Journal of Child Language (2026), 1–23
doi:10.1017/S0305000926100877

CAMBRIDGE UNIVERSITY PRESS

RESEARCH ARTICLE

Age-Related Changes in Implicit Causality and Consequentiality Pronoun Biases

Amanda Rose Yuile and Joshua K Hartshorne

Communication Sciences and Disorders, MGH Institute of Health Professions, USA
Corresponding author: Amanda Rose Yuile; Email: ayuile@mghihp.edu

(Received 10 July 2025 UTC; revised 28 May 2026 UTC; accepted 30 June 2026 UTC)

Abstract

When adults encounter pronouns in sentences like "Sally frightened Mary because she…," they typically resolve the pronoun to the inferred cause of the event, a phenomenon called implicit causality. Similarly, in sentences like "Because Sally frightened Mary, she…," pronoun interpretation often reflects the inferred consequence of the event, known as implicit consequentiality. We re-analysed data from two publicly available cross-sectional studies (Hartshorne & Snedeker, 2013, Are implicit causality pronoun resolution biases consistent across languages and cultures? *Experimental Psychology*, *60*, 179; Hartshorne et al., 2015, Development of the first-mention bias. *Journal of Child Language*, *42*, 423) to explore age-related changes in implicit causality and implicit consequentiality pronoun biases. Dataset 1 included 978 native English speakers aged 9–81 years and Dataset 2 included 2,648 native English speakers aged 8–81 years. Both implicit causality and consequentiality showed significant developmental improvements from childhood into early adulthood, with implicit causality effects continuing to mature longer, suggesting that the underlying representations supporting pronoun biases continue developing into adulthood. There was no observable decline in pronoun bias, though results after 65 were highly variable.

Keywords: language acquisition; comprehension; pronouns; implicit causality; implicit consequentiality

1. Age-related changes in implicit causality and consequentiality pronoun biases

Pronouns like *she* or *he* are ubiquitous in language yet they pose a unique challenge for comprehension: their meaning is entirely context dependent. Correctly understanding a pronoun involves generating inferences about reference from syntactic, semantic, and pragmatic discourse factors. Despite decades of research on pronoun interpretation, we still know relatively little about how this skill develops and changes across the lifespan. This article investigates third-person singular pronoun comprehension across the lifespan, focusing on two well-documented semantic biases for pronoun interpretation: implicit causality and implicit consequentiality. To motivate this work, we first consider what is known about language development across the lifespan, and what insights such research can provide. We then consider pronoun interpretation in particular as a window

© The Author(s), 2026. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/S0305000926100877 Published online by Cambridge University Press

into broader discourse processes, an area that remains underexplored in research on lifespan language development.

Research on language acquisition tends to focus on the first few years of learning. This focus is understandable: early childhood is marked by dramatic and easily observable developmental milestones, while more subtle developments that occur later in life may go undetected in the smaller scale studies traditionally implemented in the field (Asendorpf et al., 2013; Hartshorne, 2020; Stanley et al., 2020). Moreover, many of the major theoretical debates in language acquisition over the last several decades have focused on the earliest emergence of language knowledge (Ambridge & Lieven, 2011; Fisher et al., 2020; Tomasello, 2005). In part for this reason, language acquisition researchers often investigate when children show sensitivity to language phenomena, rather than when they master them (Bowerman, 1994; Brown, 1973; Crain & Thornton, 2012; Lidz et al., 2003; Marcus et al., 1999).

Nonetheless, language learning continues long after early childhood. For example, it has long been recognized that vocabulary acquisition continues well into adulthood. This is not only intuitively true – adults regularly encounter and learn new words – but also empirically supported by numerous studies (Bowles & Salthouse, 2008; Doppelt & Wallace, 1955; Hartshorne & Germine, 2015; Singer et al., 2003; Verhaeghen, 2003). In terms of syntax, studies that look at early school-age children routinely show performance well south of ceiling. For instance, even by age 5, typically developing children still produce tense and agreement errors in ~10% of relevant utterances (Leonard, 2017; Rice & Wexler, 1996; Rice et al., 1998). More recent work indicates that reaching ceiling takes many years. In production, the length and complexity of individuals' utterances steadily increases into early adulthood (Berman, 2007; Montag & MacDonald, 2015; Nippold, 2007). In a large-scale cross-sectional study of comprehension, Hartshorne and his colleagues (Chen & Hartshorne, 2021; Hartshorne et al., 2018) report that growth in syntactic competence continues until age 40. Many pragmatics abilities are also notoriously slow to develop; for example, while young children *can* understand metaphors in a non-literal way under constrained contexts (Di Paola et al., 2020; Pouscoulous & Tomasello, 2020), robust non-literal interpretation of metaphors emerges only years later (Asch & Nerlove, 1960; Billow, 1975; Martin-Gonzalez et al., 2025). More broadly, standardized tests, like the Clinical Evaluation of Language Fundamentals – Fifth Edition, show the same patterns, with expected competence and performance continuing to increase into young adulthood across multiple linguistic domains (Wiig et al., 2013).

The nature and scope of this protracted learning remains poorly understood. While the available data makes clear that some aspects of language undergo protracted development, they are too coarse-grained to say exactly what, frequently lumping together many different linguistic phenomena (Chen & Hartshorne, 2021; Hartshorne et al., 2018; Wiig et al., 2013). Take, for example, the protracted development of syntactic competence. Hartshorne et al. (2018) approached this question by averaging over knowledge of passivization, clefting, agreement, relative clauses, preposition use, verb syntactic subcategorization, pronoun gender and case, modals, determiners, subject dropping, aspect, sequence of tenses, and wh movement. Different syntactic processing theories treat these phenomena differently, and it is entirely possible that protracted development is specific to some syntactic processes and not others. While this example concerns the observation of protracted development of syntax, the ideas apply more broadly.

Once we know exactly what undergoes protracted development, there remains the question of why. Existing theories offer little guidance, remaining ambivalent as to whether protracted development would result from gradual accumulation of world

https://doi.org/10.1017/S0305000926100877 Published online by Cambridge University Press

knowledge, refinement or strengthening of linguistic representations, improvements in memory recall or cognitive inhibition, or other mechanisms entirely. Theories of general cognitive development are of little help, primarily because they have been invalidated. In the previous century, it was believed that some abilities peak in late adolescence and rapidly decline ("fluid intelligence"), while others peak in middle age and begin to decline in the 50s ("crystalized intelligence") (Doppelt & Wallace, 1955; Salthouse, 2004). This simple empirical picture was explained in terms of a dual-mechanism theory: Fluid intelligence tasks were those that relied more heavily on processing speed (which was assumed to decline across the lifespan), while crystalized intelligence task performance relied more heavily on accumulated knowledge (which continues to accumulate indefinitely). Age-related decline in crystalized intelligence occurs when processing speed has dropped so far that accumulation of knowledge cannot compensate.

This fluid/crystalized intelligence dichotomy never fit what we knew about language, and it turns out it does not explain non-linguistic cognition, either. The theory was based on a coarse-grained picture of a half-dozen cognitive tasks. Recent years have revealed a much more complex picture (Bleidorn et al., 2013; Erb et al., 2023; Germine et al., 2011; Hampshire et al., 2012; Hartshorne & Germine, 2015; Logie & Maylor, 2009; Maylor & Logie, 2010). For instance, face memory peaks at around 30 before an immediate decline (Germine et al., 2011). In contrast, working memory peaks at around 25 and remains stable for a decade (Hartshorne & Germine, 2015). Some measures of social intelligence peak in middle age but show little subsequent decline (e.g., Hartshorne & Germine, 2015). Even different aspects of cognitive control show different lifespan trajectories (Erb et al., 2023). Unfortunately, no alternative theory of cognitive development has yet emerged (see, for instance, Germine et al., 2011; Hampshire et al., 2012; Hartshorne & Germine, 2015), as researchers have been focused on first determining empirically what such a theory should have to explain.

Thus, existing theories, whether of language acquisition or of cognitive development more broadly, either provide no explanation for our increasingly complex picture of language development, or they make the wrong predictions. The biggest need right now is more data, to better clarify the empirical picture. Unfortunately, increases in performance after 5 are often numerically small compared to the dramatic improvements in the first few years of life, making later learning empirically difficult to measure, despite representing meaningful change in linguistic competence (for related discussion, see Nippold, 2007). To see the importance of "small" improvements, consider that someone who did not begin learning English until their 20s is easily distinguishable from a native speaker, even though their grammatical proficiency still compares favourably to that of a native-speaking 10-year-old (see figures in Hartshorne et al., 2018). That native speakers detect even subtle differences such as these indicates that these variations represent meaningful development in linguistic proficiency. To give another example, over-regularizations such as *runned* or *breaked* occur in fewer than 1% of preschoolers' utterances, yet these statistically rare errors are salient, frequently remarked upon characteristics of child speech (Marcus et al., 1992). In short, an extremely high level of performance is a fundamental characteristic of mature human linguistic capacity, but these smaller changes are difficult to detect in typically sized studies (20–50 participants per cell, with sample size usually based on a heuristic approach; Lakens, 2022), which have large error bars.

Yet another empirical challenge concerns our limited ability to test a wide range of ages for any given linguistic phenomenon. Studies in this area typically rely on cross-sectional comparisons of discrete age groups (e.g., younger vs. older kids or younger vs. older

https://doi.org/10.1017/S0305000926100877 Published online by Cambridge University Press

adults) rather than sampling more densely across a continuous range of development. This leaves a lot unknown, most notably what happens in between those age groups. For instance, if 18-year-olds score higher than 5-year-olds, we do not know whether learners reach ceiling at 6, at 18, or even later in adulthood. It is even logically possible that the 18-year-olds are past the peak, showing age-related decline (outside of language, such phenomena have been observed; Germine et al., 2011). In other words, our current understanding of the shape and timing of language learning beyond early childhood remains quite vague.

These observations highlight two inter-related open questions, which motivate the current study. First, how wide is this late-learning trajectory phenomenon? We know very little about which aspects of language continue to develop beyond early childhood; perhaps protracted development is the rule rather than the exception. Second, how do we characterize the learning processes that are active during these periods of small but significant change?

These questions are much larger than a single study, or even a single dozen. In this article, we push the ball forward with a large-scale investigation of a well-characterized, theoretically important phenomenon involving third-person singular pronoun comprehension. In contrast to other domains of language (such as syntax), later in development discourse-level learning remains particularly unexplored.[1] Examining pronoun comprehension allows us to ask about longer term developmental trajectories in discourse-level learning. This is because pronoun comprehension involves high-level inferences driven by the interplay of syntactic, semantic, and pragmatic discourse factors. As Rohde (2019) aptly puts it, "pronouns are the fruit flies of discourse…their interpretation reflects the deeper semantic meaning of a passage and can thus provide a window into the forces at play in establishing meaning across clauses" (p. 1). Thus, pronouns provide insight into how comprehenders find meaning across connected sequences in discourse and help to fill an important gap in current research on late language learning.

We specifically focus on pronoun interpretation biases, and on implicit causality and implicit consequentiality in particular (described next). There are several advantages to this approach: First, pronoun interpretation biases are easily quantifiable and appear robust across many different age groups (Arnold et al., 2007; Arnold et al., 2019; Francey & Cain, 2023; Song & Fisher, 2005, 2007; Yuile, 2022). Pronoun biases are additionally sensitive to small changes in language knowledge (Arnold et al., 2018; Arnold et al., 2019; Johnson & Arnold, 2021). Finally, among the many proposed influences on pronoun interpretation, implicit causality and consequentiality biases are particularly well attested and lie at the center of an ongoing theoretical debate about the nature and origin of pronoun biases (Guan & Arnold, 2021; Hartshorne & Snedeker, 2013; Hartshorne et al., 2015b; Solstad & Bott, 2022). Together, these factors make developmental comparisons of these particular biases especially informative.

In the remainder of this introduction, we first briefly review the implicit causality and consequentiality pronoun biases. We then review the developmental data that have been reported to date. Finally, we review proposed mechanisms of pronoun comprehension. Although our data will not distinguish between theories of pronoun comprehension, consideration of these theories helps contextualize the current project, and we return to implications of our findings for theory in the Discussion.

[1]While there is some work on discourse-level learning beyond age 5 (see, e.g., Currie & Cain, 2023; Lee et al., 2022; Paris & Paris, 2003), this work focuses on the production of narratives or inference generation rather than coreference.

https://doi.org/10.1017/S0305000926100877 Published online by Cambridge University Press

1.1. Implicit causality and consequentiality pronoun biases

Pronouns, such as *he*, *she*, or *they*, present a fundamental challenge for language comprehension. Formally, they are ambiguous and can refer to any entity of the corresponding gender. Nonetheless, people rarely seem to notice the ambiguity, resolving the reference quickly and without fuss. Researchers have identified a number of contextual factors that people rely on, including order-of-mention (people often expect pronouns to refer to the first noun phrase in the previous sentence; Kaiser & Trueswell, 2008) and grammatical role (people also often expect pronouns to refer to the subject of the previous sentence, independent of whether it is first mentioned; Arnold, 1998). In this article, we focus on two related semantics-based influences that have captured researchers' attention for over half a century: implicit causality and implicit consequentiality (*IC biases*). We note that this work is distinct from the "implicit causality" studied in social attribution research, which shares the name but is unrelated to pronoun interpretation (see Hartshorne, 2014, for extended discussion).

Implicit causality refers to a pronoun interpretation bias triggered by certain verbs that affects how comprehenders interpret ambiguous pronouns in the context of explanations. For example, comprehenders usually think that the pronoun refers to Sally in (1), but to Mary in (2).

(1) Sally frightened Mary because she… [she = Sally]
(2) Sally feared Mary because she… [she = Mary]

Implicit consequentiality is a similar phenomenon in which comprehenders usually assign ambiguous pronouns to the entity experiencing the consequence of the event. For example, in (3) comprehenders most often think that the pronoun refers to Mary, while in (4) they typically think that the pronoun refers to Sally.

(3) Because Sally frightened Mary, she… [she = Mary]
(4) Because Sally feared Mary, she… [she = Sally]

Note that while in these cases *fear* and *frighten* have different and opposite biases in implicit causality and consequentiality, which is not true of all verbs (Hartshorne et al., 2015b).

1.1.1. Developmental evidence on pronoun biases

Implicit causality and consequentiality pronoun biases are widely studied in adults, with research consistently demonstrating that these biases are common among mature language users and arise from the interplay between verb semantics and discourse structure (Garvey & Caramazza, 1974; Guan & Arnold, 2021; Garvey et al., 1974; Hartshorne & Snedeker, 2013; Hartshorne, 2014; Hartshorne et al., 2015b; Johnson & Arnold, 2021; Järvikivi et al., 2017; Koornneef & Sanders, 2013; Koornneef & Vanberkum, 2006; Lyu & Wang, 2022; Pickering & Majid, 2007; Solstad & Bott, 2022; van den Hoven & Ferstl, 2018; Weatherford & Arnold, 2021; Ye & Arnold, 2025). In contrast, when and how exactly these biases are acquired remains unclear and minimally investigated, though there are some tantalizing hints and theoretical considerations.

For example, individual differences studies suggest room for a prolonged learning trajectory. Less skilled comprehenders are slower to incorporate implicit causality in

https://doi.org/10.1017/S0305000926100877 Published online by Cambridge University Press

pronoun interpretation (Long & De Ley, 2000), and second language learners show weaker implicit causality biases than native speakers (Wang & Gabriele, 2023). Among 7- to 11-year-olds, comprehension skill predicts anticipatory eye movements to pronoun referents (Engelen et al., 2014). Although the studies earlier were designed to test development directly, these findings are consistent with the possibility that pronoun biases strengthen as language skills mature in later childhood and adolescence.

Further reason to suspect extended developmental trajectories for implicit causality and consequentiality pronoun biases comes from developmental studies of other pronoun biases. For example, children become faster at using a different pronoun bias (order of mention) between the toddler and preschool years, though 5-year-olds are still noticeably slower than young adults (Hartshorne et al., 2015a). More relevant to implicit causality and consequentiality, which involve semantic processing, Francey and Cain (2023) explored changes in sensitivity to aspect – grammatical and lexical markers that indicate whether an event is ongoing or complete – on pronoun interpretation. They found increasing sensitivity to the effects of lexical aspect on pronoun interpretation across three age groups (7–11, 13–14, and 18–23 years). Sensitivity to grammatical aspect also increased with age, but showed no clear difference between the 13- to 14- and 18- to 23-year-old groups, potentially suggesting learning is complete by early adolescence. However, given that the study had low power for detecting an interaction and that the age groups are broad and inconsistently spaced, some caution should be taken in interpreting a null effect as evidence against continued learning. In any case, these findings at a minimum suggest development in semantic pronoun interpretation biases beyond age 5.

Several studies provide preliminary direct evidence that implicit causality and consequentiality pronoun biases may change across the lifespan, but the nature and direction of these changes remains unclear. Bittner (2019) compared younger and older adults and found that younger adults had more robust implicit causality biases than did older adults, but only for a small subset of the verbs tested, suggesting a possible decline in this bias in older adults within specific contexts. However, in light of the large number of comparisons the authors made without correction, these results provide only tentative evidence for age-related change. Blything et al. (2022) examined 7- to 10-year-olds' use of order of mention, grammatical role, and implicit causality when interpreting pronouns. Children prioritized order of mention and implicit causality over grammatical role. Using the same materials in a comparable adult sample, Schumacher et al. (2017) found that adults prioritized implicit causality over both order of mention and grammatical role. Together, these findings suggest developmental shifts in how semantic information is used to support pronoun interpretation from middle childhood to adulthood, and provide clues that implicit causality and consequentiality biases continue to develop beyond age 5.

To summarize, the literature offers some preliminary albeit inconclusive evidence about whether pronoun interpretation based on semantic cues, such as implicit causality and consequentiality, may continue to develop beyond age 5. The available studies test discrete age groups rather than a continuous developmental span, making it difficult to characterize the shape or timing of any developmental trajectory. What is more, while some findings suggest further strengthening into adulthood (Blything et al., 2022; Schumacher et al., 2017), others suggest stability from middle childhood (Francey & Cain, 2023; Garvey et al., 1974), and still others point to possible decline later in life (Bittner, 2019). As such, it remains an open question whether these pronoun interpretation biases emerge early and remain constant, or shift across the lifespan.

https://doi.org/10.1017/S0305000926100877 Published online by Cambridge University Press

1.1.2. Theoretical accounts of pronoun biases and implications for protracted development

While there is disagreement about exactly what causes semantic pronoun biases, all the major theories are consistent with an extended learning trajectory. In this section, we briefly review theories of pronoun comprehension to further support our investigation into late-learning trajectories of implicit causality and consequentiality pronoun biases. We note, however, that the current study is not positioned to tease apart these theories, in part, because none make sufficiently specific predictions about learning trajectories.

Several sources of implicit causality and consequentiality have been proposed. *World knowledge* accounts argue that pronoun biases emerge as a consequence of creating semantically rich conceptual representations of the events described in the discourse, including predicate-argument structures as well as the temporal, causal, and spatial relations within the event (Pickering & Majid, 2007; see also situation models: Zwaan & Radvansky, 1998). Implicit causality and consequentiality biases, along this account, are attributable to higher order inferences about the likely causes and consequences of events grounded in background knowledge about how the world works. These accounts are consistent with protracted learning to the extent that the relevant world knowledge continues to be acquired or refined with experience. To make these ideas concrete, consider the example in (5) adapted from Winograd (1972):

(5) The city council denied the demonstrators a permit because they…

In this example, correctly identifying the referent of "they" depends on inferring who was more likely to have caused the denial, the city council or the protesters. Making this inference involves knowledge about political institutions, the dynamics of protest movements, and the typical sources of conflict in such situations – representations that are usually refined with age. Thus, to the extent that pronoun biases rely on knowledge about how the world works, developmental change is likely.

Semantic structure accounts (Hartshorne et al., 2015b; see also Solstad & Bott, 2022), in contrast, argue that implicit causality and consequentiality biases do not involve inferences about the world but rather derive from the literal meaning of the sentence. These accounts draw on well-developed and articulated linguistic theories of semantics and discourse, which we will not do justice to here. Briefly, it is widely accepted that sentence structure is organized around core semantic notions of causes, affected entities, and so forth Thus, comprehenders need not infer who the cause of an event was or who was affected by it, because that is often explicitly entailed by the verb. What aspects of the sentence meaning are relevant depends on the discourse relation between the two clauses (for in depth discussion of discourse structure, see Kehler, 2002). In (1–2), the second clause explains the first, and thus the cause of the first event highlighted by the semantics is likely to be mentioned. In (3–4), the second clause describes a result of the first, and so the entity specifically highlighted as being affected by the first event is likely to be mentioned in the second.

For development, this means that implicit causality and consequentiality biases emerge and change in concert with the relevant linguistic competences. While these representations are available to some degree early in acquisition, semantics in particular may continue to be refined with experience, potentially leading to gradual strengthening or adjustment of pronoun biases over time (akin to developmental increases in semantic

https://doi.org/10.1017/S0305000926100877 Published online by Cambridge University Press

specificity observed for nouns in infancy; see, e.g., Bergelson & Aslin, 2017). Using example (5) from before, one would expect change insofar as representations of the argument structure of "deny" changes. Continued development could also come from discourse processing. The biases depend on identifying the correct discourse structure, and currently there is no theory as to how this is done, much less how the ability emerges in development. Thus, within the limits of current knowledge, semantic structure accounts are consistent with a protracted developmental trajectory.

Finally, *statistical* accounts (Arnold, 1998, 2001; Arnold et al., 2018) argue that pronoun biases reflect probabilistic learning over patterns of co-reference observed in the linguistic input. This can involve learning, for instance, that pronouns often co-refer with the subject of the previous clause. It can also involve learning patterns of co-reference involving the semantic and discourse structures invoked by the semantic structure account; for instance, that the subject of an explanation clause tends to refer to the cause of the event in the previous clause. At the same time – and in contrast to the semantic structure account – if the statistics in the input supported it, one could learn that explanations tend to refer to affected entities while results refer to causes.

In (5), on a linguistic heuristics account, pronoun bias is predicted to change as comprehenders get more examples of sentences involving explanations, denying, city councils, and protesters. In principle, this could continue to be refined throughout the lifetime, depending on how often learners encounter relevant examples and how noisy the input is. (In principle, this could be calculated, but currently there are no relevant corpora of sufficient size.)

To summarize, all accounts are consistent with continued learning past early childhood, but none require it. As noted earlier, the ability of theories to make specific predictions about how long learning persists depends on ancillary factors: measurements about changes in children's world knowledge or semantic representations, assumptions about learning rates and accumulation of evidence, and so forth. Conversely, though, measuring the learning trajectory could provide important constraints on all three accounts (see Discussion).

1.2 The current study

In the analyses that follow, we re-analyse two large, publicly available datasets of implicit causality and implicit consequentiality pronoun biases in native English speakers across a broad (cross-sectional) age range. Previewing our findings, we observe that these pronoun interpretation biases continue to develop into early adulthood, with subtle differences in implicit causality and consequentiality effects. The first dataset comes from Hartshorne and Snedeker (2013, Experiment 2; osf.io/2f763/files/osfstorage) and will be referred to as Dataset 1. The second dataset, referred to as Dataset 2, comes from Hartshorne et al. (2015b; at osf.io/fw437/files/osfstorage). Dataset 1 includes data on pronoun interpretation in contexts that give rise to an implicit causality effect (ICaus A). Dataset 2 includes both implicit causality (ICaus B) and implicit consequentiality (ICons) pronoun biases.

These datasets offer several important advantages for investigating developmental change. First, both span a wide and continuous age range, from middle childhood through late adulthood. As reviewed earlier, prior work suggests potential for developmental shifts in implicit causality and consequentiality pronoun biases from childhood to adulthood. However, because previous studies have not examined a continuous and broad age range, the precise nature of these shifts remains unclear. The current datasets, by including a

https://doi.org/10.1017/S0305000926100877 Published online by Cambridge University Press

broad (albeit cross-sectional) range of ages, enable a more nuanced examination of how these biases unfold across the lifespan. No other large datasets include children.

Second, the scale and design of these datasets supports both replicability and generalizability. The smallest contains nearly a 1000 participants, providing sufficient statistical power to detect subtle developmental trends. In addition, both datasets test an unusually large set of verbs, allowing for more robust generalization of observed effects beyond a narrow set of experimental items.

2. Methods

2.1. Participants

Dataset 1 (ICaus A). A total of 978 native English speakers from 9 to 81 years of age from Hartshorne and Snedeker (2013) are included in Dataset 1 (*M* age in years = 30, SD = 12, 637 female). Additional participants were tested but excluded due to reporting that English was not their native language (127), reporting implausible ages (4), or not completing all trials (118). See Figure 1 for distributions of participants by age.

Dataset 2 (ICaus B, ICons). A total of 2,648 native English speakers from 8 to 81 years of age from Hartshorne et al. (2015b) are included in Dataset 2 (*M* age in years = 30, SD = 14, 1,623 female). Additional participants were tested but excluded due to reporting that English was not their native language (849), reporting implausible ages (10), or not completing all trials and/or failing more than 1 "catch" trial (190). See Figure 1 for distributions of participants by age.

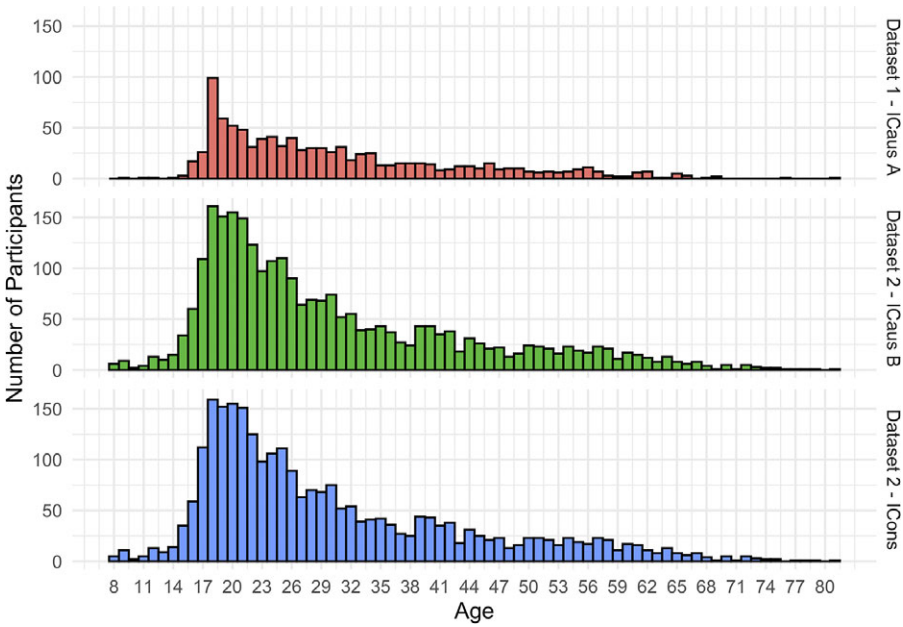


Figure 1. Count of participants at each age by dataset.

https://doi.org/10.1017/S0305000926100877 Published online by Cambridge University Press

2.2. Materials and procedures

Because full methodological details are available in the original publications, here we only summarize immediately relevant aspects of the materials and procedures.

Dataset 1 (ICaus A). In Dataset 1, participants read sentences of the following form:

(6) Sally VERBed Mary, because she is a dax.
Who do you think is the dax?
Sally Mary

Each participant was tested on 25 verbs out of a total of 265 verbs (responses per verb: $M = 88.57$, range = 1–297). All verbs were transitive emotion verbs, which as a class give rise to strong implicit causality biases. Participants were uncompensated volunteers who completed the experiment online through the Games with Words platform.

Dataset 2 (ICaus B, ICons). The materials and procedure for Dataset 2 matched the procedure of Dataset 1 except in the following ways. Hartshorne et al. (2015) tested two discourse structures – implicit causality and implicit consequentiality – by moving "because":

(7) Implicit *Causality*: Sally VERBed Mary, because she daxed.
(8) *Implicit Consequentiality*: Because Sally VERBed Mary, she daxed.

Note that the dependent clause included a novel verb instead of a novel noun – something that Hartshorne and colleagues (2015) expected not to matter but proved to systematically shift the biases. Each participant read a total of 14 implicit causality sentences and 14 implicit consequentiality sentences, with verbs chosen from a set of 502 (responses per verb: $M = 134.28$, range = 106–170). These verbs covered a range of semantic domains, including transitive emotion verbs (for complete list and motivation, see Hartshorne et al., 2015).

2.3. Outcome measure

Our goal in this analysis is to capture potential changes in how participants converge on shared interpretations for pronouns in implicit causality and consequentiality contexts. From this perspective, the direction of the bias (subject versus object) is not central. Thus, we code each response in terms of whether it matched the majority response for that item (restricted to 18–30-year-olds). For convenience, we label such responses as "correct." No strong theoretical claims are intended by this terminology. Note that analysing subject- and object-biased verbs separately (requested by a reviewer) did not reveal any meaningful difference in developmental trajectories, so these results are relegated to the Supplementary Materials.

Table 1 describes the observed response accuracy (proportion of responses selecting modal referent) averaged across all included participants in each dataset. In each dataset, a small minority of verbs were equi-biased and so are not included in our analyses (Dataset 1 [ICaus A]: $N = 10$, 4%; Dataset 2 [ICaus B, Icons]: $N = 17$, 3%). However, we retained items with even small biases because they may still contain information (some people may be sensitive to even these slight tendencies). At worst, these items will increase noise, and given our large dataset in both participants and items, noise is not a major concern.

https://doi.org/10.1017/S0305000926100877 Published online by Cambridge University Press

Table 1. Average pronoun interpretation accuracy

Effect	Mean	SD	Min	Max
ICaus A	0.75	0.16	0.08	1.00
ICaus B	0.67	0.23	0.00	1.00
ICons	0.75	0.18	0.00	1.00

Because participants completed different (though partially overlapping) sets of items in each experiment, we fit a two-parameter Bayesian item response theory (IRT) model to each dataset and discourse structure (i.e., three IRT models: ICausA, ICaus B, ICons). These IRT models provide an estimate of latent ability for each participant, normalizing for differences in item difficulty across subjects (Burkner, 2021). While not commonly used in psychology, it is increasingly recommended (Frank, 2018). Participant ability estimates from the IRT model were used as the outcome measure in the main regression models. The code for each IRT model can be found in the reproducible manuscript.

3. Results

3.1. Approach to analysis

Full details and code are available in the reproducible manuscript (https://osf.io/5xdj2/overview?view_only=6d3798ef6a614f6f92e8fbc94994f658). Our approach to analysis closely follows Erb et al. (2023). That is, to explore age-related changes in pronoun interpretation for each dataset and discourse structure (ICaus A, ICaus B, ICons), we fit separate Bayesian regression models using the brms package in R, with a thin-plate spline term to model non-linear effects of age (Burkner, 2017). In all models, we used default priors and ran four chains, each with 4,000 samples, half of which were reserved as warm-up and not included in results. We then calculated the local slope of the curve at each 1/100th of a year. This procedure was repeated for each sample, providing an estimate of uncertainty – 95% credible interval (CI) – at each interval of age. Each age interval in which the CI did not overlap with 0 represents significant developmental change.

For each model, we plot the estimated latent ability (Figure 2) and slope (Figure 3) across age, highlighting the regions with significant developmental change. While all available ages were included in the Bayesian regression models, samples were small and thus error bars are extremely large at the youngest and oldest ages. In the main text, we limit figures and discussion to the age range with reasonable error bars. Specifically, we start at the youngest age for which there is a significant slope and use 65 as the upper age limit. Figures depicting results across the full age range can be found in Appendix.

3.2. ICaus A

In ICaus A, we observed significant developmental change in pronoun interpretation starting at approximately age 15 and lasting until approximately age 26. Across these ages, performance increased. In other words, participants' responses were more variable at younger ages, but became more uniform as age increased up until the mid-20s.

https://doi.org/10.1017/S0305000926100877 Published online by Cambridge University Press

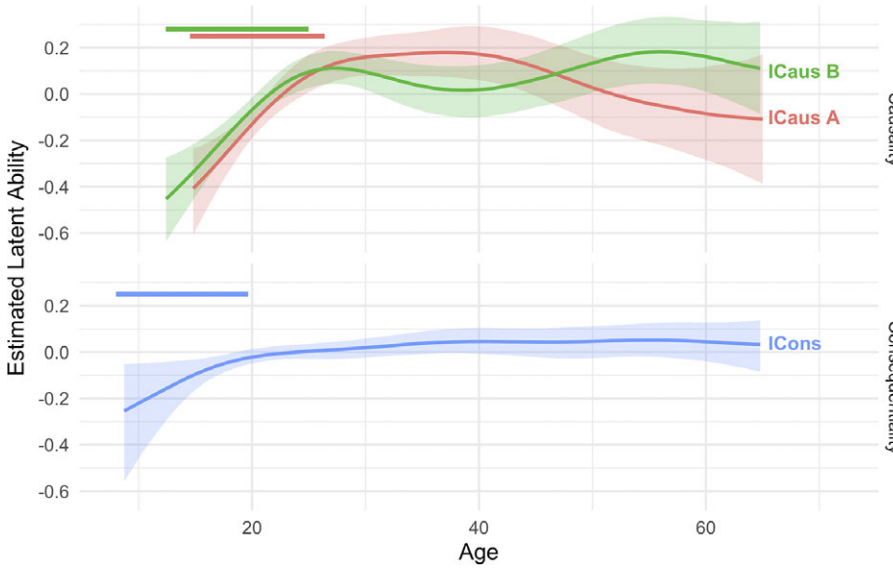


Figure 2. Estimated latent ability across age. Shading represents 95% credible intervals. Horizontal bars show age ranges where the estimated change in ability was significant. For the slope estimates themselves, see Figure 3.

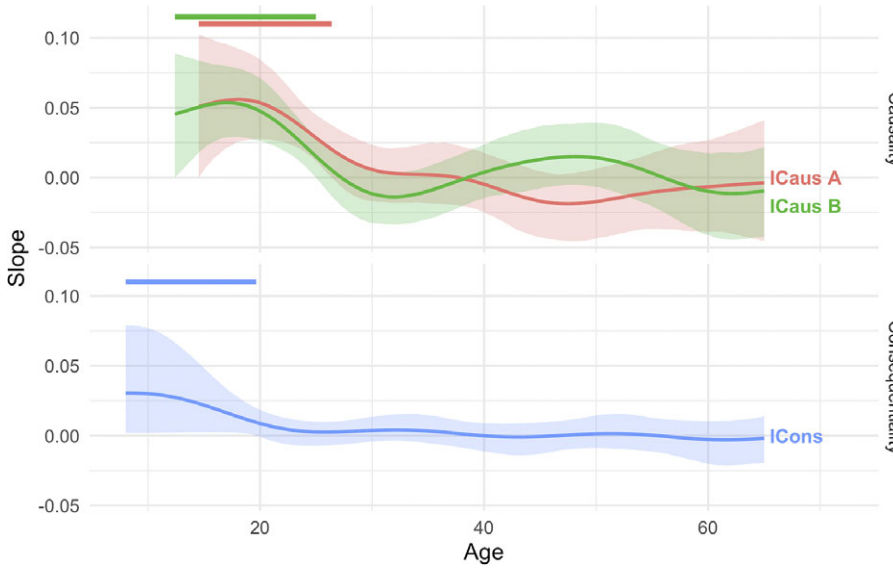


Figure 3. Estimated slope across age. Shading represents 95% confidence intervals. Horizontal bars show age ranges where the estimated change in ability was significant.

The youngest age for which we have data is 9. While we do not observe a significant increase in age starting until 15, the lack of significance at earlier ages seems to be driven by high uncertainty (see Appendix) due to the fact we have only four children under 15 years old.

https://doi.org/10.1017/S0305000926100877 Published online by Cambridge University Press

3.3. ICaus B

In ICaus B, we observed significant developmental change in pronoun interpretation from approximately ages 12 to 25. During this period, performance increased. Thus, as in ICaus A, participants' responses were more variable at younger ages, but became more uniform as age increased up until the mid-20s.

In this dataset, our youngest participants are 8 years old, thus, we did not observe significant development in pronoun bias at the youngest end of the age range. As before, this appears to be due to high uncertainty (see Appendix), again due to the small sample at these ages ($N = 21$). Note, however, that overall accuracy at age 8 is low (48%, CI = [31%, 65%]) enough that we cannot rule out the possibility that children younger than this age have not acquired implicit causality, or alternatively have not acquired enough of the verbs in our sample to exhibit implicit causality on them.

3.4. ICons

In ICons, we observed significant developmental change in pronoun interpretation from the earliest age in our dataset (8) to age 20. During this period, performance increased. Thus, as was observed in Datasets 1 and 2a which both contain implicit causality sentences, in the current dataset, pronoun interpretation strategies were more varied at younger ages, but became increasingly more uniform across age. However, unlike in implicit causality sentences (ICaus A and ICaus B), performance stopped increasing significantly at an earlier age (before 20 years old).

3.5. Comparing rate of change in slope across datasets and discourse structures

To assess whether slope increased at a similar rate across datasets, we averaged the slope values across age ranges showing significant developmental change (i.e., where credible intervals excluded zero). The resulting averages were 0.05 for ICaus A, 0.04 for ICaus B, and 0.02 for ICons. This suggests steeper change for implicit causality than for implicit consequentiality. However, these comparisons should be interpreted with caution, as the datasets differ in participants and items. We provide a fuller discussion of potential differences between implicit causality and consequentiality patterns as well as the limitations of this comparison in the Discussion section next.

3.6. Increasing bias versus matching bias

Our analysis earlier focuses on whether people's biases become stronger with age. A reviewer raised a compelling question: do learners overshoot? That is, do learners converge on the population-level bias, or do they become more biased than the typical speaker? This is reminiscent of a common issue in the decision-making literature: although choosing the most likely answer (probability maximization) is the optimal strategy if what one wants is to be correct as often as possible, humans often probability match instead, choosing an answer in proportion to how likely it is to be correct (Herrnstein, 1997). If development of IC biases is targeting probability maximization, learners should overshoot. If it is targeting probability matching, learners should converge on the population-level bias.

To examine this question, we estimated the population-level bias by computing a weighted population-level response distribution for each verb using US Census

https://doi.org/10.1017/S0305000926100877 Published online by Cambridge University Press

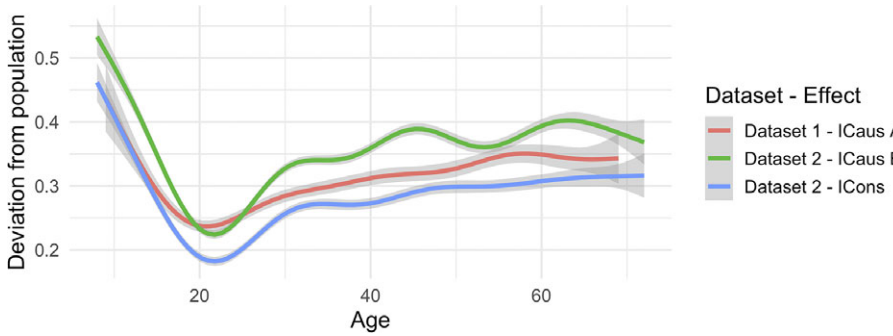


Figure 4. Deviation from the population-level verb bias distribution as a function of age, shown separately for each dataset and effect type (Dataset 1 – ICaus A; Dataset 2 – ICaus B; Dataset 2 – ICons). Lines reflect generalized additive model (GAM) fits; shaded regions indicate 95% confidence intervals.

age-distribution data, weighting individual participants' contributions proportionally to their age group's representation in the United States population. For each verb at each participant age, we calculated the proportion of subject-biased responses and measured the absolute deviation of this age-specific proportion from the corresponding population-level proportion. We modelled the relationship between age and deviation using generalized additive models (GAMs), fitted separately for each dataset (Dataset 1 – ICaus A; Dataset 2 – ICaus B; Dataset 2 – ICons).

Across all three datasets, deviation from the population-level distribution decreases with age over the period of significant developmental change identified in the main analysis, indicating that development targets probability matching rather than probability maximization (Figure 4).

4. Discussion

We asked whether implicit causality and consequentiality pronoun biases experience a late-learning trajectory. Across all datasets, we observed that these pronoun biases strengthened into early adulthood, with some subtle differences between implicit causality and consequentiality, which we discuss next. This fact raises challenges for current theories of pronoun interpretation and of language acquisition more broadly.

In considering the theoretical implications of this work, several limitations in our method should be taken into account. First, all datasets were cross-sectional. While we observed clear trends with regard to pronoun bias across age, it is always possible that these trends reflect cohort differences as opposed to age-related changes. These cohort differences could include differences in familiarity or comfort with technology, which is relevant in the current context as the experiments were all delivered using an online platform (Games with Words). However, previous research in other domains which directly compared online to offline task results across a broad range of ages found no evidence for cohort effects, giving us reason to think this concern is at most very minor (Hartshorne & Germine, 2015). Additionally, the sample of younger children was limited (see Figure 1), which restricts our ability to characterize the early stages of development. Notably, the pronoun interpretation task, while widely used, involves stimuli presented in isolation rather than embedded in discourse and a context lacking a communicative goal.

https://doi.org/10.1017/S0305000926100877 Published online by Cambridge University Press

These constraints, though common in this literature, may limit generalizability and, in particular, underestimate the influence of contextual cues that support pronoun resolution in more dynamic settings (for a related discussion, see Zerkle et al., 2017). Finally, our data are restricted to English. The degree to which implicit causality and consequentiality biases differ across languages remains debated and unclear (Fedele et al., 2015; Hartshorne et al., 2013; Ozge et al., 2025; Román, 2020; Ueno & Kehler, 2016), and thus caution should be taken in generalizing these results to other languages.

Alongside these methodological limitations, it is worth considering the consequences of our choice to use a large, broadly representative set of verbs.[2] One consequence is that verbs vary widely in bias, with some having only very weak biases. This raises the question of whether our developmental results are specific to either strongly- or weakly biased verbs. However, exploratory analyses showed similar trajectories for both (see Supplementary Online Materials). Similarly, the verbs varied substantially in frequency, with some being extremely rare and unlikely to be known by children (e.g., *niggle at*). Perhaps age-related differences in implicit causality and consequentiality biases merely reflected growth in vocabulary. This is unlikely, given that the developmental trajectories here do not match those seen for vocabulary, which shows continued development into the 60s (Hartshorne & Germine, 2015). In any case, exploratory analyses showed that the developmental trajectory of IC bias was the same for both high- and low-frequency verbs, suggesting continued learning even for verbs that are acquired early in childhood.

With these considerations in mind, next we begin by discussing two aspects of our results in more detail: the apparent differences between implicit causality and consequentiality, and the apparent lack of age-related decline in either pronoun bias. We then step back and consider the broader implications of our findings for theories of pronoun interpretation and of language acquisition.

4.1. Comparing learning trajectories for different IC biases

As noted earlier, we found subtle statistical differences between the development of implicit causality and consequentiality. Specifically, our earliest evidence of development is later for implicit causality (age 12 vs. 8) and learning peters out later for implicit causality as well (mid-20s rather than 20). The most parsimonious explanation for the differences in early development is simply low statistical power, as we have relatively few younger subjects (Figure 1). For instance, only thirty-seven 8- to 12-year-olds in Dataset 2 (ICaus B, ICons). Because of this, we do not consider this early difference further.

The later end to learning for implicit causality is comparatively more robust, as we have no shortage of subjects in their 20s. Moreover, continued learning into the mid-20s for implicit causality is replicated across ICaus A (Dataset 1) and ICaus B (Dataset 2), despite some differences in method.[3] One potential explanation for this pattern of results is that our measures of implicit causality are more sensitive or that the implicit consequentiality stimuli were subject to more robust ceiling effects. Some suggestive evidence comes from the fact that the range of inferred abilities is much wider for implicit causality than consequentiality. Note, however, that it is difficult to disentangle differences in

[2]We thank an anonymous reviewer for calling these considerations to our attention.

[3]Note also that even within Dataset 2, the implicit causality and implicit consequentiality analyses involved slightly different participant groups based on trial inclusion criteria (N = 2,616 for implicit causality, N = 2,621 for implicit consequentiality).]

https://doi.org/10.1017/S0305000926100877 Published online by Cambridge University Press

measurement sensitivity from actual differences in the underlying constructs. Therefore, systematic psychometric study is necessary to tease these possibilities apart.

If, however, the difference is psychologically real, one possibility is that implicit causality is simply harder to learn. Depending on how acquisition of pronoun interpretation works (see also the *Theoretical implications* section next), this could be because there are simply fewer relevant examples to learn from or because the underlying representations (e.g., discourse structure, event structure, or argument structure semantics) take longer to learn. With regard to the latter point, it seems plausible that consequences are conceptually simpler to represent than causes are. That is, causal representations of events encode interactions among event participants and include inferences about the intentions or underlying mental states of those participants. In contrast, consequences may be inferred from simpler associations between event sequences. This possibility suggests that future work that more explicitly manipulates the complexity of causal structures within events could help provide insight with regard to these observed differences.

4.2. Age-related decline in IC bias

We find it notable that we did not observe evidence for a decline in pronoun bias sensitivity in adulthood through age 65 (we lack the statistical power to investigate after the age of 65; see Appendix). The error bars in Figure 3 reveal that we can only rule out large declines on similar order of magnitude to the improvements in late childhood. Note that this is despite the fact that our sample size compares favourably to typical studies of ageing.

If this lack of decline holds up to further study, what would it mean? Prior work suggests age-related declines in general discourse processing and elaborative inference generation (see Stine-Morrow & Radvansky, 2017 for review). Thus, a lack of decline in IC bias would suggest that the processes supporting implicit causality and consequentiality pronoun biases do not rely on such inferencing. Interestingly, Lee and Lai (2024) argue that older adults engage less in effortful referential inference such as processing gender agreement, but not more elaborative pronoun processing strategies. Our findings of stable pronoun bias in older adults may suggest that implicit causality and consequentiality biases are tightly coupled to verb-level representations and do not involve online elaborative inference, a possibility that is consistent with a semantic structure account (see our *Introduction* and the *Theoretical implications* section before).

4.3. Theoretical implications of protracted pronoun bias learning patterns

As mentioned earlier, these findings have important theoretical implications. While all accounts have a role for learning, what varies across accounts is what specifically individuals are learning or in what way the underlying representations are being refined. The present data cannot distinguish between any of the three theoretical accounts of pronoun bias reviewed in the *Introduction*, partly because theory is under-specified with regard to protracted developmental patterns. This pattern of theoretical under-specification is not unique to pronoun bias acquisition, but observed across sub-domains of psychological research, prompting calls for the development of newer or more complete theories (see e.g., Ambridge & Wagner, 2021). Nonetheless, we speculate here about the ways these findings may constrain theory in an effort to encourage future progress on these questions (for a general discussion of how psychological phenomena may constrain theory, see Eronen & Bringmann, 2021). In particular, we consider which

https://doi.org/10.1017/S0305000926100877 Published online by Cambridge University Press

aspects of language and event representation need to be considered in more detail to generate concrete and testable hypotheses regarding timelines of learning for each account.

On the world knowledge account (Koornneef & Vanberkum, 2006; Pickering & Majid, 2007), pronoun biases result from inferences about who was most causally associated with the event (for implicit causality) or most likely to be affected within the event (for implicit consequentiality), based on what one knows about how the world works. Changes in pronoun biases are driven by either one's knowledge of the world or one's reasoning abilities (to the extent the two can be distinguished). To account for our results, the world knowledge accounts would predict that people's beliefs about likely causes and consequences of events changes measurably into early adulthood and then stabilizes. Note that while world knowledge theories invoke a variety of representations that may be relevant (e.g., salience and focus), to date these have not been organized into a mechanistic theory of how world knowledge reasoning works or informs pronoun interpretation (for what such a mechanistic theory might look like, see Gerstenberg et al., 2021).

Alternatively, semantic structure accounts (Hartshorne et al., 2015b) posit that pronoun biases emerge from verb-specific semantic representations within supportive discourse contexts. Thus, developmental change results either from the refinement of the semantic representations of verbs, including how event participants map onto syntactic positions to determine who does what to whom, or to the representation of and ability to recognize discourse structure. Exactly how long semantic structure representations continue to develop is unknown, though in at least some cases it continues at least into the early school years (Hartshorne et al., 2023). For discourse, even less is known; as noted in the *Introduction*, there is currently no account of how people recognize discourse structures, much less how that ability develops. However, to account for our results, semantic structure accounts would predict that learning the semantic representations of the relevant verbs or development of discourse processing continues into early adulthood.

Finally, along statistical learning accounts (Arnold, 1998, 2001; Arnold et al., 2018), development reflects increasing exposure to relevant verbs and/or discourse structures. To account for our results, these accounts would predict that by early adulthood, people have encountered enough examples in the input that further learning does not meaningfully change their estimates. That is, learning becomes a sampling problem: As we know from the Law of Large Numbers, as your sample gets larger, your statistical estimates approach ground truth. However, generating precise predictions about the timeline of this learning requires more explicit assumptions about how these statistical patterns are tracked. For example, do learners track patterns of reference by word or by semantic role (for a more detailed discussion, see Ye & Arnold, 2023)? This question can also be posed another way: What would have to be true of human statistical learning mechanisms such that implicit causality and consequentiality biases stabilize in early adulthood? We know of only two prior corpus studies of implicit causality, which do not address these questions fully, but provide a useful starting point Guan and Arnold (2021).

Thus, each of these accounts makes distinct predictions about the conditions under which developmental change should occur, and future work may be able to leverage these differences to tease them apart and make finer-grained predictions about protracted learning trajectories. In the meantime, these results also provide further grist for developing a replacement to the now-defunct fluid/crystalized intelligence theory of cognitive development.

https://doi.org/10.1017/S0305000926100877 Published online by Cambridge University Press

4.4. Conclusion

We provide evidence from two large-scale datasets that sensitivity to implicit causality and consequentiality in pronoun interpretation increases into early adulthood. This implies that the representations that support implicit causality and consequentiality pronoun biases experience a protracted developmental trajectory. Further investigation is needed to determine whether this is attributable to accumulating event knowledge (Koornneef & Van Berkum, 2006), argument and discourse structure (Hartshorne & Snedeker, 2013), or probabilistic distributions of pronoun reference (Arnold, 1998).

Acknowledgements. We thank Steven Pinker and Jennifer Arnold for providing comments.

Supplementary material. The supplementary material for this article can be found at http://doi.org/10.1017/S0305000926100877.

Author contribution. The authors made the following contributions. Conceptualization: J.K.H.; formal analysis: A.R.Y.; supervision: J.K.H.; writing – original draft preparation: A.R.Y.; writing – review and editing: A.R.Y., J.K.H.

Competing interests. The authors declare none.

References

Ambridge, B., & **Lieven, E. V.** (2011). *Child language acquisition: Contrasting theoretical approaches*. Cambridge University Press. https://doi.org/10.1017/CBO9780511975073

Ambridge, B., & **Wagner, L.** (2021). Testable theories of core first language acquisition. *Journal of Child Language*, **48**(5), 859–861. https://doi.org/10.1017/S0305000921000581

Arnold, J. E. (1998). *Reference form and discourse patterns* (**PhD thesis**). Stanford University

Arnold, J. E. (2001). The effect of thematic roles on pronoun use and frequency of reference continuation. *Discourse Processes*, **31**(2), 137–162. https://doi.org/10.1207/S15326950DP3102_02

Arnold, J. E., **Brown-Schmidt, S.**, & **Trueswell, J.** (2007). Children's use of gender and order-of-mention during pronoun comprehension. *Language and Cognitive Processes*, **22**(4), 527–565. https://doi.org/10.1080/01690960600845950

Arnold, J. E., **Castro-Schilo, L.**, **Zerkle, S.**, & **Rao, L.** (2019). Print exposure predicts pronoun comprehension strategies in children. *Journal of Child Language*, **46**(5), 863–893. https://doi.org/10.1017/S0305000919000102

Arnold, J. E., **Stangmann, I.**, **Hwang, H.**, **Zerkle, S.**, & **Nappa, R.** (2018). Linguistic experience affects pronoun interpretation. *Journal of Memory and Language*, **102**, 41–54. https://doi.org/10.1016/j.jml.2018.05.002

Asch, S., & **Nerlove, H.** (1960). The development of double function terms in children: An exploratory investigation. In B. Kaplan & S. Wapner (Eds.), *Perspectives in psychological theory: Essays in honor of Heinz Werner* (pp. 47–60). International Universities Press.

Asendorpf, J. B., **Conner, M.**, **De Fruyt, F.**, **De Houwer, J.**, **Denissen, J. J. A.**, **Fiedler, K.**, **Fiedler, S.**, **Funder, D. C.**, **Kliegl, R.**, **Nosek, B. A.**, **Perugini, M.**, **Roberts, B. W.**, **Schmitt, M.**, **Van Aken, M. A. G.**, **Weber, H.**, & **Wicherts, J. M.** (2013). Recommendations for increasing replicability in psychology. *European Journal of Personality*, **27**(2), 108–119. https://doi.org/10.1002/per.1919

Bergelson, E., & **Aslin, R.** (2017). Semantic specificity in one-year-olds' word comprehension. *Language Learning and Development*, **13**(4), 481–501. https://doi.org/10.1080/15475441.2017.1324308

Berman, R. A. (2007). Developing linguistic knowledge and language use across adolescence. In E. Hoff & M. Shatz (Eds.), *Blackwell handbook of language development* (pp. 347–367). Blackwell Publishing Ltd. https://doi.org/10.1002/9780470757833.ch17

Billow, R. M. (1975). A cognitive developmental study of metaphor comprehension. *Developmental Psychology*, **11**(4), 415–423. https://doi.org/10.1037/h0076668

Bittner, D. (2019). Implicit causality in younger and older adults. *Linguistics Vanguard*, **5**(s2), 20180023. https://doi.org/10.1515/lingvan-2018-0023

https://doi.org/10.1017/S0305000926100877 Published online by Cambridge University Press

Bleidorn, W., **Klimstra, T. A.**, **Denissen, J. J.**, **Rentfrow, P. J.**, **Potter, J.**, & **Gosling, S. D.** (2013). Personality maturation around the world: A cross-cultural examination of social-investment theory. *Psychological Science*, **24**(12), 2530–2540. https://doi.org/10.1177/0956797613498396

Blything, L. P., **Iraola Azpiroz, M.**, **Allen, S.**, **Hert, R.**, & **Järvikivi, J.** (2022). The influence of prominence cues in 7- to 10-year-olds' pronoun resolution: Disentangling order of mention, grammatical role, and semantic role. *Journal of Child Language*, **49**(5), 930–958. https://doi.org/10.1017/S0305000921000349

Bowerman, M. (1994). From universal to language-specific in early grammatical development. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, **346**(1315), 37–45. https://doi.org/10.1098/rstb.1994.0126

Bowles, R. P., & **Salthouse, T. A.** (2008). Vocabulary test format and differential relations to age. *Psychology and Aging*, **23**(2), 366–376. https://doi.org/10.1037/0882-7974.23.2.366

Brown, R. (1973). *A first language: The early stages*. Harvard University Press

Burkner, P. (2017). Brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, **80**(1), 1–28. https://doi.org/10.18637/jss.v080.i01

Burkner, P. (2021). Bayesian item response modeling in R with brms and Stan. *Journal of Statistical Software*, **100**(5), 1–54. https://doi.org/10.18637/jss.v100.i05

Chen, T., & **Hartshorne, J. K.** (2021). More evidence from over 1.1 million subjects that the critical period for syntax closes in late adolescence. *Cognition*, **214**, 104706. https://doi.org/10.1016/j.cognition.2021.104706

Crain, S., & **Thornton, R.** (2012). Syntax acquisition. *WIREs Cognitive Science*, **3**(2), 185–203. https://doi.org/10.1002/wcs.1158

Currie, N. K., & **Cain, K.** (2023). Developmental differences in children's generation of knowledge-based inferences. *Discourse Processes*, **60**(6), 440–456. https://doi.org/10.1080/0163853X.2023.2225980

Di Paola, A., **Domaneschi, F.**, & **Pouscoulous, N.** (2020). Metaphorical developing minds: The role of multiple factors in the development of metaphor comprehension. *Journal of Pragmatics*, **156**, 235–251. https://doi.org/10.1016/j.pragma.2019.08.008

Doppelt, J. E., & **Wallace, W. L.** (1955). Standardization of the Wechsler adult intelligence scale for older persons. *The Journal of Abnormal and Social Psychology*, **51**(2), 312–330. https://doi.org/10.1037/h0044391

Engelen, J. A. A., **Bouwmeester, S.**, **De Bruin, A. B. H.**, & **Zwaan, R. A.** (2014). Eye movements reveal differences in children's referential processing during narrative comprehension. *Journal of Experimental Child Psychology*, **118**, 57–77. https://doi.org/10.1016/j.jecp.2013.09.005

Erb, C. D., **Germine, L.**, & **Hartshorne, J. K.** (2023). Cognitive control across the lifespan: Congruency effects reveal divergent developmental trajectories. *Journal of Experimental Psychology: General*, **152**(11), 3285–3291. https://doi.org/10.1037/xge0001429

Eronen, M. I., & **Bringmann, L. F.** (2021). The theory crisis in psychology: How to move forward. *Perspectives on Psychological Science*, **16**(4), 779–788. https://doi.org/10.1177/1745691620970586

Fedele, E., & **Kaiser, E.** (2015). Resolving null and overt pronouns in Italian: An experimental investigation of syntax-semantics interactions. In C. Brown, Q. Gu, C. Loos, J. Mielens, & G. Neveu (Eds.), *TLS: Proceedings of the 15th Texas linguistic society* (pp. 53–72). University of Texas

Fisher, C., **Jin, K.**, & **Scott, R. M.** (2020). The developmental origins of syntactic bootstrapping. *Topics in Cognitive Science*, **12**(1), 48–77. https://doi.org/10.1111/tops.12447

Francey, G., & **Cain, K.** (2023). The influence of the temporal characteristics of events on adults' and children's pronoun resolution. *Journal of Child Language*, **50**(2), 391–416. https://doi.org/10.1017/S0305000921000878

Frank, M. C. (2018). With great data comes great (theoretical) opportunity. *Trends in Cognitive Sciences*, **22**(8), 669–671. https://doi.org/10.1016/j.tics.2018.05.011

Garvey, C., & **Caramazza, A.** (1974). Implicit causality in verbs. *Linguistic Inquiry*, **5**(3), 459–464.

Garvey, C., **Caramazza, A.**, & **Yates, J.** (1974). Factors influencing assignment of pronoun antecedents. *Cognition*, **3**(3), 227–243. https://doi.org/10.1016/0010-0277(74)90010-9

Germine, L. T., **Duchaine, B.**, & **Nakayama, K.** (2011). Where cognitive development and aging meet: Face learning ability peaks after age 30. *Cognition*, **118**(2), 201–210. https://doi.org/10.1016/j.cognition.2010.11.002

Gerstenberg, T., **Goodman, N. D.**, **Lagnado, D. A.**, & **Tenenbaum, J. B.** (2021). A counterfactual simulation model of causal judgments for physical events. *Psychological Review*, **128**(5), 936–975. https://doi.org/10.1037/rev0000281

https://doi.org/10.1017/S0305000926100877 Published online by Cambridge University Press

Guan, S., & **Arnold, J. E.** (2021). The predictability of implicit causes: Testing frequency and topicality explanations. *Discourse Processes*, **58**(10), 943–969. https://doi.org/10.1080/0163853X.2021.1974690

Hampshire, A., **Highfield, R. R.**, **Parkin, B. L.**, & **Owen, A. M.** (2012). Fractionating human intelligence. *Neuron*, **76**(6), 1225–1237. https://doi.org/10.1016/j.neuron.2012.06.022

Hartshorne, J. K. (2014). What is implicit causality? *Language, Cognition and Neuroscience*, **29**(7), 804–824. https://doi.org/10.1080/01690965.2013.796396

Hartshorne, J. K. (2020). How massive online experiments (MOEs) can illuminate critical and sensitive periods in development. *Current Opinion in Behavioral Sciences*, **36**, 135–143. https://doi.org/10.1016/j.cobeha.2020.09.005

Hartshorne, J. K., & **Germine, L. T.** (2015). When does cognitive functioning peak? The asynchronous rise and fall of different cognitive abilities across the life span. *Psychological Science*, **26**(4), 433–443. https://doi.org/10.1177/0956797614567339

Hartshorne, J. K., **Huang, Y.**, & **Skorb, L.** (2023). Some puzzling findings regarding the acquisition of verbs. *Language Development Research*, **3**(1). https://doi.org/10.34842/2023.535

Hartshorne, J. K., **Nappa, R.**, & **Snedeker, J.** (2015a). Development of the first-mention bias. *Journal of Child Language*, **42**(2), 423–446. https://doi.org/10.1017/S0305000914000075

Hartshorne, J. K., **O'Donnell, T. J.**, & **Tenenbaum, J. B.** (2015b). The causes and consequences explicit in verbs. *Language, Cognition and Neuroscience*, **30**(6), 716–734. https://doi.org/10.1080/23273798.2015.1008524

Hartshorne, J. K., & **Snedeker, J.** (2013). Verb argument structure predicts implicit causality: The advantages of finer-grained semantics. *Language and Cognitive Processes*, **28**(10), 1474–1508. https://doi.org/10.1080/01690965.2012.689305

Hartshorne, J. K., **Sudo, Y.**, & **Uruwashi, M.** (2013). Are implicit causality pronoun resolution biases consistent across languages and cultures? *Experimental Psychology*, **60**(3), 179–196. https://doi.org/10.1027/1618-3169/a000187

Hartshorne, J. K., **Tenenbaum, J. B.**, & **Pinker, S.** (2018). A critical period for second language acquisition: Evidence from 2/3 million English speakers. *Cognition*, **177**, 263–277. https://doi.org/10.1016/j.cognition.2018.04.007

Herrnstein, R. J. (1997). *The matching law*. Harvard University Press

Järvikivi, J., **Van Gompel, R. P. G.**, & **Hyönä, J.** (2017). The interplay of implicit causality, structural heuristics, and anaphor type in ambiguous pronoun resolution. *Journal of Psycholinguistic Research*, **46**(3), 525–550. https://doi.org/10.1007/s10936-016-9451-1

Johnson, E., & **Arnold, J. E.** (2021). Individual differences in print exposure predict use of implicit causality in pronoun comprehension and referential prediction. *Frontiers in Psychology*, **12**, 672109. https://doi.org/10.3389/fpsyg.2021.672109

Kaiser, E., & **Trueswell, J. C.** (2008). Interpreting pronouns and demonstratives in finnish: Evidence for a form-specific approach to reference resolution. *Language and Cognitive Processes*, **23**(5), 709–748. https://doi.org/10.1080/01690960701771220

Kehler, A. (2002). *Coherence, reference, and the theory of grammar* (Vol. **380**). CSLI publications

Koornneef, A., & **Vanberkum, J.** (2006). On the use of verb-based implicit causality in sentence comprehension: Evidence from self-paced reading and eye tracking. *Journal of Memory and Language*, **54**(4), 445–465. https://doi.org/10.1016/j.jml.2005.12.003

Koornneef, A. W., & **Sanders, T. J. M.** (2013). Establishing coherence relations in discourse: The influence of implicit causality and connectives on pronoun resolution. *Language and Cognitive Processes*, **28**(8), 1169–1206. https://doi.org/10.1080/01690965.2012.699076

Lakens, D. (2022). Sample size justification. *Collabra: Psychology*, **8**(1), 33267. https://doi.org/10.1525/collabra.33267

Lee, R., **Chambers, C. G.**, **Heuttig, F.**, & **Ganea, P. A.** (2022). Children's and adults' use of fictional discourse and semantic knowledge for prediction in language processing. *PLoS ONE*, **17**(4), e0267297. https://doi.org/10.1371/journal.pone.0267297

Lee, C.L., & **Lai, C.H.** (2024). Age-related differences in understanding pronominal reference in sentence comprehension: An electrophysiological investigation. *Psychology and Aging*, **39**(3), 275–287. https://doi.org/10.1037/pag0000775

Leonard, L. B. (2017). *Children with specific language impairment*. MIT press

https://doi.org/10.1017/S0305000926100877 Published online by Cambridge University Press

Lidz, J., **Gleitman, H.**, & **Gleitman, L.** (2003). Understanding how input matters: Verb learning and the footprint of universal grammar. *Cognition*, **87**(3), 151–178. https://doi.org/10.1016/S0010-0277(02)00230-5

Logie, R. H., & **Maylor, E. A.** (2009). An internet study of prospective memory across adulthood. *Psychology and Aging*, **24**(3), 767–774. https://doi.org/10.1037/a0015479

Long, D. L., & **De Ley, L.** (2000). Implicit causality and discourse focus: The interaction of text and reader characteristics in pronoun resolution. *Journal of Memory and Language*, **42**(4), 545–570. https://doi.org/10.1006/jmla.1999.2695

Lyu, S., & **Wang, L.** (2022). Implicit causality and pronoun resolution in intersubjective discourse relations. *Frontiers in Psychology*, **13**, 866103. https://doi.org/10.3389/fpsyg.2022.866103

Marcus, G. F., **Pinker, S.**, **Ullman, M.**, **Hollander, M.**, **Rosen, T. J.**, **Xu, F.**, & **Clahsen, H.** (1992). Overregularization in language acquisition. *Monographs of the Society for Research in Child Development*, **57**(4), 1–182.

Marcus, G. F., **Vijayan, S.**, **Bandi Rao, S.**, & **Vishton, P. M.** (1999). Rule learning by seven-month-old infants. *Science*, **283**(5398), 77–80. https://doi.org/10.1126/science.283.5398.77

Martin-Gonzalez, I., **Ronderos, C. R.**, **Castroviejo, E.**, **Schroeder, K.**, **Lossius-Falkum, I.**, & **Vicente, A.** (2025). That kid is a grasshopper! Metaphor development from 3 to 9 years of age. *Journal of Child Language*, **52**(4), 945–970. https://doi.org/10.1017/S0305000924000187

Maylor, E. A., & **Logie, R. H.** (2010). A large-scale comparison of prospective and retrospective memory development from childhood to middle age. *Quarterly Journal of Experimental Psychology*, **63**(3), 442–451. https://doi.org/10.1080/17470210903469872

Montag, J. L., & **MacDonald, M. C.** (2015). Text exposure predicts spoken production of complex sentences in 8- and 12-year-old children and adults. *Journal of Experimental Psychology: General*, **144**(2), 447–468. https://doi.org/10.1037/xge0000054

Nippold, M. A. (2007). *Later language development: School-age children, adolescents, and young adults* (3rd ed.). PRO-ED

Ozge, D., **Evcen, E.**, & **Hartshorne, J. K.** (2025). Referential form, word order, and implicit causality in Turkish emotion verbs. *Proceedings of the Annual Meeting of the Cognitive Science Society*. https://escholarship.org/uc/item/4pb6b6gg

Paris, A. H., & **Paris, S. G.** (2003). Assessing narrative comprehension in young children. *Reading Research Quarterly*, **38**(1), 36–76. https://doi.org/10.1598/RRQ.38.1.3

Pickering, M. J., & **Majid, A.** (2007). What are implicit causality and consequentiality?. *Language and Cognitive Processes*, **22**(5), 780–788. https://doi.org/10.1080/01690960601119876

Pouscoulous, N., & **Tomasello, M.** (2020). Early birds: Metaphor understanding in 3-year-olds. *Journal of Pragmatics*, **156**, 160–167. https://doi.org/10.1016/j.pragma.2019.05.021

Rice, M. L., & **Wexler, K.** (1996). Toward tense as a clinical marker of specific language impairment in English-speaking children. *Journal of Speech, Language, and Hearing Research*, **39**(6), 1239–1257. https://doi.org/10.1044/jshr.3906.1239

Rice, M. L., **Wexler, K.**, & **Hershberger, S.** (1998). Tense over time: The longitudinal course of tense acquisition in children with specific language impairment. *Journal of Speech, Language, and Hearing Research*, **41**(6), 1412–1431. https://doi.org/10.1044/jslhr.4106.1412

Rohde, H. (2019). Pronouns. In *The Oxford handbook of experimental semantics and pragmatics* (1st ed., pp. 452–473). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198791768.013.21

Román, D. C. (2020). *Department of linguistics* (PhD thesis). University of Alberta

Salthouse, T. A. (2004). What and when of cognitive aging. *Current Directions in Psychological Science*, **13**(4), 140–144. https://doi.org/10.1111/j.0963-7214.2004.00293.x

Schumacher, P. B., **Roberts, L.**, & **Järvikivi, J.** (2017). Agentivity drives real-time pronoun resolution: Evidence from German er and der. *Lingua*, **185**, 25–41. https://doi.org/10.1016/j.lingua.2016.07.004

Singer, T., **Verhaeghen, P.**, **Ghisletta, P.**, **Lindenberger, U.**, & **Baltes, P. B.** (2003). The fate of cognition in very old age: Six-year longitudinal findings in the Berlin aging study (BASE). *Psychology and Aging*, **18**(2), 318–331. https://doi.org/10.1037/0882-7974.18.2.318

Solstad, T., & **Bott, O.** (2022). On the nature of implicit causality and consequentiality: The case of psychological verbs. *Language, Cognition and Neuroscience*, **37**(10), 1311–1340. https://doi.org/10.1080/23273798.2022.2069277

https://doi.org/10.1017/S0305000926100877 Published online by Cambridge University Press

Song, H., & **Fisher, C.** (2005). Who's "she"? Discourse prominence influences preschoolers' comprehension of pronouns. *Journal of Memory and Language*, **52**(1), 29–57. https://doi.org/10.1016/j.jml.2004.06.012

Song, H., & **Fisher, C.** (2007). Discourse prominence effects on 2.5-year-old children's interpretation of pronouns. *Lingua*, **117**(11), 1959–1987. https://doi.org/10.1016/j.lingua.2006.11.011

Stanley, T. D., **Carter, E. C.**, & **Doucouliangos, H.** (2020). What meta-analyses reveal about the replicability of psychological research. *Psychological Bulletin*, **144**(12), 1325–1346. https://doi.org/10.1037/bul0000169

Stine-Morrow, E. A. L., & **Radvansky, G. A.** (2017). Discourse processing and development through the adult lifespan. In M. F. Schober, D. N. Rapp, & M. A. Britt (Eds.), *Routledge handbooks in linguistics. The Routledge handbook of discourse processes* (2nd ed., pp. 247–268). Routledge. https://doi.org/10.4324/9781315687384-14

Tomasello, M. (2005). *Constructing a language: A usage-based theory of language acquisition*. Harvard University Press

Ueno, M., & **Kehler, A.** (2016). Grammatical and pragmatic factors in the interpretation of Japanese null and overt pronouns. *Linguistics*, **54**(6), 1165–1221. https://doi.org/10.1515/ling-2016-0027

van den Hoven, E., & **Ferstl, E. C.** (2018). The roles of implicit causality and discourse context in pronoun resolution. *Frontiers in Communication*, **3**, 53. https://doi.org/10.3389/fcomm.2018.00053

Verhaeghen, P. (2003). Aging and vocabulary score: A meta-analysis. *Psychology and Aging*, **18**(2), 332–339. https://doi.org/10.1037/0882-7974.18.2.332

Wang, T., & **Gabriele, A.** (2023). Individual differences modulate sensitivity to implicit causality bias in both native and nonnative processing. *Studies in Second Language Acquisition*, **45**(4), 853–881. https://doi.org/10.1017/S0272263122000468

Weatherford, K. C., & **Arnold, J. E.** (2021). Semantic predictability of implicit causality can affect referential form choice. *Cognition*, **214**, 104759. https://doi.org/10.1016/j.cognition.2021.104759

Wiig, E. H., **Semel, E.**, & **Secord, W. A.** (2013). *Clinical evaluation of language fundamentals–fifth edition (CELF-5)* (5th ed.). NCS Pearson

Winograd, T. (1972). Understanding natural language. *Cognitive Psychology*, **3**(1), 1–191. https://doi.org/10.1016/0010-0285(72)90002-3

Ye, Y., & **Arnold, J. E.** (2023). Learning the statistics of pronoun reference: By word or by category? *Cognition*, **239**, 105546. https://doi.org/10.1016/j.cognition.2023.105546

Ye, Y., & **Arnold, J. E.** (2025). Implicit causality can affect pronoun use in fragment completion tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. https://doi.org/10.1037/xlm0001435

Yuile, A. R. (2022). Children's use of event representations in reference comprehension. *[Doctoral Dissertation, UIUC]. IDEALS @ Illinois.* https://hdl.handle.net/2142/117820

Zerkle, S. A., **Rosa, E. C.**, & **Arnold, J. E.** (2017). Thematic role predictability and planning affect word duration. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, **8**(1), 17–28. https://doi.org/10.5334/labphon.98

Zwaan, R. A., & **Radvansky, G. A.** (1998). Situation models in language comprehension and memory. *Psychological Bulletin*, **123**(2), 162–185. https://doi.org/10.1037/0033-2909.123.2.162

https://doi.org/10.1017/S0305000926100877 Published online by Cambridge University Press

Appendix

Estimated ability across complete age range

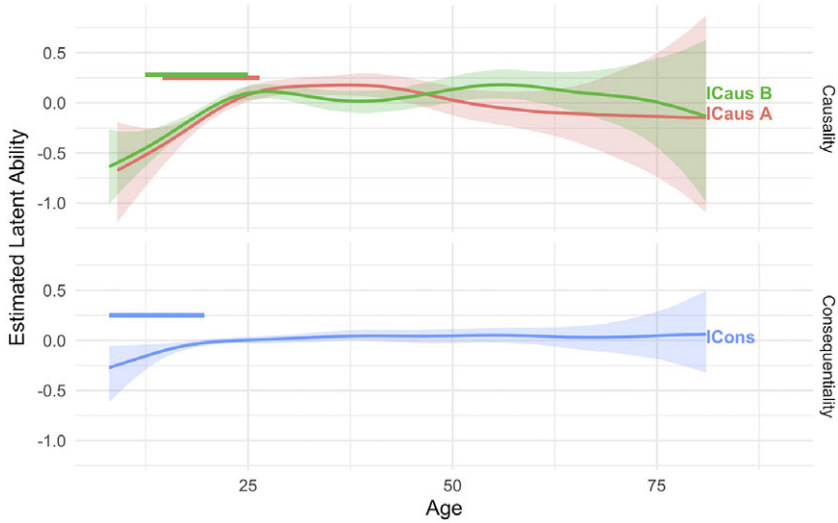


Figure A1. Estimated latent ability across complete age range. Shading represents 95% credible intervals. Horizontal bars show age ranges where the estimated change in ability was significant. Note that these analyses depend on slope estimates.

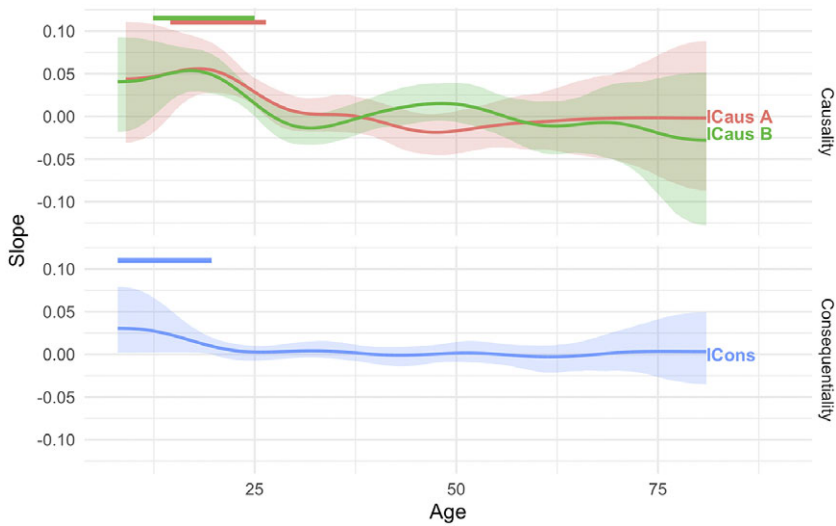


Figure A2. Estimated slope across complete age range. Shading represents 95% confidence intervals. Horizontal bars show age ranges where the estimated change in ability was significant.

Cite this article: Yuile, A.R., & Hartshorne, J.K. (2026). Age-Related Changes in Implicit Causality and Consequentiality Pronoun Biases. *Journal of Child Language* 1–23, https://doi.org/10.1017/S0305000926100877

https://doi.org/10.1017/S0305000926100877 Published online by Cambridge University Press