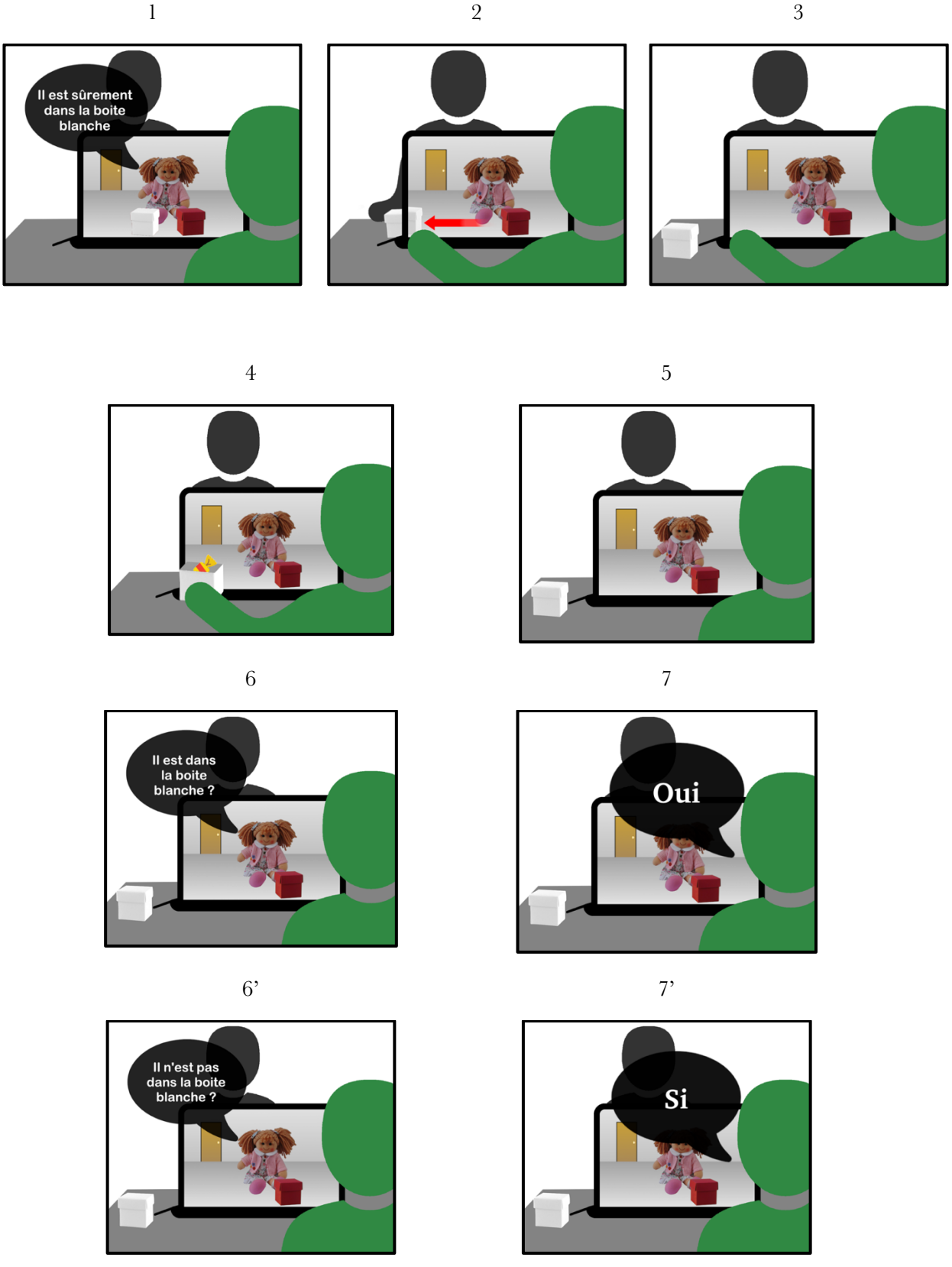


Fig. 4. A visual representation of the way a Response Reaction Time (an RRT) was measured, with an example from the Negative Si (NS) condition, i.e. after the negative question *Il n'est pas dans la boite blanche?* – " It is not in the white box?". Here the starting point, the left boundary of the RRT, begins after the first phoneme. Similarly, the right boundary is the very beginning of the consonant /s/ of the answer *Si*. For illustrative purposes, the visual representation of the starting point of the signal here is simplified in that it coincides with the end of the first phoneme; however, the determination of the starting point relied on both visual and auditory input.

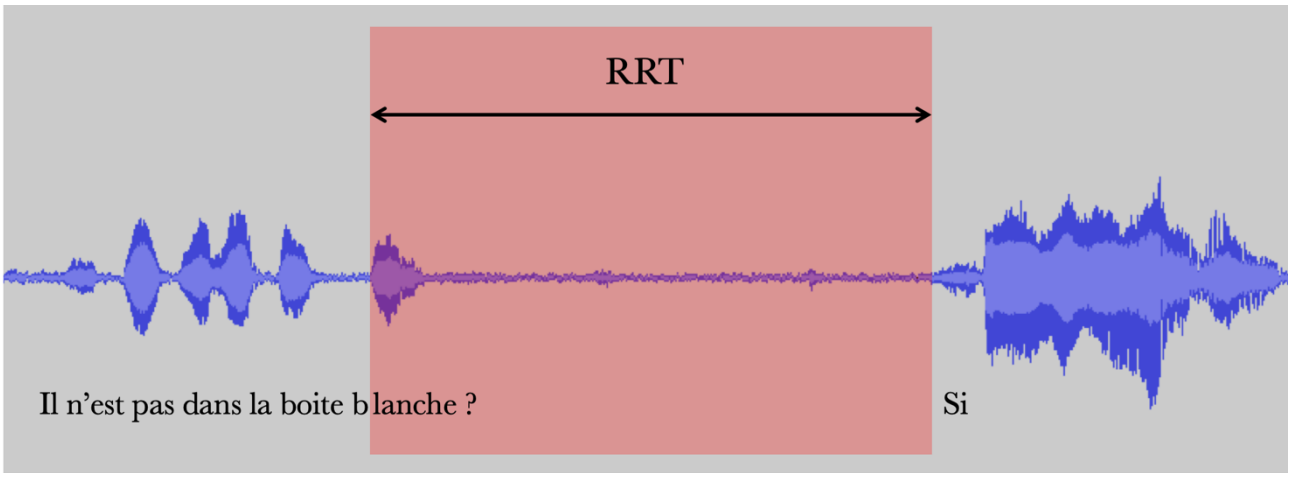


Fig. 5. Posterior distribution for each parameter in the fixed effect part of the model with respect to Accuracy in Experiment 1. The x-axis represents the estimated difference between conditions (with zero representing the "no difference" point) on a logit scale. The black dot represents the mean of the distribution, the green bar its range, and the orange line the 95% credible intervals (meaning that 95% of the distribution lies within its bars).

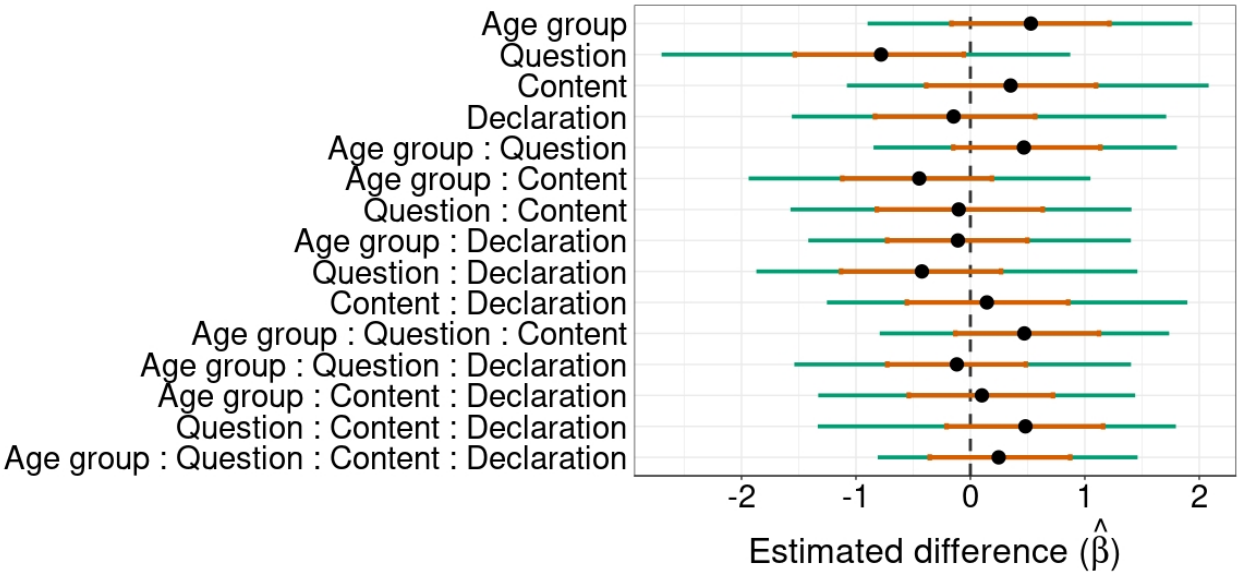
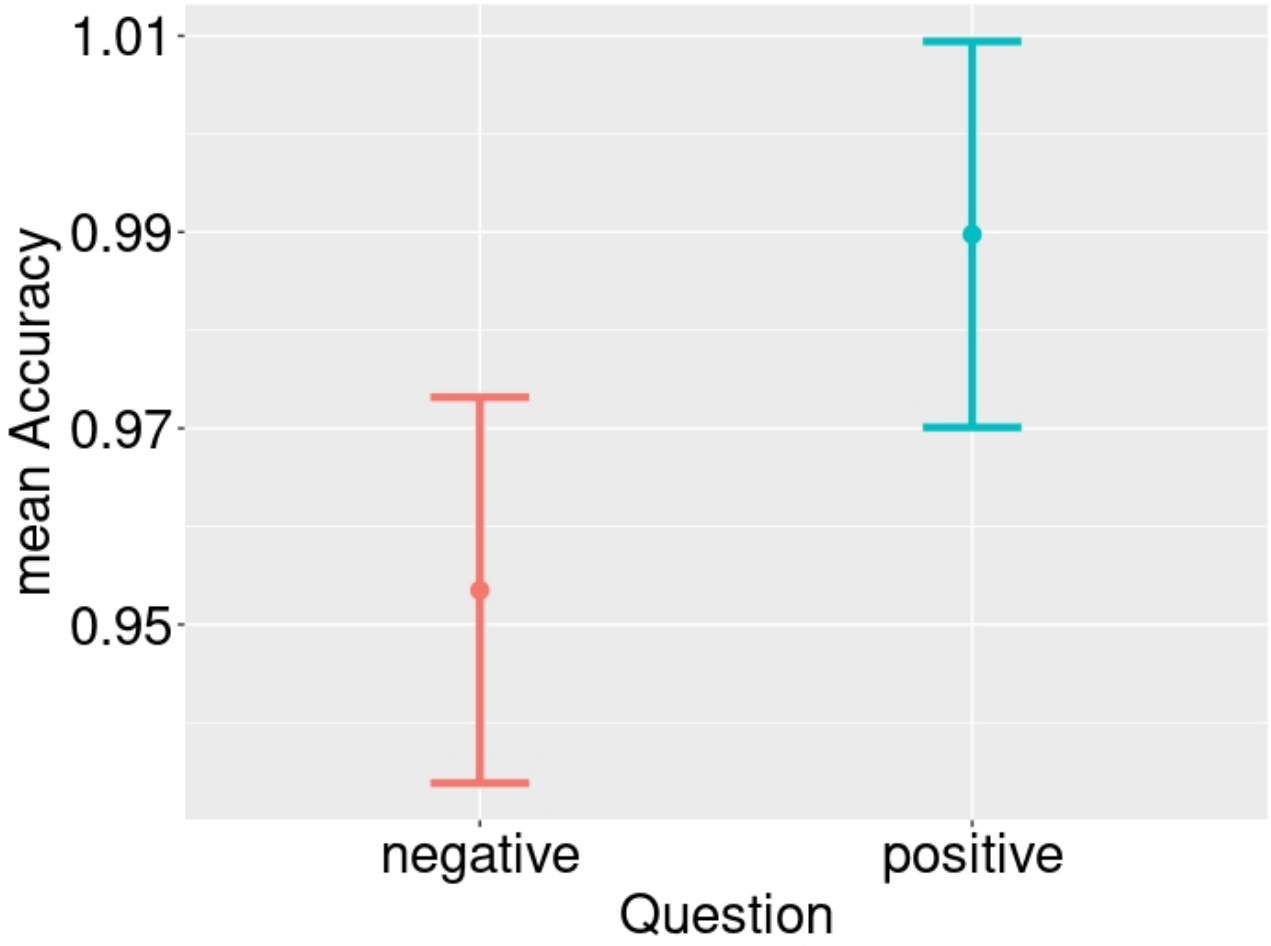


Fig. 6. Mean accuracy of Experiment 1 as function of question ("negative" vs. "positive"), with 95% confidence intervals.



Negative	Positive
0.953 (0.111)	0.989 (0.029)

Fig. 7. Posterior distribution for each parameter in the fixed effect part of the model with respect to Reaction Response times (RRTs) in Experiment 1. The x-axis represents the estimated difference between conditions (with zero representing the "no difference" point). The black dot represents the mean of the distribution, the green bar its range, and the orange line the 95% credible intervals (meaning that 95% of the distribution lies within its bars).

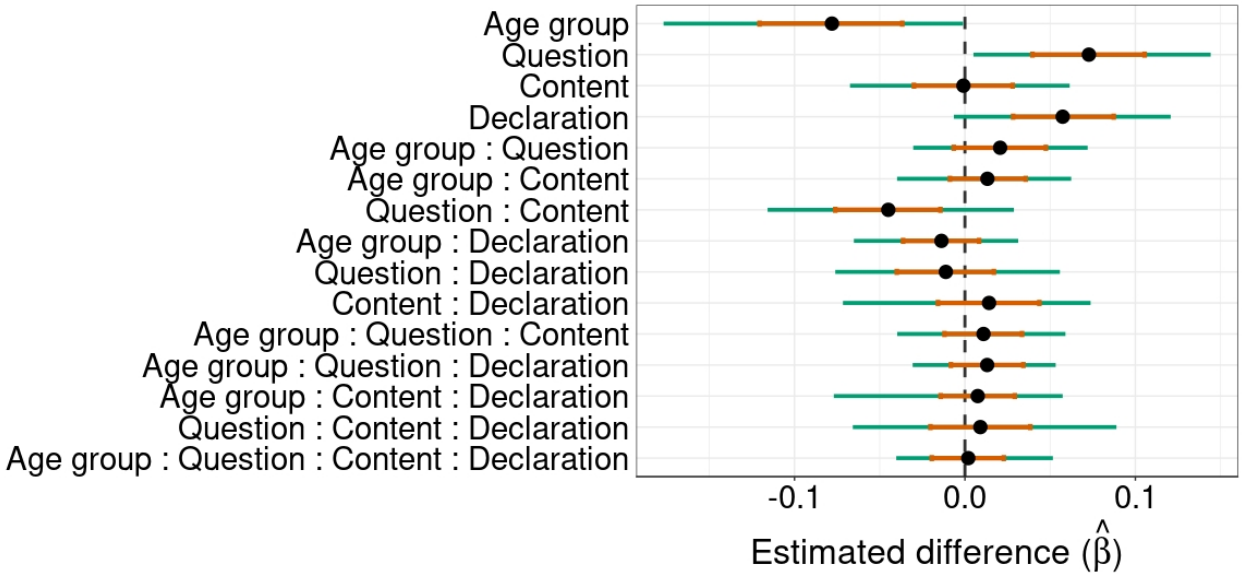
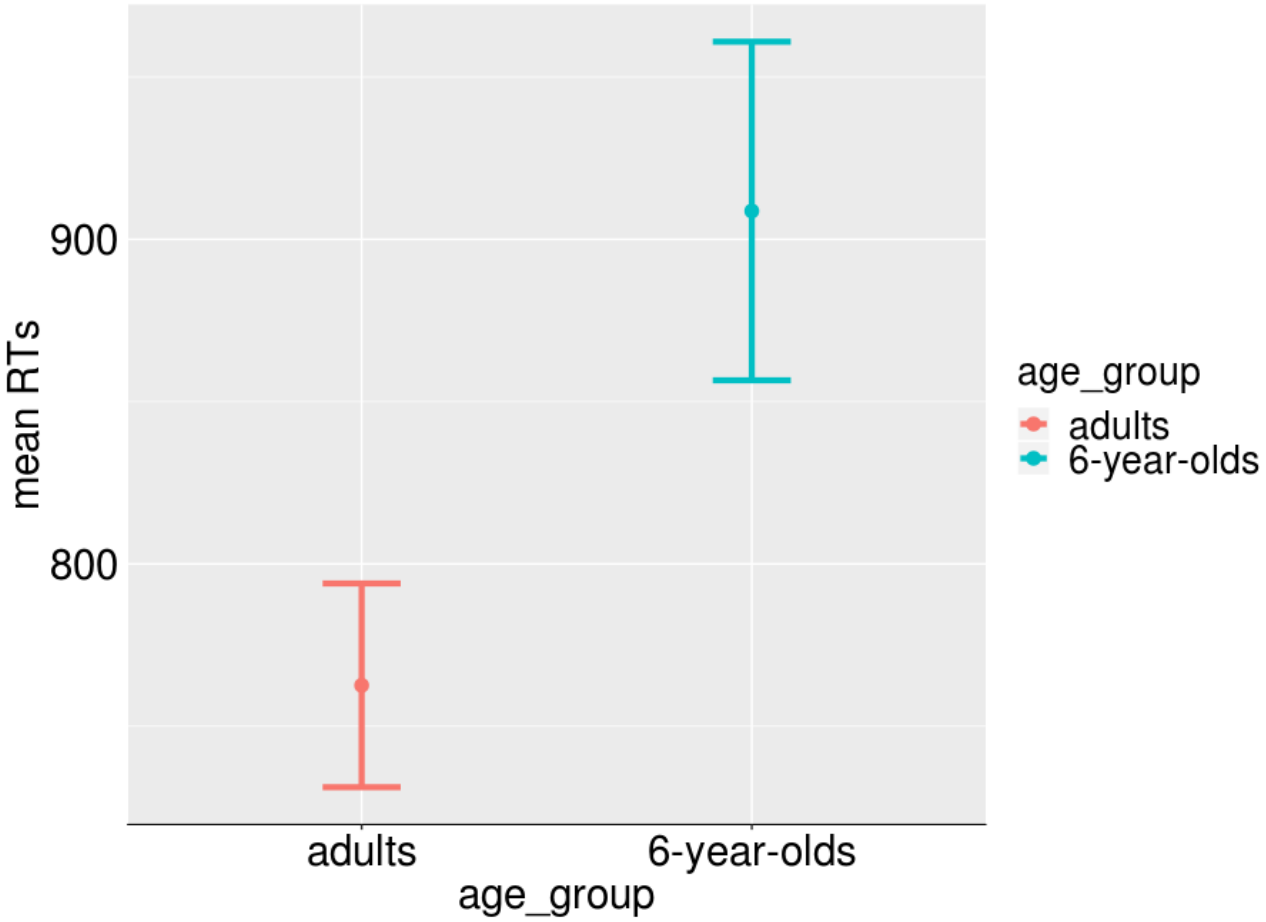


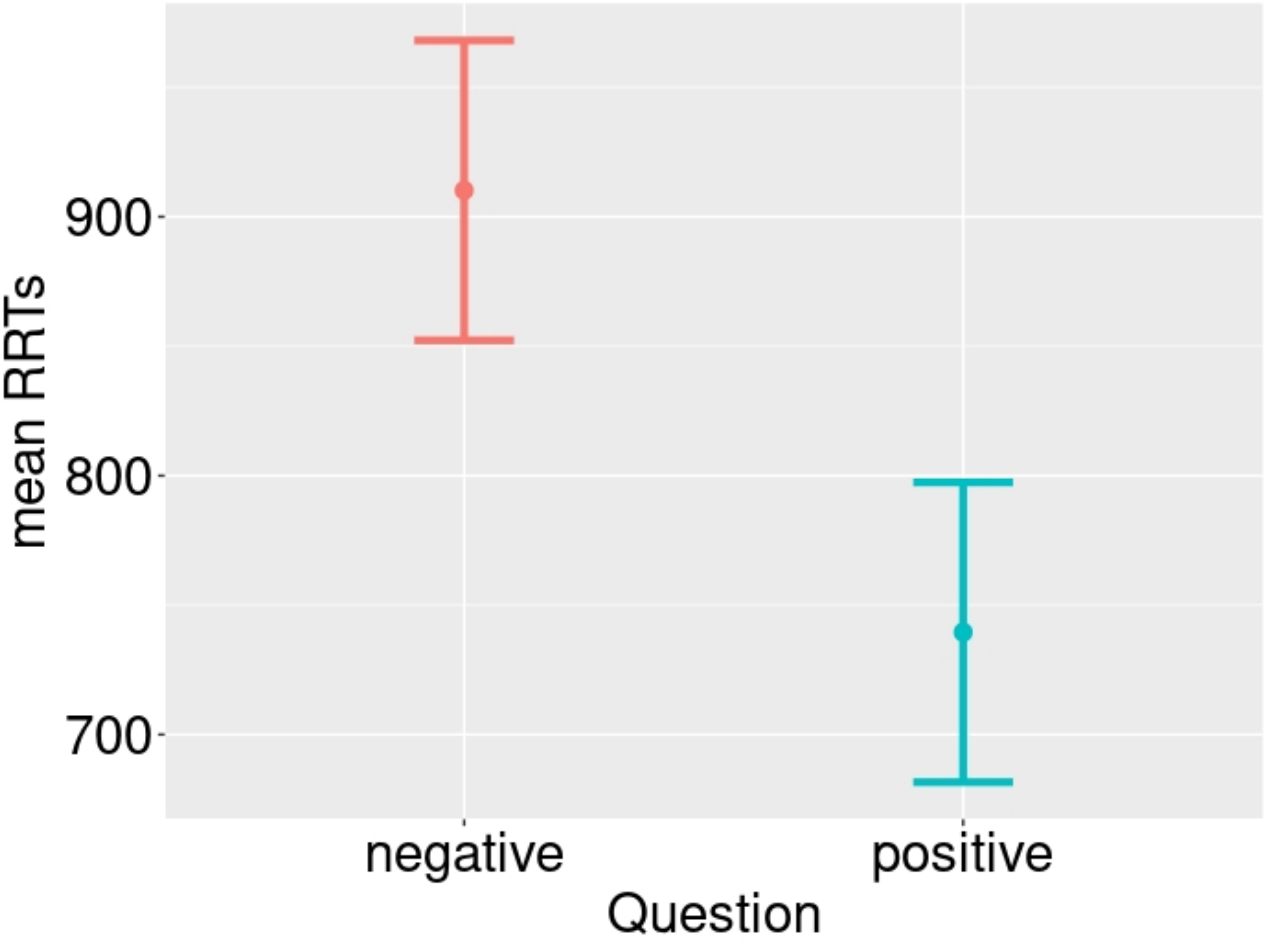
Fig. 8. Mean RRTs of Experiment 1 as function of age group ("adults" vs. "6-year-olds"), with 95% confidence intervals.



In this Table (and all tables throughout the appendix) we show the mean and the standard deviation in brackets.

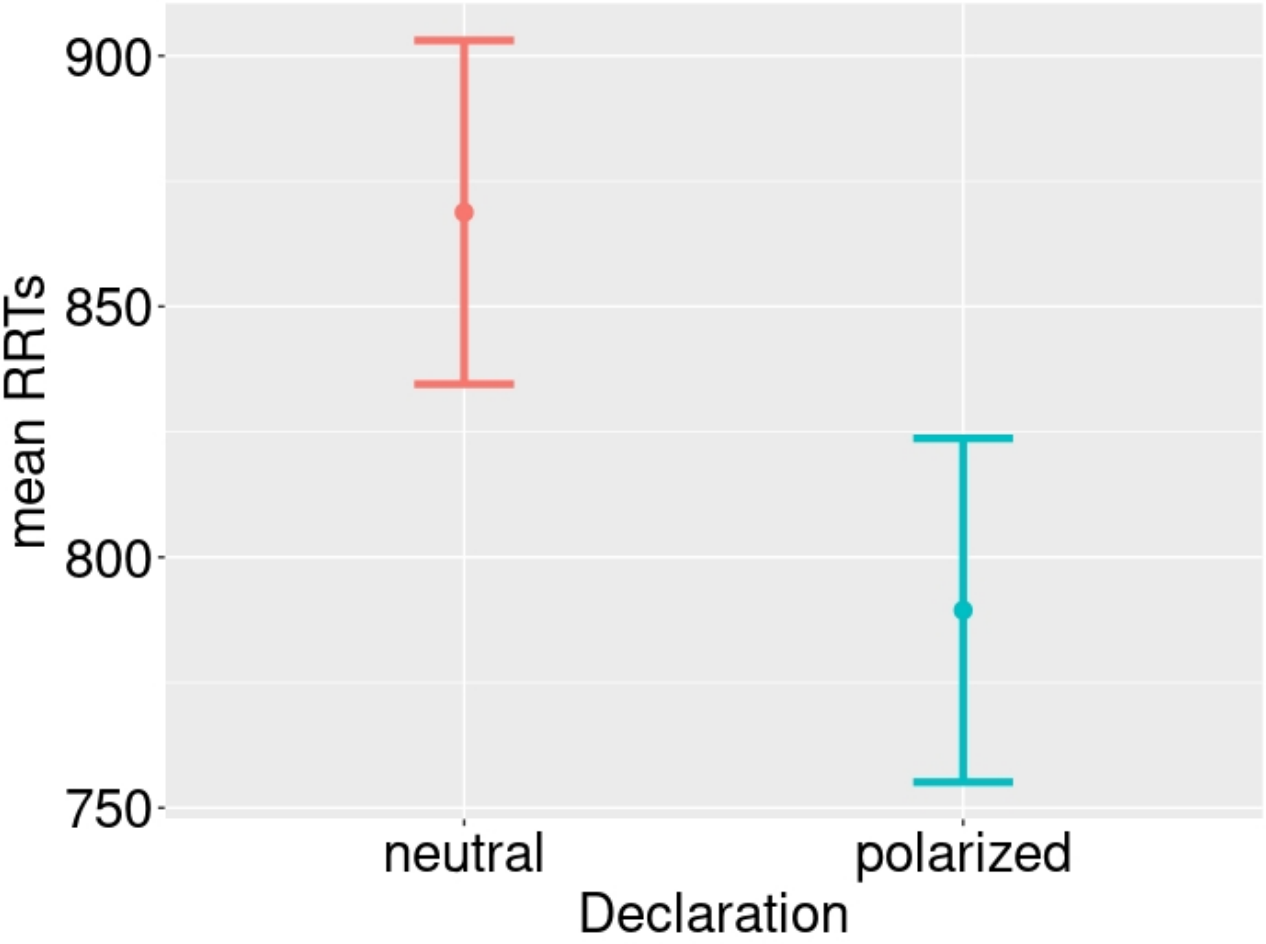
adults	**6-year-olds**
759.11 ms (142.58)	916.65 ms (175.76)

Fig. 9. Mean RRTs of Experiment 1 as function of question ("negative" vs. "positive"), with 95% confidence intervals.



negative	positive
910.17 ms (341.91)	739.53 ms (149.58)

Fig. 10. Mean RRTs of Experiment 1 as function of declaration ("neutral" vs. "polarized"), with 95% confidence intervals.



neutral	polarized
868.77 ms (211.05)	789.40 ms (185.21)

Fig. 11a. Mean RRTs of Experiment 1 as function of question ("negative" vs. "positive") and content ("empty" vs. "full" box content), with 95% confidence intervals.

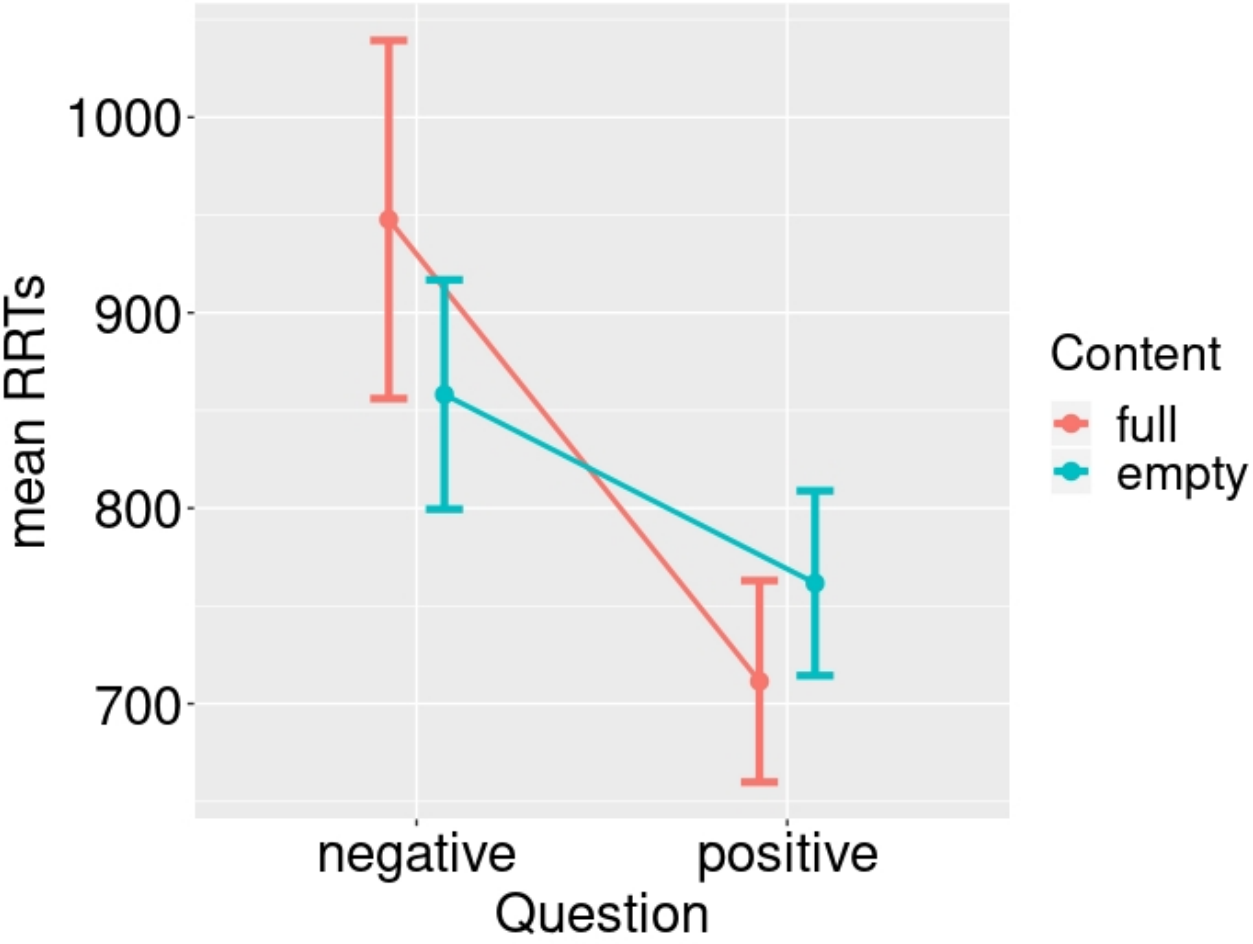
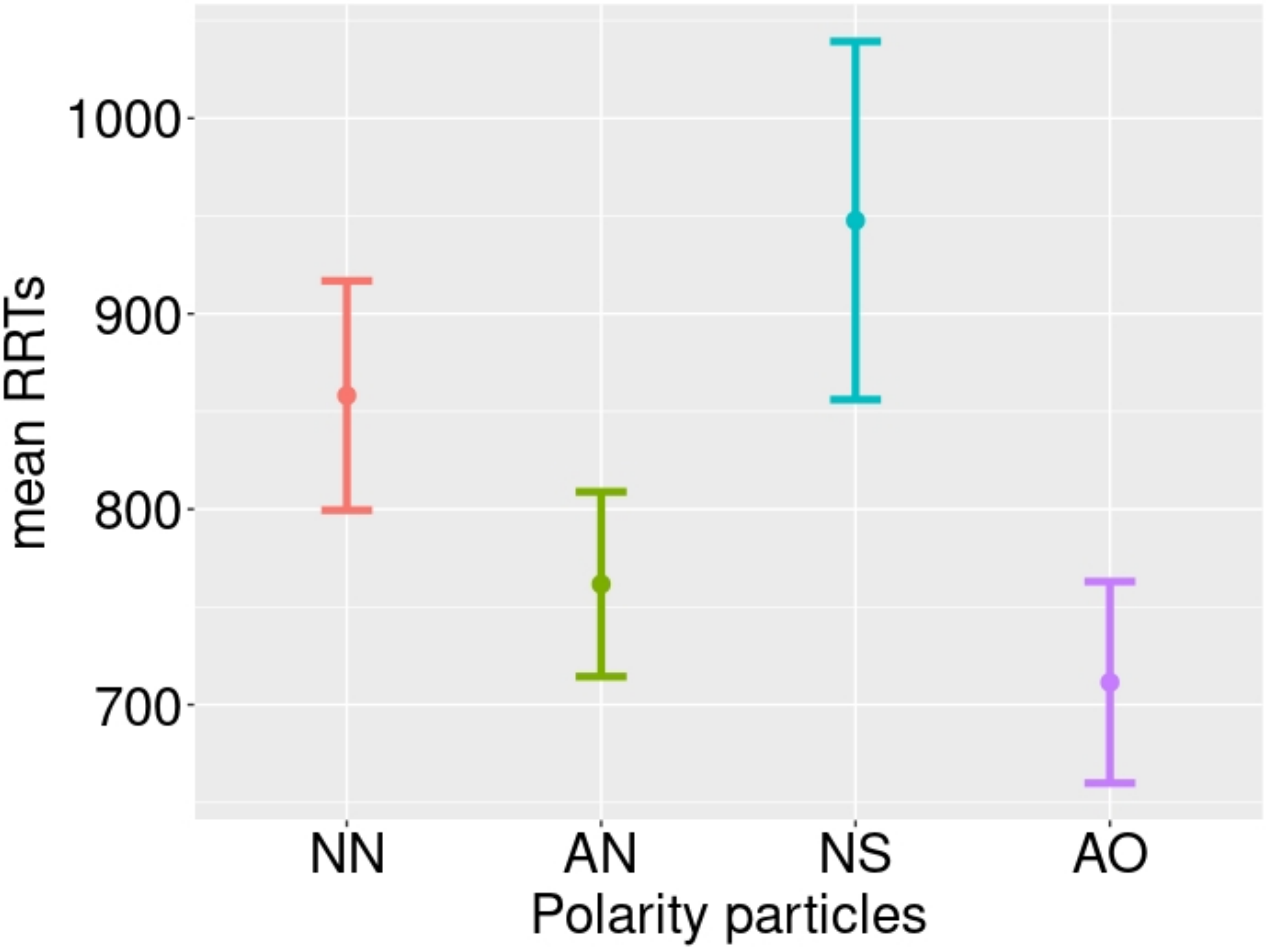


Fig. 11b. Mean RRTs of Experiment 1 of each polarity particle: *Si* responses are triggered by negative question after a full box (NS condition), *Oui* responses by affirmative questions after an empty box (AO), and *Non* responses after affirmative (AN) negative (NN) questions revealing empty boxes, with 95% confidence intervals.



NN	AN	NS	AO
858.15 (296.99)	761.64 (166.27)	947.67 (457.52)	711.42 (175.31)

Fig. 12. Secondary Analyses of Accuracy in Experiment 1: Posterior distribution for each parameter in the fixed effect part of the model with respect to Accuracy in Experiment 1. The x-axis represents the estimated difference between conditions (with zero representing the "no difference" point) on a logit scale. The black dot represents the mean of the distribution, the green bar its range, and the orange line the 95% credible intervals (meaning that 95% of the distribution lies within its bars).

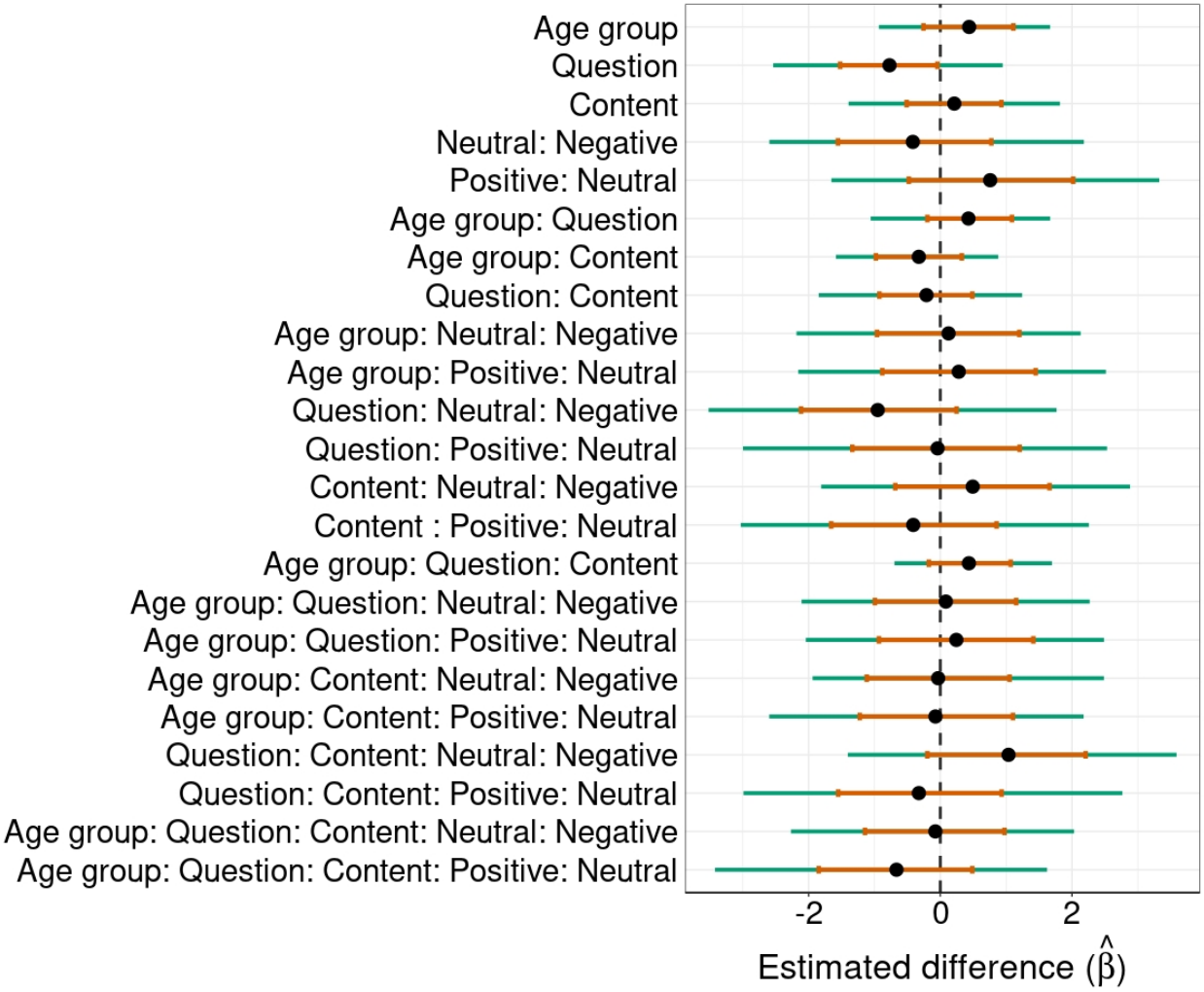
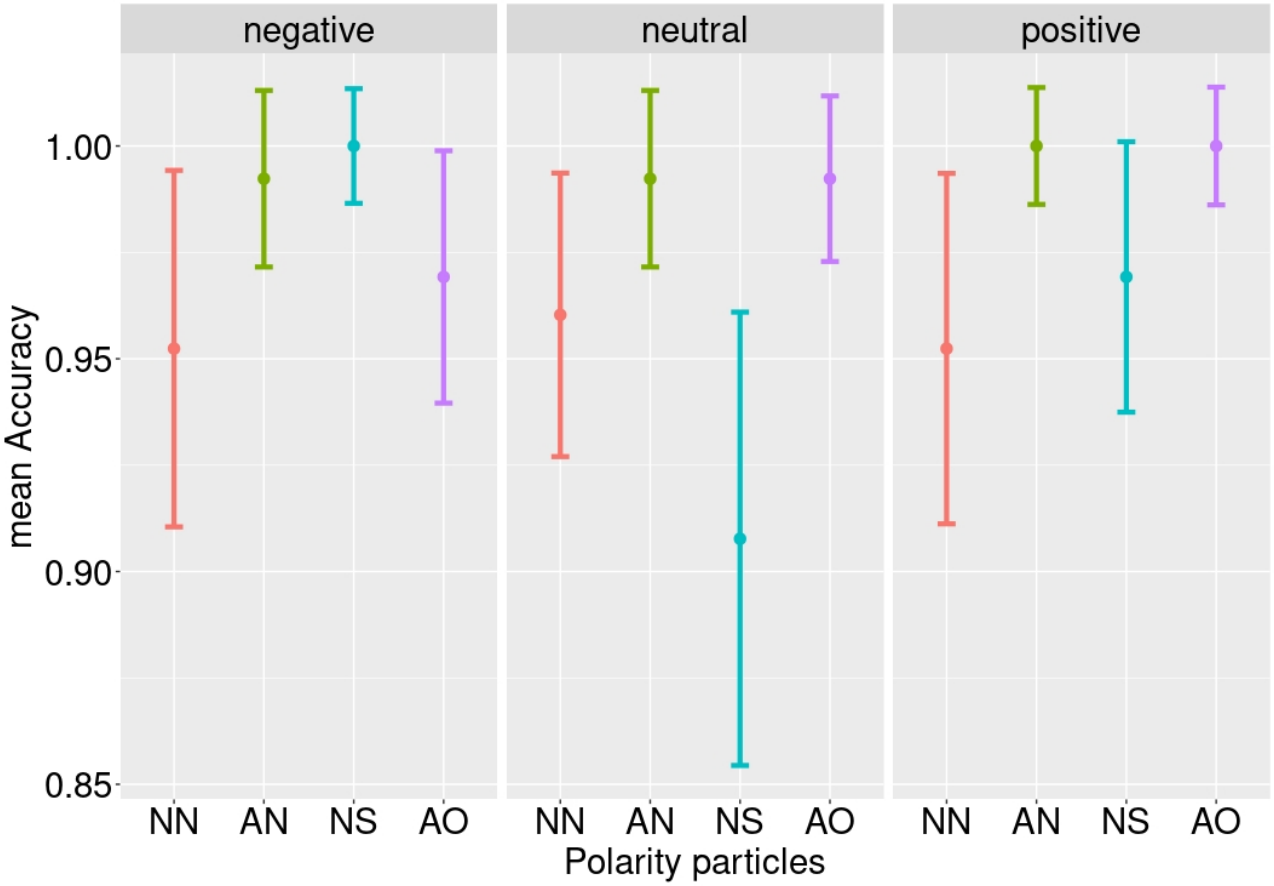


Fig. 13. Secondary analysis: Mean accuracy of Experiment 1 as function of polarity particle (Question by Content) and declaration (neutral vs. negative, neutral vs. positive) with 95% confidence intervals.



Declaration/Particles	**NN**	**AN**	**NS**	**AO**
Negative	0.95 (0.19)	0.99 (0.62)	1.00 (0.00)	0.97 (0.12)
Neutral	0.96 (0.16)	0.99 (0.62)	0.90 (0.23)	0.99 (0.06)
Positive	0.95 (0.19)	1.00 (0.00)	0.97 (0.12)	1.00 (0.00)

Fig. 14. Secondary analyses of Experiment 1 RRTs: Posterior distribution for each parameter in the fixed effect part of the model with respect to Reaction Response times (RRTs) in Experiment 1. The x-axis represents the estimated difference between conditions (with zero representing the "no difference" point). The black dot represents the mean of the distribution, the green bar its range, and the orange line the 95% credible intervals (meaning that 95% of the distribution lies within its bars).

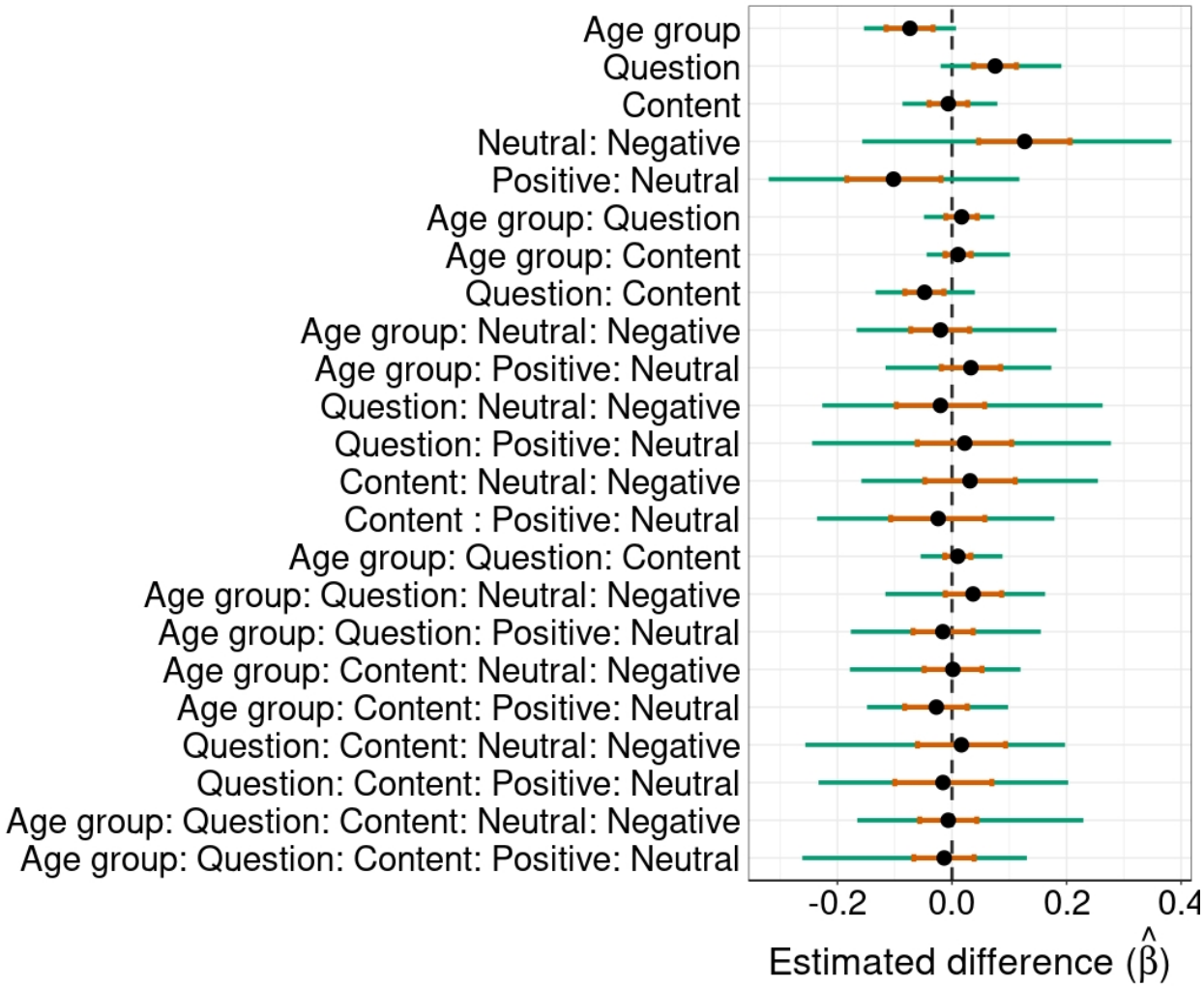
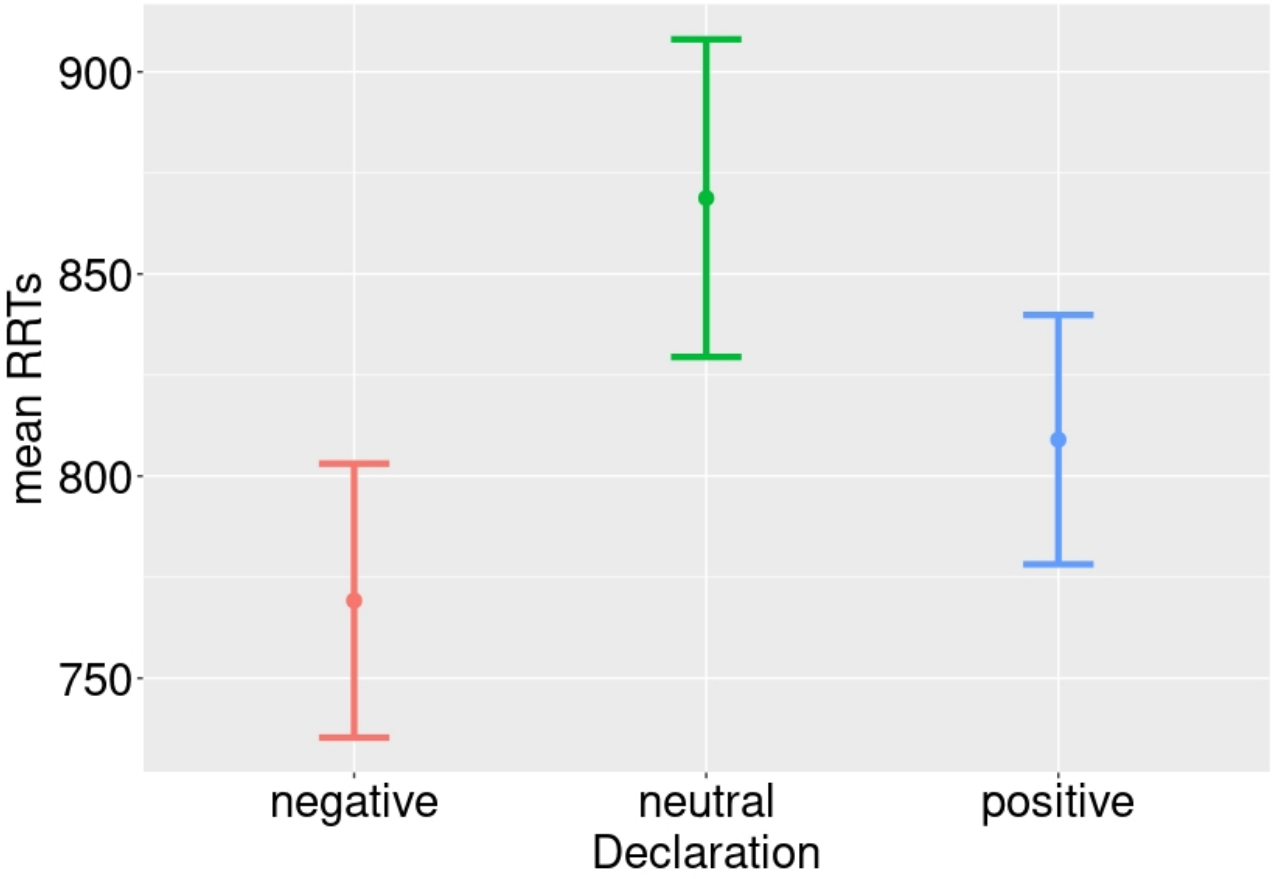
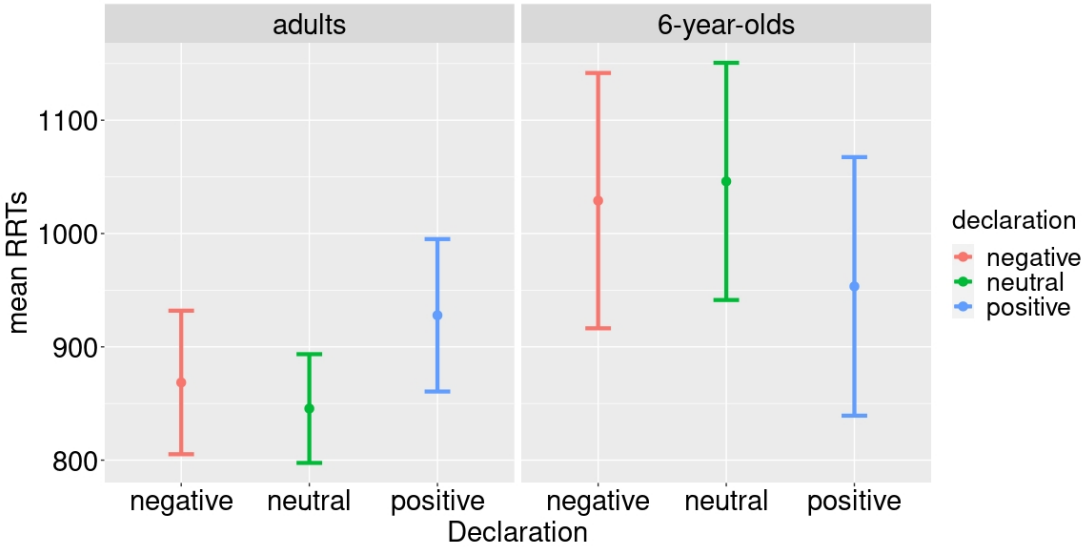


Fig. 15. Mean RRTs of Experiment 1 as function of Declaration, with 95% confidence intervals.



Negative	Neutral	Positive
769.18 (199.31)	868.77 (211.05)	809.01 (203.74)

Fig. 16. Tertiary analyses. Mean RRTs for responding *Si* in Experiment 1separately for adults and for 6-year-olds as a function of Declaration, (c), with 95% confidence intervals.



Declaration/Particles	**adults**	**6-year-olds**
Negative	868.5 (739.7)	1029 (390.6)
Neutral	845.5 (209)	1046 (310.2)
Positive	927.8 (489.2)	953.3 (264.8)

Fig. 17. Posterior distribution for each parameter in the fixed effect part of the model with respect to Accuracy in Experiment 2. The x-axis represents the estimated difference between conditions (with zero representing the "no difference" point) on a logit scale. The black dot represents the mean of the distribution, the green bar its range, and the orange line the 95% credible intervals (meaning that 95% of the distribution lies within its bars).

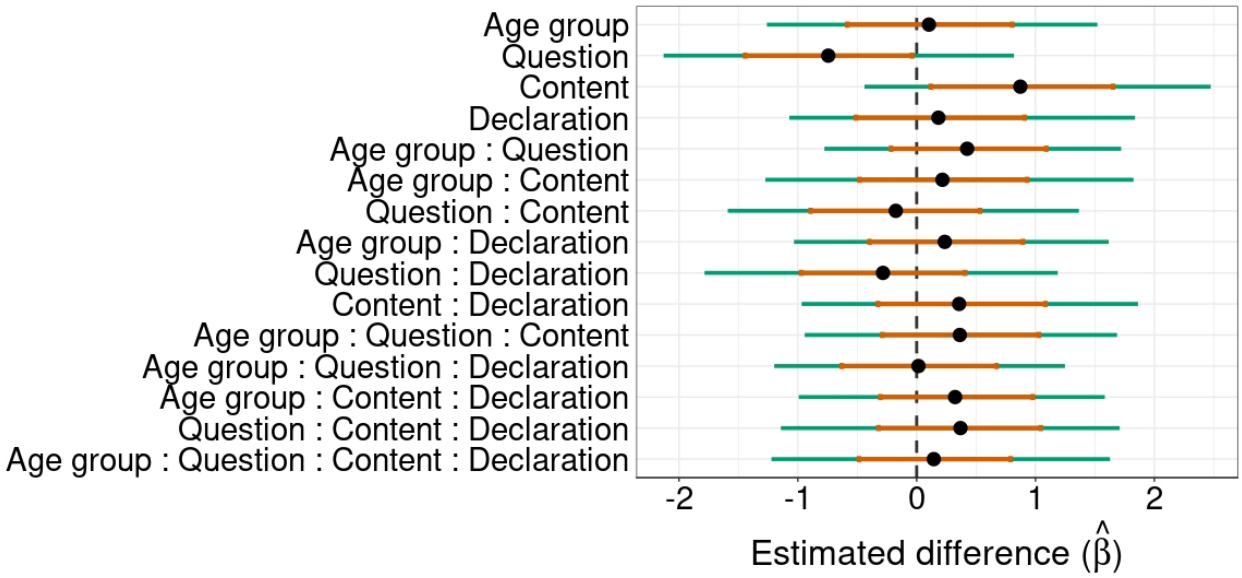
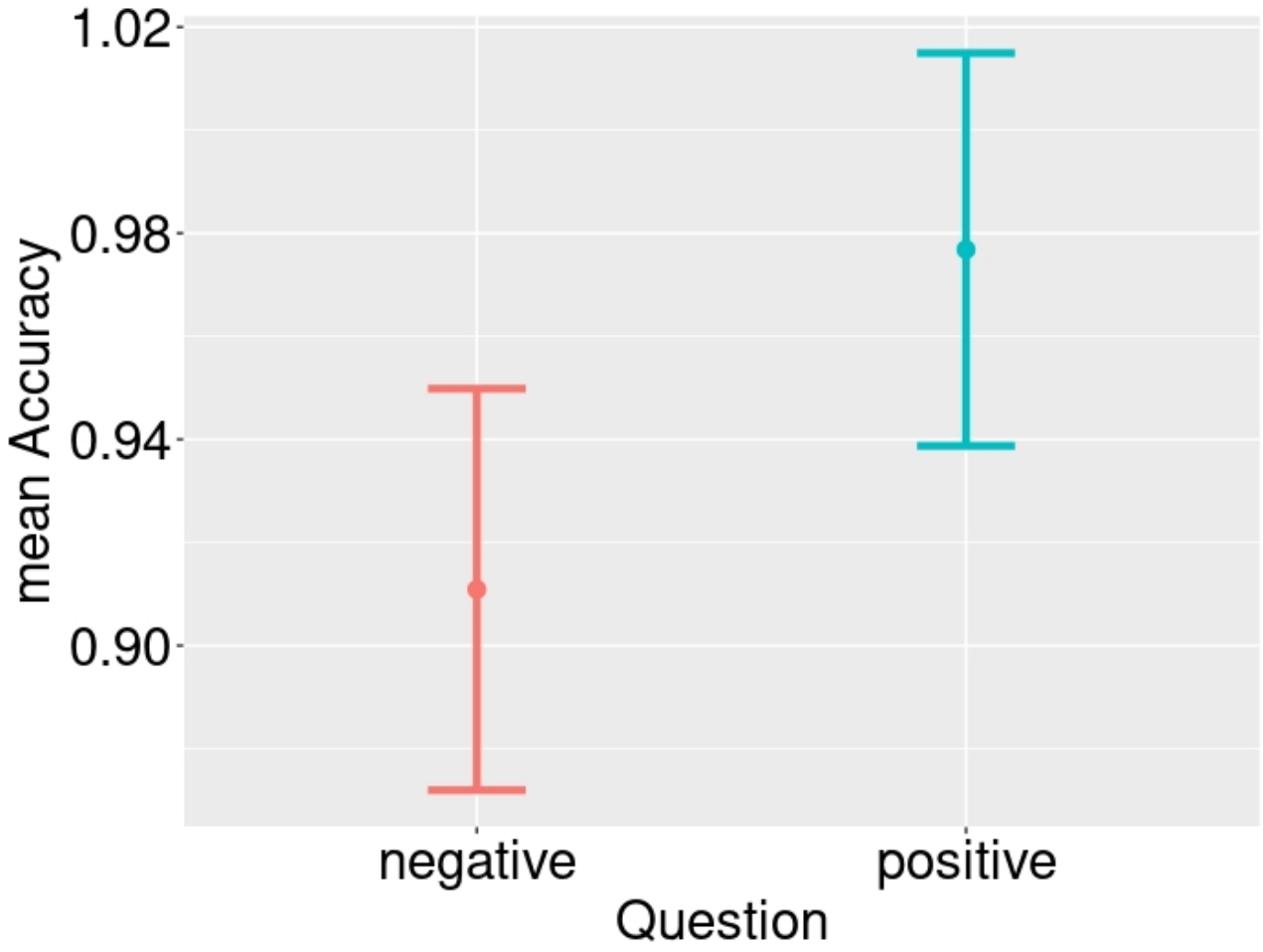
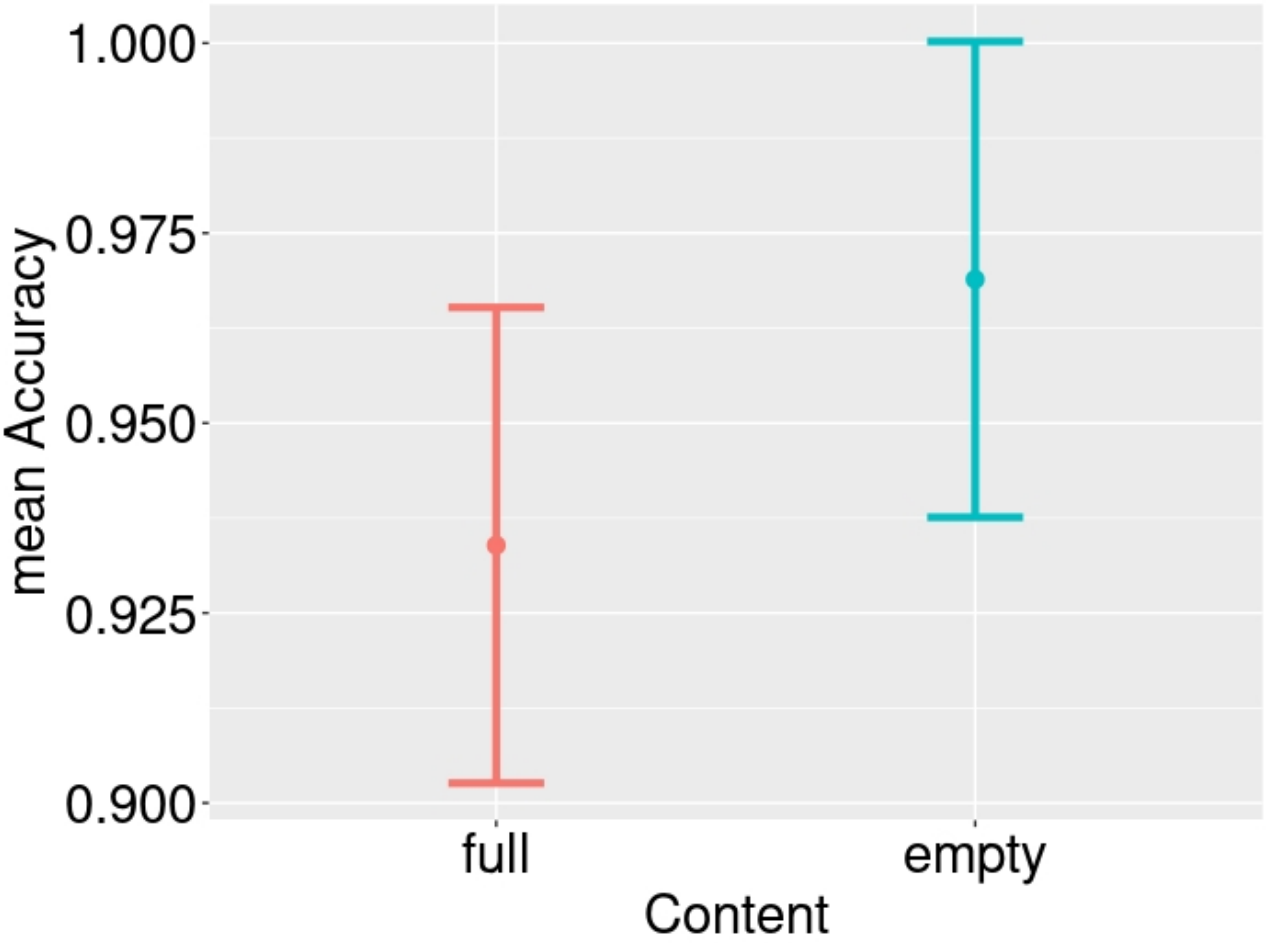


Fig. 18. Mean accuracy of Experiment 2 as function of question ("negative" vs. "positive"), with 95% confidence intervals.



negative	positive
0.910 (0.176)	0.977 (0.049)

Fig. 19. Mean accuracy of Experiment 2 as function of content ("full" vs. "empty"), with 95% confidence intervals.



full	empty
0.933 (0.103)	0.968 (0.093)

Fig. 20. Posterior distribution for each parameter in the fixed effect part of the model with respect to Response Reaction times (RRTs) in Experiment 2. The x-axis represents the estimated difference between conditions (with zero representing the "no difference" point). The black dot represents the mean of the distribution, the green bar its range, and the orange line the 95% credible intervals (meaning that 95% of the distribution lies within its bars).

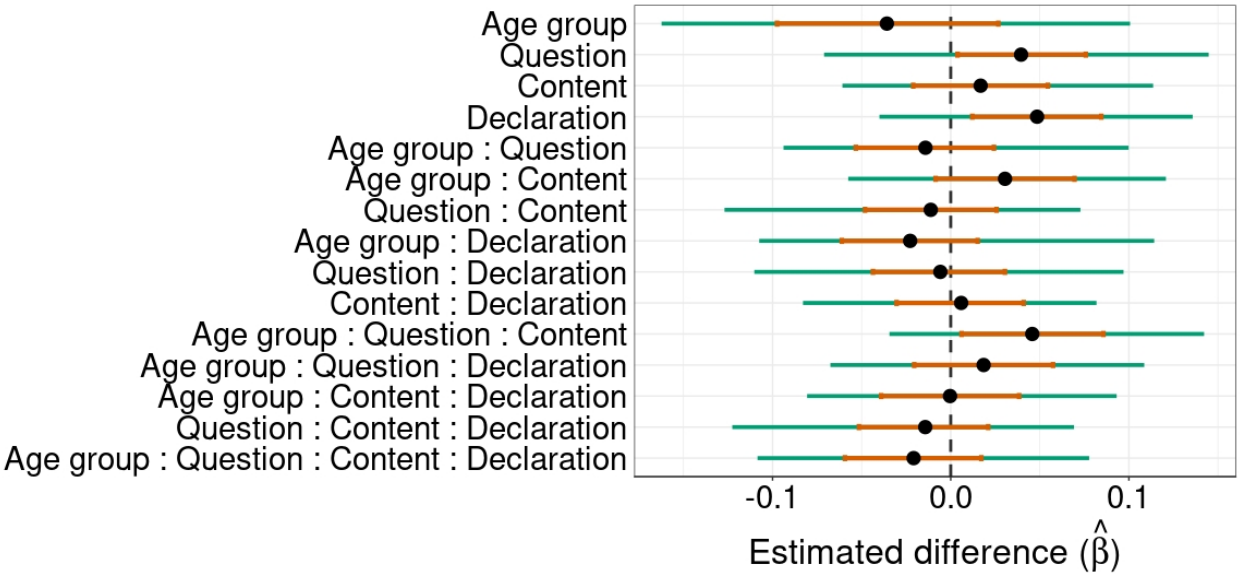
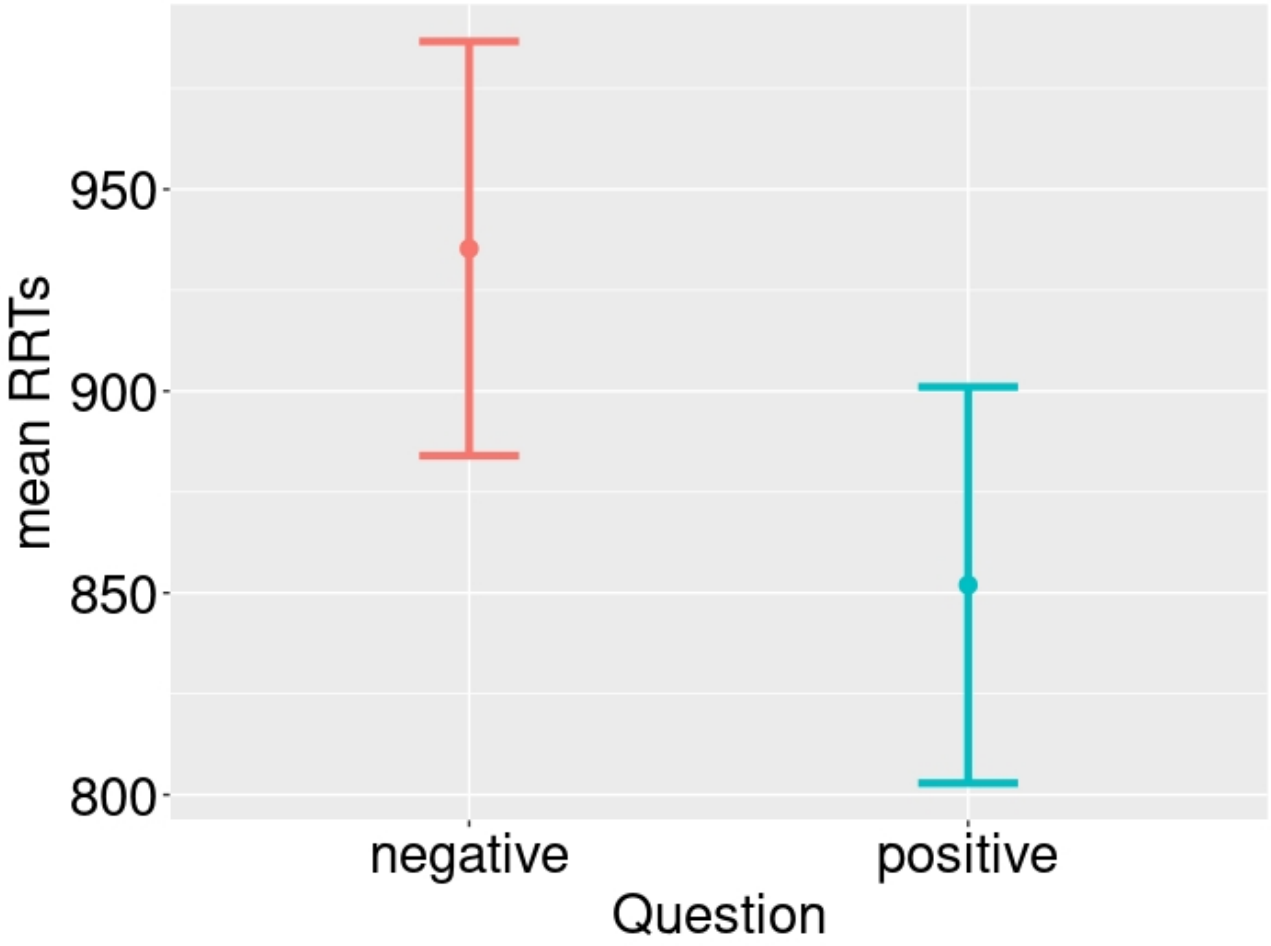
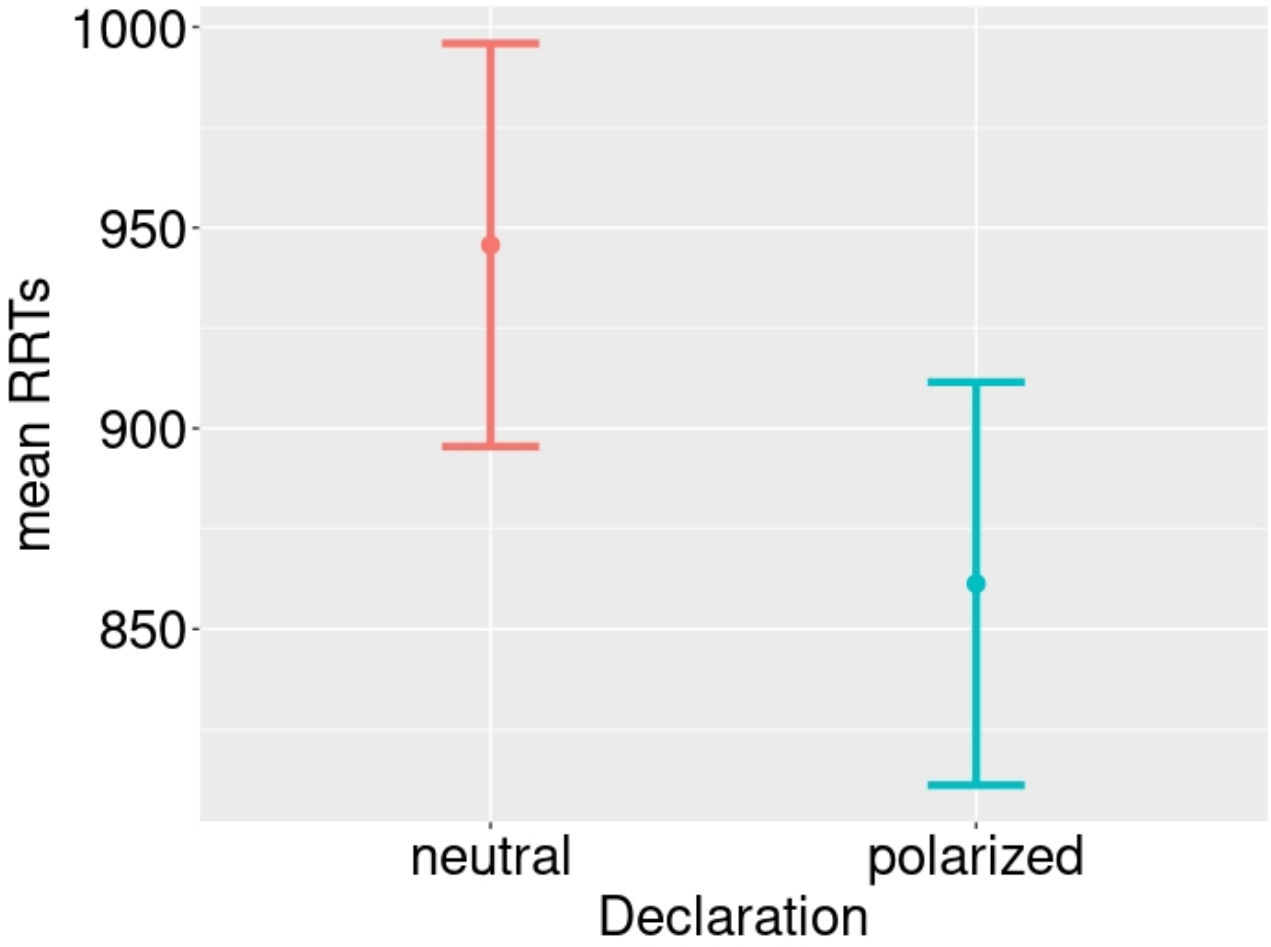


Fig. 21. Mean RRTs of Experiment 2 as function of questions ("negative" vs. "positive"), with 95% confidence intervals.



negative	positive
935.29 (249)	851.96 (217.82)

Fig. 22. Mean RRTs of Experiment 2 as function of declaration ("neutral" vs. "polarized"), with 95% confidence intervals.



neutral	polarized
945.68 (267.49)	861.32 (204.51)

Fig. 23a. Mean RRTs of Experiment 2 as a function of age group ("four-year-olds" vs. "6-year-olds"), question ("negative" vs. "positive") and content ("full" vs. "empty"), with 95% confidence intervals.

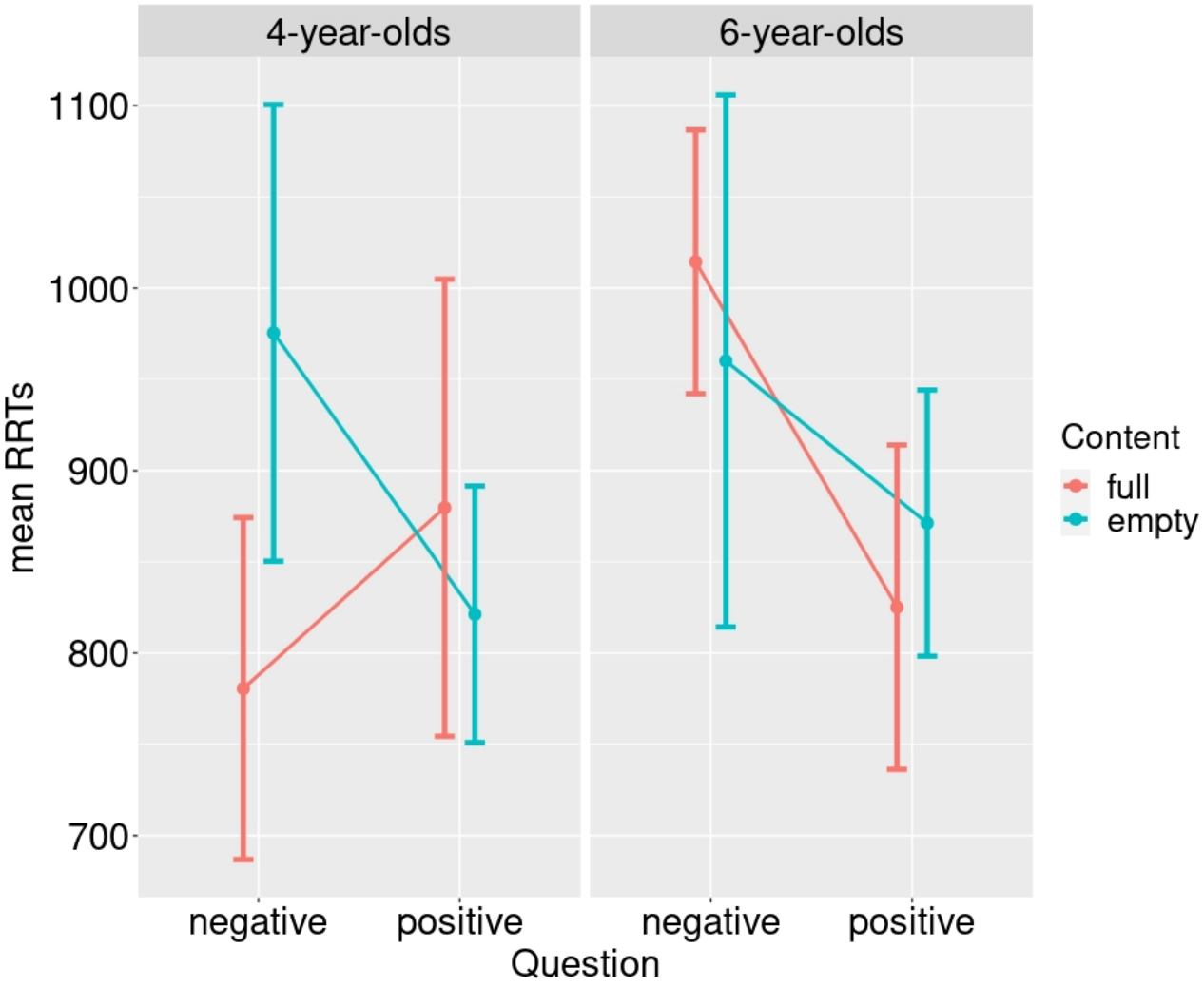
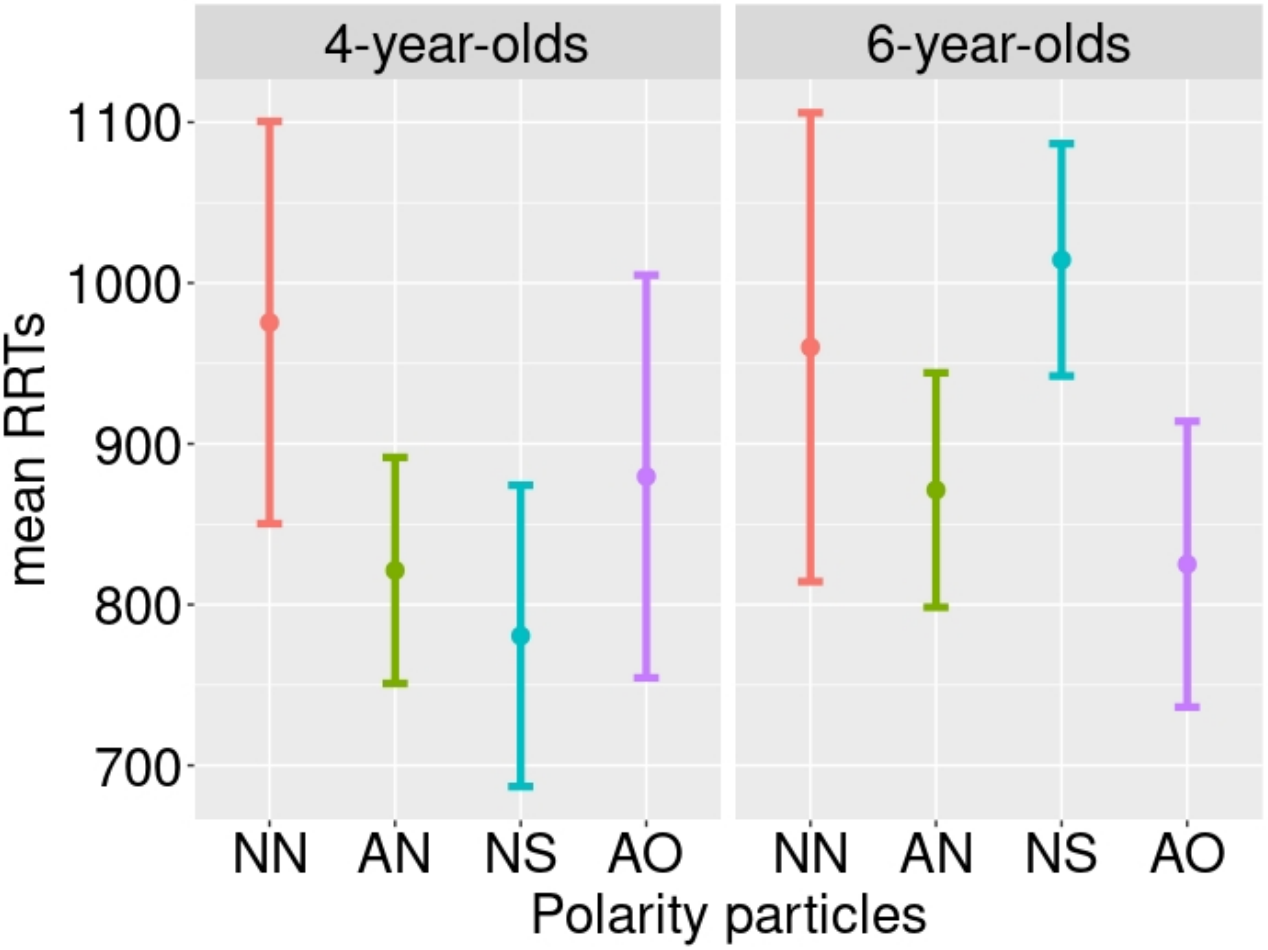


Fig. 23b. Mean RRTs of Experiment 2 as a function of age group ("four-year-olds" vs. "6-year-olds") and responses (question by content: "Negative-Non" vs. "Affirmative-Non" vs. "Negative-Si" vs. "Affirmative-Oui") with 95% confidence intervals.



Age group/Particles	**NN**	**AN**	**NS**	**AO**
6-year-olds	960.07 (430.15)	871.24 (192.82)	1014.40 (221.30)	825.10 (219.39)
4-year-olds	975.42 (374.08)	821.25 (203.15)	780.54 (202.79)	879.66 (424.95)

Fig. 24. Secondary analyses of experiment 2 Accuracy: Posterior distribution for each parameter in the fixed effect part of the model with respect to Accuracy in Experiment 2. The x-axis represents the estimated difference between conditions (with zero representing the "no difference" point) on a logit scale. The black dot represents the mean of the distribution, the green bar its range, and the orange line the 95% credible intervals (meaning that 95% of the distribution lies within its bars).

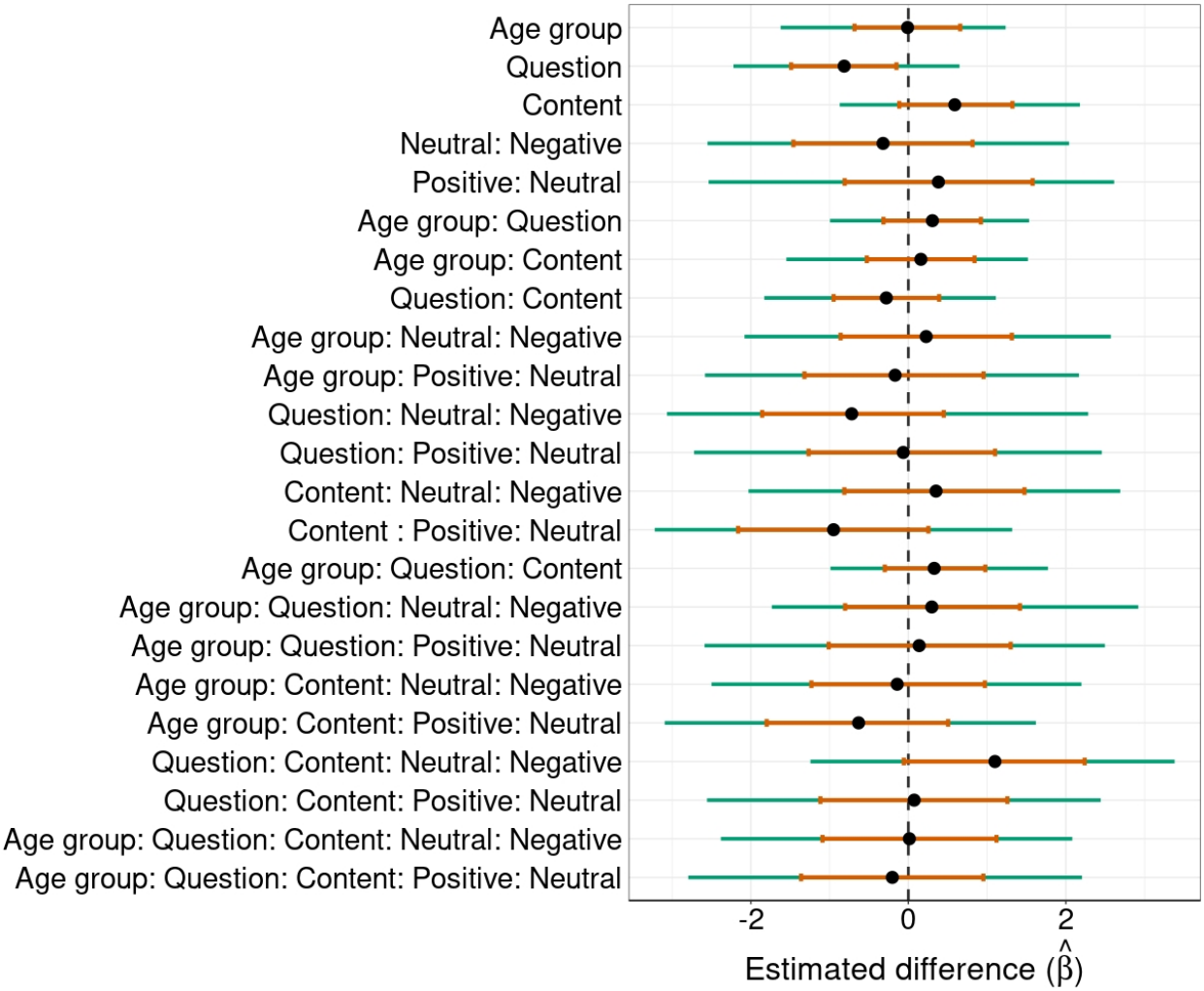
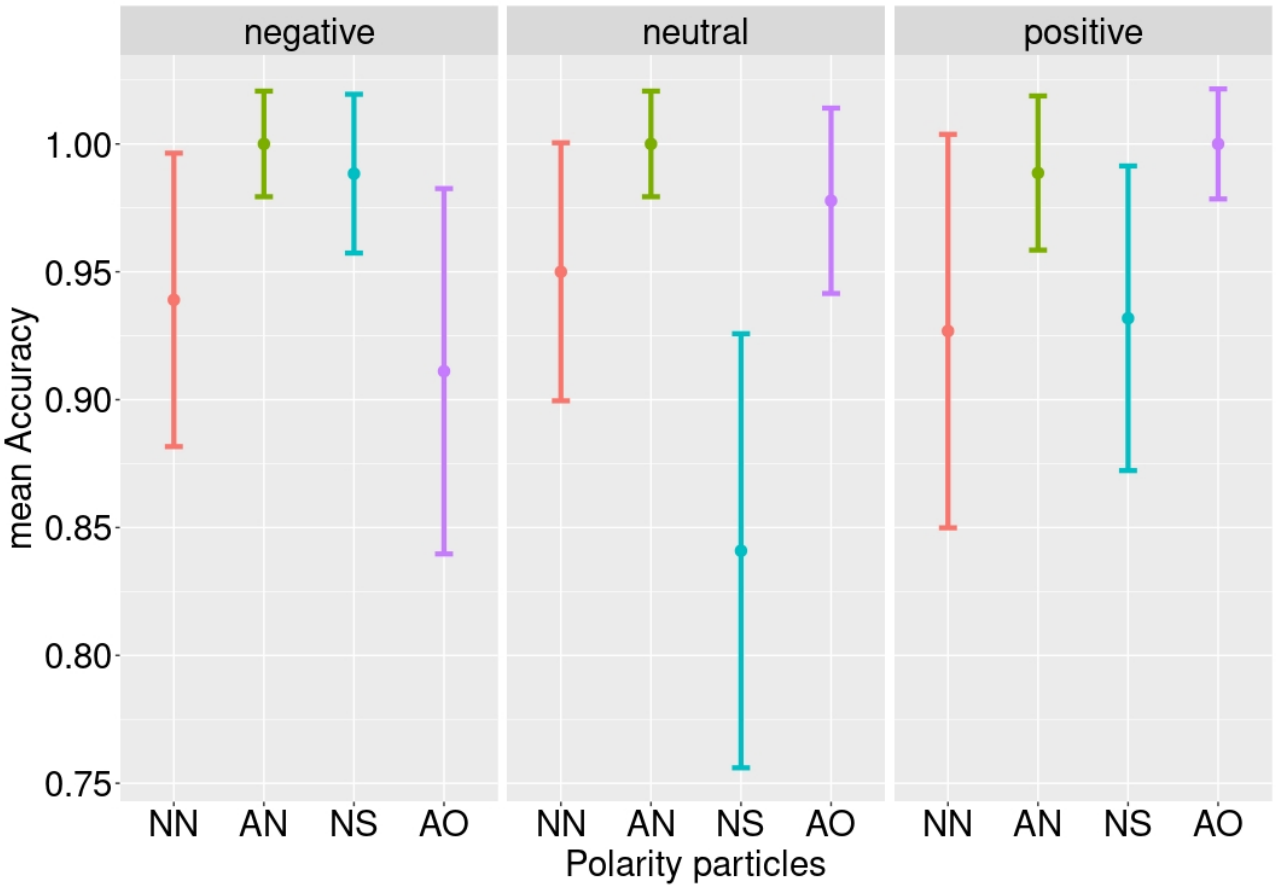


Fig. 25. Mean accuracy of Experiment 2 as function of polarity particles and declaration, with 95% confidence intervals.



Declaration/Particles	**NN**	**AN**	**NS**	**AO**
Negative	0.94 (0.19)	1.00 (0.00)	0.98 (0.07)	0.91 (0.24)
Neutral	0.95 (0.19)	1.00 (0.00)	0.84 (0.30)	0.97 (0.10)
Positive	0.93 (0.26)	0.98 (0.75)	0.93 (0.20)	1.00 (0.00)

Fig. 26. Secondary analyses of Experiment 2 RRTs: Posterior distribution for each parameter in the fixed effect part of the model with respect to Response Reaction times (RRTs) in Experiment 2. The x-axis represents the estimated difference between conditions (with zero representing the "no difference" point). The black dot represents the mean of the distribution, the green bar its range, and the orange line the 95% credible intervals (meaning that 95% of the distribution lies within its bars).

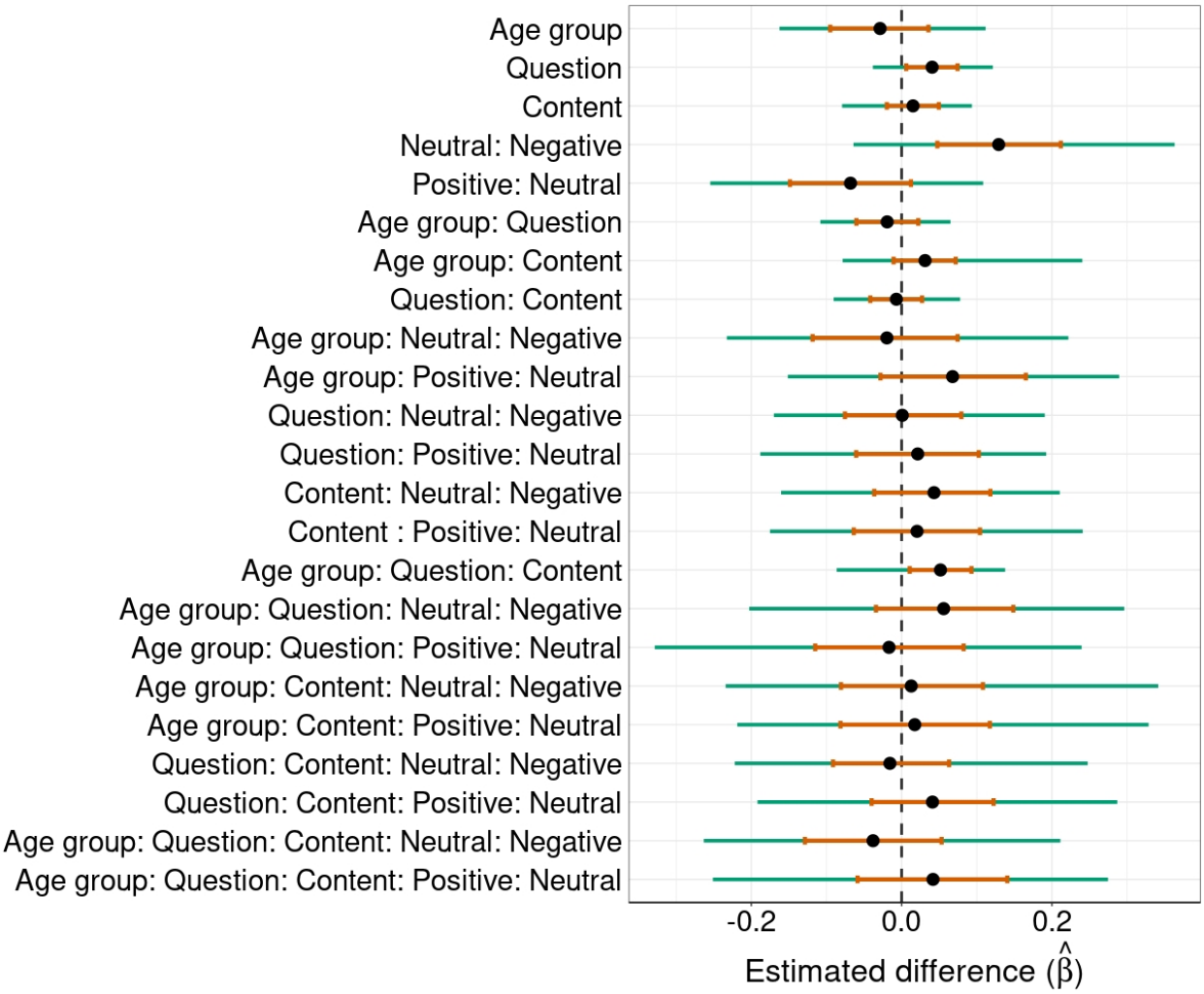
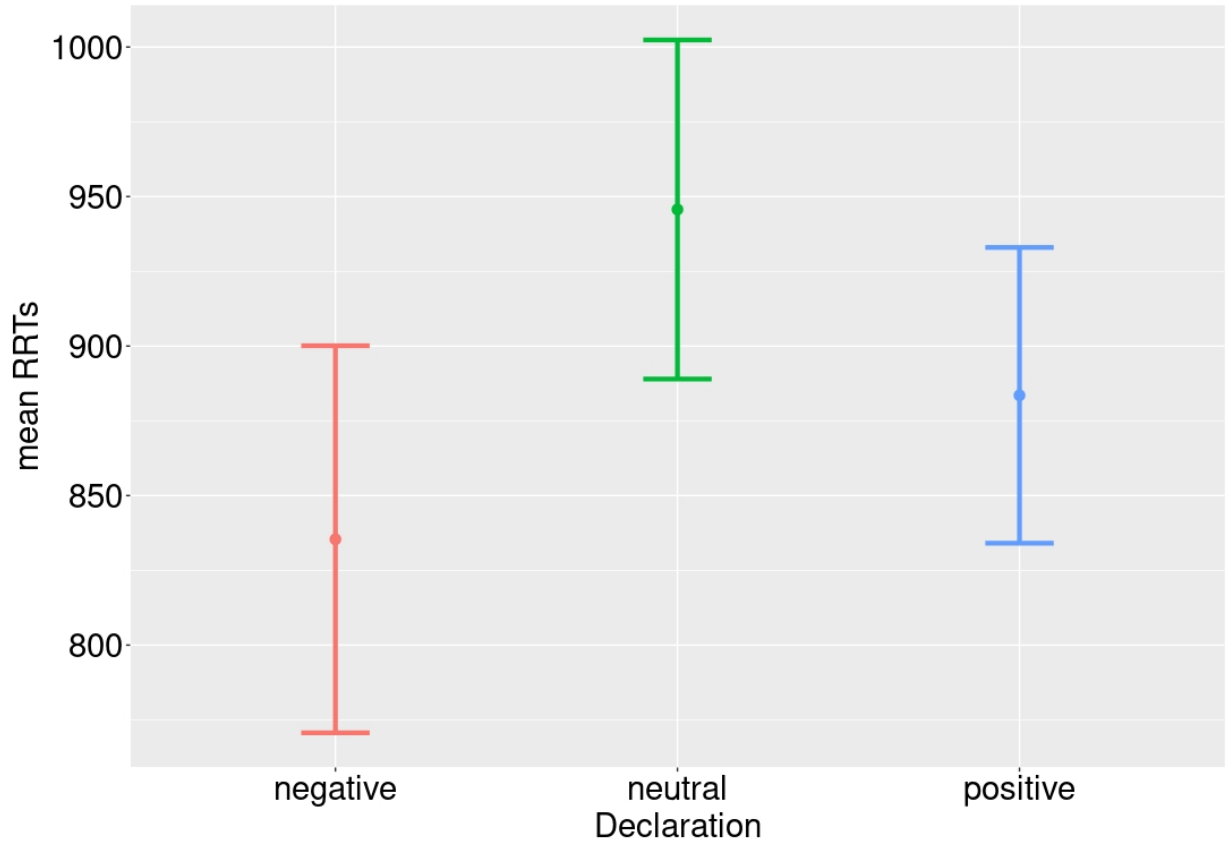


Fig. 27. Mean RRTs of Experiment 2 as function of declaration, with 95% confidence intervals.



Negative	Neutral	Positive
835.38 (248.48)	945.68 (267.49)	883.52 (231.80)