

Developmental Relations Between Cognitive Reflection and Modal Cognition

Andrew Shtulman (shtulman@oxy.edu)

Charlotte Harrington (charrington@oxy.edu)

Chloe Hetzel (chetzel@oxy.edu)

Josephine Kim (jkim3@oxy.edu)

Caroline Palumbo (cpalumbo@oxy.edu)

Theodore Rountree-Shtulman (teddyroutree@gmail.com)

Department of Psychology, Occidental College

1600 Campus Road, Los Angeles, CA 90041

Abstract

Children can be unduly skeptical of events that violate their expectations, claiming these events neither could happen nor should happen, even if the events violate no physical or social laws. Here, we explore whether children's reasoning about possibility and permissibility—modal cognition—is aided by cognitive reflection, or the disposition to privilege analysis over intuition. Ninety-nine children between the ages of 4 and 11 judged the possibility and permissibility of several hypothetical events, and their judgments were compared to their score on a developmental version of the cognitive reflection test, the CRT-D. Children's CRT-D scores predicted their ability to differentiate possible events from impossible ones and their ability to differentiate impermissible events from permissible ones, as well as their ability to differentiate possibility from permissibility in general. Such differentiations were predicted by children's CRT-D scores independent of age and executive function. These findings suggest that mature modal cognition may require the ability to reflect on, and override, the intuition that unexpected events cannot happen.

Keywords: conceptual development, possibility judgment, permissibility judgment, cognitive reflection

Introduction

Reasoning about possibility is central to learning and decision making. Events must be deemed possible to be accepted as facts, anticipated as outcomes, or pursued as goals (Phillips & Knobe, 2018). When thinking about the climate, for instance, a person must grant that global warming is possible to accept that it is happening, plan for its consequences, and strive to mitigate its effects. Despite the centrality of possibility judgment, it develops slowly. Young children fail to differentiate constraints that render events impossible from those that render them merely improbable. They deny the possibility of events that violate physical laws, like walking on water or walking through a wall, but they also deny the possibility of events that violate mere regularities, like making pickle-flavored ice cream or owning a lion for a pet (Goulding & Friedman, 2021, 2022; Nissel et al., 2023; Nolan-Reyes et al., 2016; Shtulman & Carey, 2007).

Children's conflation of improbability and impossibility is highly robust, particularly among preschoolers. Preschoolers deny the possibility of a wide range of improbable events, including walking on a wire, catching a fly with chopsticks,

reading someone's lips, having twenty siblings, or growing a beard to one's toes (Shtulman, 2009). They deny the possibility of improbable events regardless of where they learn about the events (Danovitch & Lane, 2020), who informs them of the events (Williams & Danovitch, 2022), and how the events are framed (Lane et al., 2018). Prompting preschoolers to envision improbable events in their imagination does not increase their acceptance of those events (Lane et al., 2016), nor does prompting them to explain how the events might occur (Nancekivell & Friedman, 2017).

Preschoolers not only deny the possibility of events that violate physical regularities but also those that violate social regularities, such as traditions, customs, or rules. They deny that a child could sing jingle bells at a birthday party, wear pajamas to the grocery store, or eat dinner under the table (Browne & Woolley, 2004). They deny that adults could change the side of the road we drive on, change the name of dogs to wugs, or change the color of stoplights from red to purple (Komatsu & Galotti, 1986). They even deny that a person could commit common moral infractions, like lying or stealing, which they have undoubtedly observed, if not committed, themselves (Chernyak et al., 2019).

Children's tendency to deny any event that violates their expectations was aptly demonstrated in a study by Shtulman and Phillips (2018), where children between the ages of three and ten were presented with four types of unexpected events—impossible, improbable, unconventional, and immoral—and asked to judge the events' possibility. In one vignette, children were told about a girl named Sophie, who, while baking bread for a neighbor, accidentally stirred a spider into the dough. Children were then asked to evaluate the following solutions to this problem: (a) Sophie uses a time machine to go back in time and catch the spider before it falls into the dough (an impossible event); (b) Sophie tells her neighbor that a spider fell into the dough, and her neighbor buys the bread anyway (an improbable event); (c) Sophie bakes the dough and sells the bread to her neighbor without telling him about the spider (an immoral event); and (d) Sophie bakes the dough and eats the bread herself (an unconventional event).

Younger children denied the possibility of both improbable events and impossible events, as observed in previous studies, but they also denied the possibility of unconventional

events and immoral events. That is, they denied the possibility of events that violated any kind of expectation, regardless of whether the expectation was physical or social and regardless of whether it corresponded to a law or mere regularity. Older children were better than younger children at differentiating events that violate physical laws (impossible events) from those that do not (improbable events, unconventional events, and immoral events), but even older children often denied the possibility of the latter.

In addition to judging possibility, the children in Shtulman and Phillips's (2018) study also judged permissibility. Permissibility, like possibility, is a facet of modal cognition, or reasoning about whether an event must, can, or cannot occur. Deciding whether an event is permissible corresponds both logically (Kratzer, 1977) and conceptually (Shtulman & Tong, 2013) to deciding whether it is possible. Indeed, Shtulman and Phillips found that children judged permissibility similarly to how they judged possibility: they claimed that events that violated their expectations were not only impossible but also wrong. Eating spider-bread, for instance, was judged as both impossible and impermissible. The same was found for the law-violating events—impossible events and immoral events—which should be judged differently across the two modal domains. Children claimed (correctly) that Sophie could not travel back in time, but they also claimed that it would be wrong for her to do so. And they claimed (correctly) that it would be wrong for Sophie to lie to her neighbor, but they also claimed it would be impossible for her to do so.

It thus appears that when children encounter an event that violates their expectations, their immediate assessment is that it *can't* happen in a generic sense of *can't* that conflates impossibility with improbability, unconventionality, and impermissibility. Not until the end of elementary school do children begin distinguishing events that violate laws from those that violate mere regularities, as well as the domain of the violation itself (physical or social).

This account of modal development implies that children need to reflect on their modal intuitions to determine whether those intuitions are correct—that is, they need to reflect on whether the expectation violated by an unexpected event precludes the event from occurring. Reflection could help children override a modal intuition in several ways. It could help them identify circumstances that would bring the event about or would render it morally acceptable. It could help them identify a precedent for the event—a similar event known to have occurred or known to be permissible. Or it could help them realize that the event, though unexpected, does not violate any physical or moral principles and is therefore within the realm of things that can occur. Circumstances, precedents, and principles are the kinds of considerations that adults take into account when making modal judgments (Shtulman & Tong, 2013), and it's open question when children begin to reflect on such considerations as well.

In the present study, we assessed how reflection might impact children's modal judgments by comparing those

judgments to their performance on the Cognitive Reflection Test, Developmental Version, or CRT-D (Young & Shtulman, 2020a). This test consists of nine brainteasers designed to elicit an incorrect intuitive response that can be overridden with reflection, even by young children. One such brainteaser is “What do cows drink?”, which elicits the intuitive response “milk.” However, a moment's reflection reveals that, although cows produce milk, they drink water.

The CRT-D was modeled after the adult Cognition Reflection Test, or CRT (Frederick, 2005), and like the CRT, has proven highly predictive of tasks that pit intuition against analysis. It predicts children's rational thought, as measured by heuristics-and-biases tasks like denominator neglect, base rate sensitivity, belief bias syllogisms, and otherside thinking (Gong et al., 2021). It predicts children's understanding of counterintuitive concepts in science and mathematics, such as vitalist biology and arithmetic equivalence (Young & Shtulman, 2020a). And it predicts children's ability to learn counterintuitive concepts from instruction (Young & Shtulman, 2020b). The CRT-D's relation to rational thought and conceptual understanding holds in diverse cultural contexts and persists even when controlling for age and executive function (Gong et al., 2021; Young et al., 2018).

Here, we investigated whether the CRT-D predicts children's modal cognition. If reflection aids in judging an expectation-defying event possible or permissible, then children with higher CRT-D scores should demonstrate greater differentiation of events that violate modally-relevant considerations, such as physical laws or social laws, from those that do not. Moreover, CRT-D scores should predict children's modal judgments independent of age and executive function, as reflection is neither redundant with, nor guaranteed by, either factor.

Method

Participants

Ninety-nine children between the ages of 4 and 11 were recruited from public playgrounds and tested onsite. Their average age was 8.1 years ($SD = 2.0$), and 55% were female. Five additional children were recruited but did not complete the experimental session. Children were recruited from a community that is 35% White, 35% Hispanic or Latino, 18% Asian, and 8% Black; the community is largely middle class, with 14% of the population living below the poverty line. Children completed all tasks on an iPad, administered using Qualtrics. These tasks measured cognitive reflection, modal judgment, and executive function and were administered in that order.

Cognitive Reflection

Children completed the nine-item CRT-D, developed by Young and Shtulman (2020) to measure cognitive reflection in elementary-school-aged children. The CRT-D consists of nine brainteasers that elicit an intuitive response correctable with reflection, such as “Which weighs more: a pound of rocks or a pound of feathers?”. CRT-D scores were calculated

as the number of correct responses provided, which ranged from 0 to 8. Children’s average score was 3.0 ($SD = 2.1$), and scores generally increased with age ($r = .54, p < .001$).

Modal Judgment

Children’s understanding of possibility and permissibility was assessed using six vignettes from Shtulman and Phillips (2018). Each vignette describes a character facing an interpersonal challenge, followed by solutions that exemplify five types of events: impossible, improbable, unconventional, immoral, and ordinary. The story about Sophie baking bread was one such vignette. Another was about a boy named Ralph, who was attending the birthday party of a boy named John. Children were told, “When it’s time for cake, Ralph finishes his slice and wants to eat more, but the rest of the cake is gone.” They were then asked to judge the possibility or permissibility of the following solutions:

- Ralph waves his hands over his plate and a new slice of cake instantly appears there. (impossible)
- Ralph tells John’s mother that he is still hungry, and she brings him a cake of his very own. (improbable)
- Ralph waits until everyone clears their plates and then searches the trash for any leftover cake. (unconventional)
- When John is not looking, Ralph steals John’s cake off his plate and shoves it in his mouth. (immoral)
- Ralph excuses himself from the table and goes outside to play birthday games with his friends (ordinary).
- Ralph eats a slice of watermelon instead. (ordinary)

Children made possibility judgments for three vignettes and permissibility judgments for the other three, counterbalanced across children. Vignettes were randomized within each judgment block, as were the events within each vignette. Two ordinary events were included per vignette, rather than one, to counter the expectation that all events under consideration were impossible or wrong. The full set of vignettes are included in materials posted to the Open Science Framework: <https://tinyurl.com/57ebhe5t>. All data and analyses are posted there as well.

Possibility judgments were elicited by asking whether the event “could happen in real life,” and permissibility judgments were elicited by asking whether the event was “okay to do in real life.” These questions were followed by a measure of certainty. Children who said the event could not happen were asked whether it was “kinda impossible or very impossible,” and children who said the event was not okay were asked whether it was “kinda wrong or very wrong.” These questions yielded a judgment score ranging from 0 (very impossible/wrong) to 1 (kinda impossible/wrong) to 2 (possible/okay). Our rationale for using a three-point scale, rather than summing the number of times participants selected the “possible” option, was that it provided greater sensitivity to the development of modal distinctions. Children inclined to judge unexpected events as impossible or impermissible might still recognize distinctions among them, selecting “very impossible” for impossible events but “kinda

impossible” for improbable events or selecting “very wrong” for immoral events but “kinda wrong” for unconventional events. A three-point scale allowed us to capture budding differentiations between law-violating events and regularity-violating events that would be obscured if we focused solely on the possible/permissible side of the modal divide.

Along these same lines, our primary analyses of children’s modal judgments were conducted using difference scores, where we subtracted judgments for law-violating events from those for other events. Doing so allowed us to use the former as an internal baseline for the latter, thereby amplifying modal distinctions made by children who might have perseverated on the impossible/impermissible side of the modal divide but were still attentive to domain-relevant modal distinctions. For instance, a child who judged improbable events as “kinda impossible” but judged impossible events as “very impossible” would appear to have a more nuanced understanding of possibility than a child who judged both types of events as “kinda impossible” or who vacillated between “kinda impossible” and “very impossible” at random. The same could be said for a child who judged unconventional events as “kinda wrong” but judged immoral events as “very wrong.” Difference scores could capture such emerging distinctions, which were likely more common among younger children.

Another reason we focus on difference scores is that they provide a more comprehensive measure of modal cognition—a measure that spans their reasoning about causal violations of varying strength (law-violating vs. regularity-violating) and content (physical vs. social). Such a measure allowed us to examine the distinctions children make not only within a modal domain but also across domains, given that children’s initial sense of can’t appears to conflate these domains (Chernyak et al., 2019; Shtulman & Phillips, 2018).

Executive Function

Inhibitory control. We measured inhibitory control using the day-night task (Gerstadt et al., 1994). On congruent trials, children were instructed to tap a picture of the sun when they heard “day” and a picture of the moon when they heard “night.” On incongruent trials, they made the opposite selections. Children completed 10 trials of each type, with near-ceiling accuracy (over 97% correct). Because children made few errors, we focused on their speed. Children took an average of 1.37 seconds ($SD = 0.52$) to respond to congruent trials and 1.67 seconds ($SD = 0.56$) to respond to incongruent trials—a reliable difference ($t(98) = 6.89, p < .001$). Mean response time for incongruent trials was used as our measure of inhibitory control. Response time was negatively correlated with age, as older children responded more quickly than younger ones ($r = -.56, p < .001$).

Working memory. We measured working memory with a backwards digit span task (Alloway et al., 2009), where children heard a series of digits and were instructed to repeat them in reverse order. The series ranged from 2 digits to 7,

increasing by one digit every other trial. If children responded incorrectly on two consecutive trials, the task was ended. We used the span of children's last correct response as our measure of working memory. Children who were unsuccessful at repeating 2 digits backwards (because they repeated them forwards instead) were assigned a score of 1. Scores ranged from 1 to 7 and averaged 3.1 ($SD = 1.3$). Scores generally increased with age ($r = .52, p < .001$).

Set shifting. We measured set shifting with a verbal fluency task (Snyder & Munakata, 2010). Children were asked to name as many animals as they can in a minute, followed by as many foods as they can. Successful performance requires shifting between sets of animals (e.g., pets, farm animals, zoo animals) or sets of food (e.g., breakfast foods, snacks, desserts) as each set is exhausted. The decision of when to shift sets, and what set to shift to, underlies much of the variation in children's performance (Snyder & Munakata, 2010). We used the number of category-appropriate exemplars as our measure of set shifting, averaged across the two categories. Children named an average of 15.3 exemplars ($SD = 6.1$), with older children naming more exemplars than younger ones ($r = .61, p < .001$).

EF Composite. Children who succeeded on one executive function task tended to succeed on the others as well. Inhibitory control was correlated with working memory ($r = .40, p < .001$), working memory with set shifting ($r = .56, p < .001$), and set shifting with inhibitory control ($r = .55, p < .001$). These intercorrelations provided justification for combining the three measures into a single composite. Scores on each test were standardized, with scores for inhibitory control reverse-coded. Standardized scores were then averaged, yielding EF composite scores that ranged from -2.35 to +2.28. These scores were correlated with age ($r = .69, p < .001$). They were also correlated with cognitive reflection ($r = .55, p < .001$), as observed in previous research with children (Young & Shtulman, 2020a), as well as adults (Toplak et al., 2011).

Results

Possibility Judgments

Children judged an event's possibility on a scale from very impossible (0) to kinda impossible (1) to possible (2). The normative pattern would be to judge all events possible except impossible events. While children routinely affirmed the possibility of ordinary events (M judgment score = 1.9, $SD = 0.2$) and denied the possibility of impossible ones ($M = 0.3, SD = 0.5$), they were more equivocal for improbable events ($M = 1.2, SD = 0.6$), unconventional events ($M = 1.5, SD = 0.6$), and immoral events ($M = 1.5, SD = 0.6$), judging them very impossible 21% of the time and kinda impossible 20% of the time.

Children with higher CRT-D scores were more likely to affirm the possibility of unconventional events ($r = .24, p = .017$) and immoral events ($r = .22, p = .030$) and more likely to reject the possibility of impossible events ($r = -.32, p =$

.001). A similar pattern was observed for EF scores; children with higher EF scores were more likely to affirm the possibility of ordinary events ($r = .32, p = .001$), unconventional events ($r = .50, p < .001$), and immoral events ($r = .40, p < .001$) and more likely to reject the possibility of improbable events ($r = -.49, p < .001$). Children's judgments for improbable events were not correlated with either CRT-D scores ($r = .02, p = .858$) or EF scores ($r = .06, p = .567$)—an unexpected finding that may have stemmed from the nature of our improbable events, as we elaborate in the Discussion. In general, however, children's possibility judgments were predicted by both their CRT-D scores and their EF scores.

Permissibility Judgments

Children judged an event's permissibility on a scale from very wrong (0) to kinda wrong (1) to okay (2). The normative judgment pattern would be to judge all events permissible (i.e., okay) except immoral ones. Children routinely affirmed the permissibility of ordinary events (M judgment score = 1.8, $SD = 0.3$) and denied the permissibility of immoral events ($M = 0.2, SD = 0.4$) but were equivocal for unconventional events ($M = 0.8, SD = 0.5$), improbable events ($M = 1.0, SD = 0.6$), and impossible events ($M = 1.0, SD = 0.7$), judging them very wrong 40% of the time and kinda wrong 26% of the time.

Children with higher CRT-D scores were more likely to affirm the permissibility of ordinary events ($r = .28, p = .005$) and improbable events ($r = .29, p = .004$) and more likely to reject the permissibility of immoral events ($r = -.20, p = .049$). Judgments for impossible events and unconventional events did not correlate with CRT-D scores ($r = .11, p = .288$, and $r = -.14, p = .155$, respectively). Children with higher EF scores were more likely to affirm the permissibility of ordinary events ($r = .49, p > .001$), but EF scores were not correlated with permissibility judgments for any other type of event (all $r < .16, p > .11$). Thus, in contrast to children's possibility judgments, their permissibility judgments were predicted by their CRT-D scores more consistently than by their EF scores.

Within-Domain Differentiations

To explore the relation between CRT-D scores and possibility judgments at a global level—across all events—we computed a differentiation score by subtracting mean judgment scores for impossible events from mean judgment scores for the other four events. A difference of 2 would indicate perfect differentiation of events that violate physical laws (impossible events) from those that do not, whereas a difference of 0 (or lower) would indicate a lack of differentiation. Children's difference scores ranged from -.6 to 2.0 and averaged 1.2 ($SD = 0.7$). Their correlation with children's CRT-D scores was $r = .37, p < .001$.

We computed a differentiation score for permissibility judgments in a similar manner; mean judgment scores for immoral events were subtracted from mean judgment scores for the other four events. A difference of 2 would indicate

perfect differentiation of events that violate social laws (immoral events) from those that do not, whereas a difference of 0 (or lower) would indicate a lack of differentiation. Children’s differentiation scores ranged from -9 to 1.9 and averaged 1.0 ($SD = 0.5$). Their correlation with children’s CRT-D scores was $r = .36, p < .001$.

Between-Domain Differentiations

The differentiation scores computed above reflect children’s ability to make distinctions within a modal domain. To measure their ability to make distinctions across domains, we focused on the two events that warrant opposite judgments: impossible events and immoral events. Children should claim that impossible events cannot occur but are okay, whereas they should claim that immoral events are not okay but can occur.

For impossible events, we subtracted children’s mean possibility judgment score from their mean permissibility judgment score, yielding a between-domain possibility score that could range from -2 to +2, where +2 is the normative score. Children’s actual scores ranged from -1.3 to 2.0 and averaged 0.7 ($SD = 0.8$). They were reliably correlated with children’s CRT-D scores: $r = .34, p < .001$. For immoral events, we subtracted children’s mean permissibility judgment score from their mean possibility judgment score, yielding a between-domain permissibility score that could range from -2 to +2, where +2 is once again the normative score. Children’s scores ranged from -1.0 to 2.0 and averaged 1.3 ($SD = 0.7$). They were also correlated with children’s CRT-D scores: $r = .29, p < .01$.

Comparison of Predictors

Cognitive reflection consistently predicted children’s modal differentiations within and between domains, but how does cognitive reflection compare to age and executive function? We addressed this question by regressing modal differentiation scores against CRT-D scores, EF scores, and age. We ran the regressions in three steps, entering age in the first step, EF scores in the second, and CRT-D scores in the third. We also entered the interaction between CRT-D scores and age in the third step, to assess whether CRT-D scores predicted younger children’s modal differentiation scores more strongly than older children’s or vice versa. By structuring our regression in this manner, we were able to assess whether executive function predicted modal judgments independent of age and whether cognitive reflection predicted modal judgments independent of both age and executive function.

The final models from each regression are presented in Table 1. Age, when entered by itself, explained a significant amount of variance in all four sets of differentiation scores: within-domain possibility, $R^2 = .38, F(1,97) = 58.50, p < .001$; within-domain permissibility, $R^2 = .06, F(1,97) = 6.47, p = .013$; between-domain possibility, $R^2 = .10, F(1,97) = 10.64, p = .002$; and between-domain permissibility: $R^2 = .16, F(1,97) = 17.75, p < .001$. EF scores, when entered after age, explained significantly more variance in only one set of

scores: within-domain possibility, $\Delta R^2 = .08, F(1,96) = 15.12, p < .001$. CRT-D scores, when entered after age and EF scores, explained significantly more variance in three sets of scores: within-domain possibility $\Delta R^2 = .09, F(2,94) = 8.85, p < .001$; within-domain permissibility, $\Delta R^2 = .09, F(2,94) = 4.84, p = .010$; and between-domain possibility, $\Delta R^2 = .06, F(2,94) = 3.26, p = .043$. That is, CRT-D scores and their interaction with age accounted for variance in three of four differentiation scores that was not accounted for by either age or EF scores.

Of the four predictors in the final model of each regression analysis, CRT-D scores proved the most consistent. They were a significant predictor in three of the four models (within-domain possibility, within-domain permissibility, and between-domain permissibility), whereas age was a significant predictor in only two (within-domain possibility and between-domain permissibility), and EF scores were a significant predictor in only one (within-domain possibility).

Cognitive reflection further predicted differentiation scores through its interaction with age. The CRT-D-by-age interaction was negative in all four models and significantly so in two (within-domain possibility and between-domain permissibility). We explored these interactions by comparing the correlation between CRT-D scores and modal differentiation scores for younger children (aged 4-7, $n = 46$) to that of older children (aged 8-12, $n = 53$). The former were consistently larger than the latter (within-domain possibility: $r = .29$ vs. $-.02$; within-domain permissibility: $r = .33$ vs. $.30$; between-domain possibility: $r = .41$ vs. $.14$; between-domain permissibility: $r = .19$ vs. $.16$), indicating that cognitive reflection was a stronger predictor of modal cognition for younger children than for older children. These interactions do not appear to have been driven by ceiling effects on the part of older children, as no child earned a perfect score of 9, and only 14% of children earned scores between 6 and 8.

Table 1: Regression analyses of modal differentiation scores assessing the contribution of CRT-D scores independent of age and executive function. * $p < .05$, ** $p < .01$, *** $p < .001$

| Differentiation Score | Predictor | β |
|-------------------------------|-------------|---------|
| Within-domain Possibility | Age | .31*** |
| | EF | .21** |
| | CRT-D | .41*** |
| | CRT-D x age | -.05*** |
| Within-domain permissibility | Age | .09 |
| | EF | -.05 |
| | CRT-D | .24* |
| | CRT-D x age | -.02 |
| Between-domain possibility | Age | .18 |
| | EF | -.13 |
| | CRT-D | .31 |
| | CRT-D x age | -.03 |
| Between-domain permissibility | Age | .23** |
| | EF | .11 |
| | CRT-D | .35* |
| | CRT-D x age | -.04* |

Discussion

Children tend to deny the possibility and permissibility of unexpected events, claiming that such events cannot happen, and should not happen, in the real world. Might this tendency be due to a lack of reflection? The correspondence between children's CRT-D scores and their modal judgments suggests yes. Children who were more successful at solving brainteasers like "What do cows drink?" were also more successful at differentiating events that violate laws from those that violate mere regularities. They were also more successful at differentiating distinct senses of modality, recognizing that impossible events can still be permissible and impermissible events can still be possible.

The correspondence we observed between cognitive reflection and modal judgment held even when controlling for age, which suggests that the development of modal cognition is not driven solely by older children's greater amount of event-relevant knowledge. Knowledge may be necessary to affirm the possibility or permissibility of an unexpected event, but it does not appear to be sufficient. Consider the unexpected event of eating spider-bread, from the vignette about Sophie and her baking accident. Accepting that this event could happen (and would be okay to happen) requires knowing certain facts about spiders and certain facts about baking—that spiders are edible, that some people like exotic foods, that baking kills germs, and so forth. But facts alone may not suffice; children need to search for and access these facts, to decide whether their intuition that a person can't eat spider-bread is grounded in any fundamental principles. While eating spider-bread might sound gross, children who reflect on this intuition may realize that it does violate any physical or social laws. They may also come to think of circumstances under which such an event might occur (e.g., that Sophie likes to try new foods) or actual events that establish a precedent for this hypothetical one (e.g., that people in other cultures eat insects).

Cognitive reflection predicted modal judgment independent of executive function as well. Executive function reliably predicted children's judgments about possibility but did not reliably predict their judgments about permissibility. Moreover, when controlling for executive function, cognitive reflection continued to explain variance in children's modal differentiation scores both within and between domains. These findings suggest that cognitive reflection facilitates modal judgment above and beyond the specific information-processing skills constitutive of executive function. Cognitive reflection certainly draws on such skills; one must inhibit an intuitive response in order to shift to an alternative response, all while keeping the original question in mind. But cognitive reflection transcends these skills in that children must initiate them and coordinate them, without external guidance. This kind of meta-level reasoning may be what separates children who base their modal judgments on an immediate intuition from those who override that intuition in light of countervailing considerations.

While we confirmed the expectation that children's cognitive reflection would predict their modal cognition, we discovered two unexpected findings. First, in our analysis of event-specific judgments, neither CRT-D scores nor EF scores predicted children's acceptance of improbable events. Although children denied the possibility of improbable events more often than not, they tended to view these events as "kinda impossible" rather than "very impossible." We suspect that our improbable events, borrowed from Shtulman and Phillips (2018), struck children as less extraordinary than those used in previous studies (e.g., Goulding & Friedman, 2022; Lane et al., 2016; Nolan-Reyes et al., 2016; Williams & Danovitch, 2022). Our improbable events involved atypical behaviors, such as a storeowner giving away free candy or a mother attending school with her daughter, rather than extraordinary situations, such as finding an alligator under the bed or growing a beard to one's toes. Even so, CRT-D scores predicted children's acceptance of the *permissibility* of these events, implying that reflection was still needed to counter the social expectations they elicited.

Second, cognitive reflection impacted younger children's judgments to a greater degree than older children's. While younger children may need to reflect on an unexpected event to deem it possible or permissible, older children may arrive at the same conclusion with less reflection, if any. Older children may have enough knowledge about the event—in the form of circumstances, precedents, or principles—that their initial intuitions are correct, or at least better-informed. Younger children's intuitions, on the other hand, are almost certainly more parochial; their limited knowledge presumably makes unexpected events more unexpected for them than for an older child. But as limited as young children's knowledge might be, they still have some knowledge relevant to the event at hand—some circumstance, precedent, or principle that, if accessed, might counter their immediate intuition. But accessing such considerations likely requires reflection. In other words, reflection may be required for younger children to reach the same conclusion that older children reach spontaneously, as younger children must actively search a smaller knowledgebase of relevant considerations.

In conclusion, children who are better at reflecting on their own cognition are also better at assessing the modal status of various expectation-defying events, as well as the sense of modality relevant to this assessment. Reflection may thus facilitate the emergence of mature modal cognition, allowing children to override intuitions that conflate possibility, probability, and permissibility and engage in a more analytic assessment of the modal status of unexpected events.

Acknowledgments

We would like to thank the James S. McDonnell Foundation for supporting this research through an Understanding Human Cognition Scholar Award to Andrew Shtulman. We would also like to thank Alexander Dong, Mohuli Ganguly, Matthew Lee, Abby McLaughlin, and Emmy Staples for their assistance with data collection.

References

- Alloway, T. P., Gathercole, S. E., Kirkwood, H., & Elliott, J. (2009). The cognitive and behavioral characteristics of children with low working memory. *Child Development, 80*, 606-621.
- Browne, C.A., & Woolley, J.D. (2004). Preschoolers' magical explanations for violations of physical, social, and mental laws. *Journal of Cognition and Development, 5*, 239-260.
- Chernyak, N., Kang, C., & Kushnir, T. (2019). The cultural roots of free will beliefs: How Singaporean and US Children judge and explain possibilities for action in interpersonal contexts. *Developmental Psychology, 55*, 866-876.
- Danovitch, J.H., & Lane, J.D. (2020). Children's belief in purported events: When claims reference hearsay, books, or the internet. *Journal of Experimental Child Psychology, 193*, 104808.
- Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives, 19*, 25-42.
- Gerstadt, C. L., Hong, Y. J., & Diamond, A. (1994). The relationship between cognition and action: performance of children 3½-7 years old on a Stroop-like day-night test. *Cognition, 53*, 129-153.
- Gong, T., Young, A. G., & Shtulman, A. (2021). The development of cognitive reflection in China. *Cognitive Science, 45*, 12966.
- Goulding, B. W., & Friedman, O. (2021). A similarity heuristic in children's possibility judgments. *Child Development, 92*, 662-671.
- Goulding, B.W., Stonehouse, E.E., & Friedman, O. (2022). Causal knowledge and children's possibility judgments. *Child Development, 93*, 794-803.
- Kratzer, A. (1977). What "must" and "can" must and can mean. *Linguistics and Philosophy, 1*, 337-355.
- Lane, J.D., Ronfard, S., & El-Sherif, D. (2018). The influence of first-hand testimony and hearsay on children's belief in the improbable. *Child Development, 89*, 1133-1140.
- Lane, J.D., Ronfard, S., Francioli, S.P., & Harris, P.L. (2016). Children's imagination and belief: Prone to flights of fancy or grounded in reality? *Cognition, 152*, 127-140.
- Nancekivell, S.E., & Friedman, O. (2017). She bought the unicorn from the pet store: six-to seven-year-olds are strongly inclined to generate natural explanations. *Developmental Psychology, 53*, 1079-1087.
- Nissel, J., Li, H., Cramer, A., & Woolley, J. D. (2023). Three men make a tiger: The effect of consensus testimony on Chinese and U.S. children's judgments about possibility. *Journal of Cognition and Culture, 23*, 98-126.
- Nolan-Reyes, C., Callanan, M.A., & Haigh, K.A. (2016). Practicing possibilities: Parents' explanations of unusual events and children's possibility thinking. *Journal of Cognition and Development, 17*, 378-395.
- Phillips, J., & Knobe, J. (2018). The psychological representation of modality. *Mind & Language, 33*, 65-94.
- Shtulman, A. (2009). The development of possibility judgment within and across domains. *Cognitive Development, 24*, 293-309.
- Shtulman, A., & Carey, S. (2007). Improbable or impossible? How children reason about the possibility of extraordinary events. *Child Development, 78*, 1015-1032.
- Shtulman, A., & Phillips, J. (2018). Differentiating "could" from "should": Developmental changes in modal cognition. *Journal of Experimental Child Psychology, 165*, 161-182.
- Shtulman, A., & Tong, L. (2013). Cognitive parallels between moral judgment and modal judgment. *Psychonomic Bulletin & Review, 20*, 1327-1335.
- Snyder, H. R., & Munakata, Y. (2010). Becoming self-directed: Abstract representations support endogenous flexibility in children. *Cognition, 116*, 155-167.
- Toplak, M. E., West, R. F., & Stanovich, K. E. (2011). The Cognitive Reflection Test as a predictor of performance on heuristics-and-biases tasks. *Memory & Cognition, 39*, 1275-1289.
- Williams, A.J., & Danovitch, J.H. (2022). Is what Mickey Mouse says impossible? Informant reality status and children's beliefs in extraordinary events. *Journal of Cognition and Development, 23*, 323-339.
- Young, A., & Shtulman, A. G. (2020a). Children's cognitive reflection predicts conceptual understanding in science and mathematics. *Psychological Science, 31*, 1396-1408.
- Young, A. G., & Shtulman, A. (2020b). How children's cognitive reflection shapes their science understanding. *Frontiers in Psychology, 11*, 1247.