Early Evaluations of Caregivers Who Help and Hinder Safe and Dangerous Goals

Rodney Tompkins (rtompkins@ucsd.edu)
Department of Psychology
University of California, San Diego

Lindsey J. Powell (ljpowell@ucsd.edu)
Department of Psychology
University of California, San Diego

Abstract

Two experiments collected third-party evaluations from U.S. 4–5-year-old children (N = 80) who heard stories about caregivers helping or hindering their infants' achievement of safe, dangerous, or ambiguous goals. Children's evaluations were sensitive to danger: They switched from positively evaluating parents who helped access safe objects, to negatively evaluating those who helped access dangerous objects. Older children offered robustly positive evaluations of parents who protectively hindered access to dangerous objects, but younger participants were more likely to negatively evaluate these parents. Given a moderately risky goal that participants themselves judged as unsafe, children's evaluations of helping and hindering were mixed, though there was preliminary evidence of a developmental shift. These findings show that young children go beyond basic inferences about whether an act promotes or hampers another agent's goal when considering whether the action was good or bad. Instead, young children consider the broader consequences for the target's welfare.

Keywords: protection; caregiving; safety; risk; danger; morality; social cognitive development

Introduction

Young children are a danger to themselves. Accidents involving a wide range of circumstances – like falls, cuts, burns, ingestion of poisons, or drowning – are the leading cause of child injury and death (Agran et al., 2003). It is thus unsurprising that caregivers invest time, energy, and resources in protecting children from their own dangerous actions (e.g., Dahl, 2016). This protection also involves another cost: To enact protection, caregivers must counter their child's immediate wishes, such as grasping an intriguing yet sharp knife, potentially upsetting them. This raises a question: Do children appreciate protective hindering? Or do they disapprove of any action that involves hindering a child's goal pursuit and achievement, even if it also prevents an overtly dangerous situation? Across two preregistered experiments, we investigated 4–5-year-old children's third-party evaluations of caregivers who help and hinder their children from achieving safe and dangerous goals.

Typically, children endorse helping and disapprove of hindering. One account of these evaluations is that children, and potentially infants, use expectations of rational action to identify acts of helping as those intended to promote others' goals and identify hindering as actions intended to block others' goals (Hamlin et al., 2013; Jara-Ettinger et al., 2016; Schlingloff-Nemecz et al., 2023). This account is corroborated by evidence that children's evaluations take intentions into account. Confronted with adults who fail to help either because they are unwilling or unable, 1-year-old toddlers preferentially help the unable actor (Behne et al., 2005; Dunfield & Kuhlmeier, 2010). Infants may preferentially reach for or look at actors who intend to help, whether or not they are successful (Hamlin, 2013; Hamlin et al., 2013; Woo & Spelke, 2023a, 2023b; Woo et al., 2017). And children incorporate intentions into their explicit endorsements of helping and judgments of harm (Cushman et al., 2013; Van de Vondervoort & Hamlin, 2017; though see Lucca et al., 2025). Thus, even early in life, humans' third-party social evaluations are sensitive to the kind of outcome an actor intends to bring about.

Sensitivity to goals and intentions is not sufficient, however, to provide an appreciation of protective hindering. Oftentimes, a protector is intentionally hindering the outcome they know the target wants because they believe it is bad for the target's ultimate welfare (i.e., safety from injury or harm). Making negative evaluations of risky helping and positive evaluations of protective hindering requires a similar understanding of the good and bad impacts of those actions beyond their effect on goal achievement, as well as an integration of this understanding with social evaluation. Infants and children do seem to recognize dangerous actions as costly, as they use willingness to take risks to reason about others' desires (see Gjata et al., 2022; Liu et al., 2022). However, there is evidence that concepts of safety and danger are still solidifying over the elementary school years (see Crittenden, 2003; Grieve & Williams, 1985; Hill et al., 2000; Peterson et al., 1986; Pfeffer, 1989).

Even if children do represent dangerous outcomes as costly or harmful, this may not impact evaluations of helping and hindering. Young children judge that the obligatory nature of helping extends to immoral goals, such as helping a friend steal (Dahl et al., 2020). Three-year-old children also endorse those who agree to help with antisocial goals; by age 5 children instead evaluate hinderers of antisocial goals more positively (Szarek et al., 2023). Age was also a significant factor in children's negative evaluations of helping that incurred Covid-related risks (Marshall et al., 2023). Thus, young children may not appreciate protection both because they do not always understand the potential negative consequences of pursuing unfamiliar, dangerous goals, and because they do not always integrate their understanding of these consequences into their evaluations.

There is some evidence that young children will themselves hinder others' goals or ignore their requests in order to protect them: 3-year-old children selectively deny

In D. Barner, N.R. Bramley, A. Ruggeri and C.M. Walker (Eds.), *Proceedings of the 47th Annual Conference of the Cognitive Science Society*

providing and warn adults about dysfunctional items (Martin & Olson, 2013), and 5-year-old children override another child's request for a desirable snack (chocolate) that will make them feel sick, so long as they can provide an alternative, still-desired snack (fruit snacks) that will not make them feel sick (Martin et al., 2016). However, children's willingness to override others' goals is still limited: When the desired snack would make the child feel sick, but the alternative is less desirable (carrots), children provide the more desirable yet sickness-inducing snack (Martin et al., 2016). Thus, young children only subvert others' goals if they can help in another way that also seems appealing. This limit on their own helping may, however, underestimate their ability to recognize the value of protective hindering when enacted by a caregiver.

The Current Investigation

To appreciate protective caregivers, children must be able to recognize that helping dangerous goal pursuit risks harm and that hindering such goals leads to better ultimate outcomes. The current investigation tested for this recognition by asking children to evaluate caregivers who helped or hindered their infants' unsafe goals, comparing such evaluations to caregivers who helped and hindered in safe circumstances.

Experiment 1

Methods

Transparency & Openness The procedure and analysis plan for this experiment were preregistered, and all planned components are presented here. The preregistration and experimental materials can be found online via Open Science Framework (https://osf.io/p65nb/).

Participants Thirty-two 4-year-old children ($M = 4.52$, $SD = 0.29$ [7 birthdays not reported]; 13 boys, 11 girls, 0 gender-diverse children [8 not reported]) participated at a local science museum from April through June of 2024. We used pilot data ($N = 16$ 5-year-old children) to conduct a simulation-based power analysis using R package 'simr' (Green & MacLeod, 2016; R Core Team, 2021); given highly consistent pilot data this indicated we would have the power to observe the interaction of interest at $N = 4$. To collect a prudent sample size as well as to fully counterbalance the experiment, we collected data from 32 children.

Following preregistered protocol, 10 participants were excluded for the following reasons: not completing the experiment ($n = 2$), not being fluent in English ($n = 1$), failing scale training questions ($n = 5$), report of atypical development ($n = 1$), experimenter error ($n = 2$), and/or substantial parental interference ($n = 1$). Per parent report, participants were of the following racial and/or ethnic backgrounds: 6.25% African American or Black, 12.50% Asian American or Asian, 3.13% Middle Eastern, 3.13% Hispanic, 25.00% White, 28.13% two or more races and/or ethnicities, 21.88% not reported.

Procedure

Scale Training & Prelude Participants first completed training on two scales: a six-point 'Evaluation' Likert scale [ranging from *Very Bad* (0) to *Very Good* (5)] and a dichotomous 'Goal' scale (with the phrases Safe and Not Safe, respectively accompanied with pictures of a checkmark and of an X). Participants passed scale training if they reported that it is 'safe' to read a book and 'not safe' to play with fire; otherwise, they were excluded from analyses as preregistered. Then, participants were told a prelude to the experimental vignettes, detailing a party in which moms and their infants were in attendance.

Overtly Safe or Not Safe Goal Vignettes & Evaluations Participants heard four experimental vignettes following a 2 (Action: Help vs. Hinder) x 2 (Goal: Safe vs. Not Safe) within-participant design. Each vignette featured a new mom and baby. In the Safe vignettes the baby wanted to play with a basket full of age-appropriate toys. In the Not Safe vignettes the baby wanted to play with a basket full of tools and sharp objects for adults (see Figure 1). In one vignette of each Goal type, the mom helped her son achieve his goal by bringing him closer to the basket. In the other vignette of each Goal type, the mom hindered her son from achieving his goal by carrying him away from the basket. In both helping vignettes, the babies appeared happy after being helped to reach basket; in both hindering vignettes, the babies appeared in distress while getting taken away. Vignette order was blocked by Goal; Goal block order and Action order within the Goal blocks were counterbalanced. At the start of each block, participants confirmed the basket items for that block were safe or not safe for babies to play with. After hearing each vignette, participants used the six-point evaluation scale to evaluate the mom's action.

Ambiguously Safe or Not Safe Goal Vignettes & Evaluations In a final pair of vignettes, participants evaluated caregiver helping and hindering in the context of a (presumably) more ambiguous item: balloons. Participants first judged whether they thought balloons were safe or not safe for infants to play with. Then they evaluated one mom

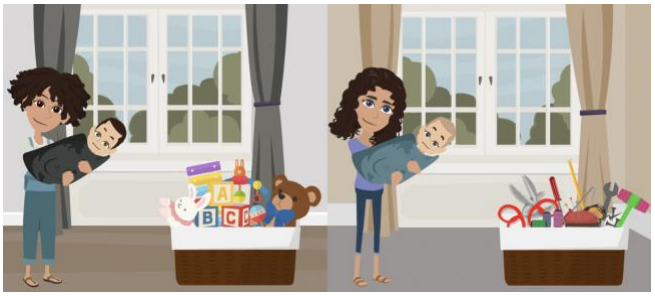

Figure 1: Vignette picture examples in which the infants have an overtly 'safe' goal (left) and an overtly 'not safe' goal (right) of playing with the respective basket items. Images are copyrighted by and used by permission of VYOND, trademark of GoAnimate, Inc.

who helped her infant play with the balloons, and one mom who hindered her infant from playing with the balloons in the context of vignettes like the ones described above (presentation order counterbalanced). Unlike the overtly safe or not safe goal vignettes, these vignettes did not inform participants about whether the moms themselves thought the balloons were safe or not safe for babies to play with.

Results

Analysis Plan All analyses were conducted using R Statistical Software (R Core Team, 2021). For the four overtly safe or not safe trials, we ran a linear regression model containing Evaluation (six-point scale ranging from *Very Bad* = 0, to *Very Good* = 5) as the dependent variable, with the main effects of and interaction between Action (2: Help vs. Hinder) and Goal (2: Safe vs. Not Safe) as predictor variables. A random effect of participant was included to account for the within-participant experimental design. If participants' evaluations of parents' helping and hindering took explicitly stated safety of a goal into account, then the model should yield a significant interaction between Action and Goal. For the two ambiguously safe or not safe trials, we followed a similar analysis plan but replaced the predictor variable Goal with participants' own Safety Judgment (2: Safe vs. Not Safe). To further examine whether inclusion of the interaction in each model improved the model fit, we conducted separate nested model comparisons and calculated Bayes factors for each model.

Then, to unpack the directional pattern of results compared to the inflection between Good and Bad at the midpoint of the scale (= 2.50), we conducted analyses under Frequentist and Bayesian frameworks. We ran two-tailed t-tests, as well as calculated predicted probabilities and estimates along with 95% confidence and credible intervals using R packages "ggeffects" (Lüdecke, 2018) and "brms" (Bürkner, 2017). For Bayesian analyses, we interpreted predicted intervals that were above and did not overlap with the midpoint as evidence of positive evaluations; intervals that were below and did not overlap with the midpoint as evidence of negative evaluations; intervals that overlapped with chance as not providing any particular directional evidence. Unless otherwise noted, the inferences drawn from the Frequentist and Bayesian framework analyses substantiate one another. For the sake of space, we only discuss discrepancies between the Frequentist and Bayesian analyses (for full details, see preregistration).

Overtly Safe or Not Safe Goal Vignettes There was a significant interaction between Action and Goal, χ^2 (1, N = 32) = 85.40, BF > 1,000.00, p < .001, supporting our hypothesis that children's evaluations of helpful and hindering actions attend to the safety of a goal (see Figure 2).

As predicted, participants positively evaluated the Safe goal helper (M = 4.59, 95% CI [4.32, 4.87]; t(31) = 15.67, p < .001) and negatively evaluated the Safe goal hinderer (M =1.06, 95% CI [0.54, 1.58]; t(31) = -5.67, p < .001). Participants also negatively evaluated the Not Safe goal helper (M = 0.84, 95% CI [0.32, 1.37]; t(31) = -6.41, p < .001). Evaluations of the protector, however, yielded mixed results. Under the Frequentist framework analysis, participants did not positively evaluate the parent who protectively hindered their child from achieving the Not Safe goal (M = 3.09, 95% CI [2.34, 3.85]; t(31) = 1.69, p = .101); under the Bayesian framework analysis, the confidence and credible intervals were above and did not overlap with chance (by .06), suggesting the mean of 4-year-old children's evaluations is reliably, though weakly, positive. Erring on the side of caution, we interpret 4-year-olds' evaluations of the protector as being no different from chance responding, or at least not robustly and reliably positive.

Ambiguously Safe or Not Safe Goal Vignettes As expected, the interaction between Action (Help vs. Hinder) and Safety Judgment [whether participants judged ballons as Safe (n = 16) vs. Not Safe (n = 16) for babies to play with] was significant, χ^2 (1, N = 32) = 9.09, BF = 192.13, p = .003. Participants who judged the balloons as safe positively evaluated the helper (M = 4.31, 95% CI [3.59, 5.03]; t(15) = 5.36, p < .001) and negatively evaluated the hinderer (M = 1.06, 95% CI [0.12, 2.01]; t(15) = -3.25, p = .005). However, participants who judged the balloons as not safe did not have directional evaluations: They were no different from the midpoint in their evaluations of either the helper (M = 2.44, 95% CI [1.27, 3.60]; t(15) = -0.11, p = .911) or the hinderer (M = 2.13, 95% CI [0.90, 3.35]; t(15) = -0.65, p = .525).

Discussion

Experiment 1 demonstrates that 4-year-old children's evaluations of helping and hindering are sensitive to the safety or danger of the helpee's goal. However, they did not robustly approve of protective caregivers, despite the ubiquity of this role in early childhood caregiving. In order to

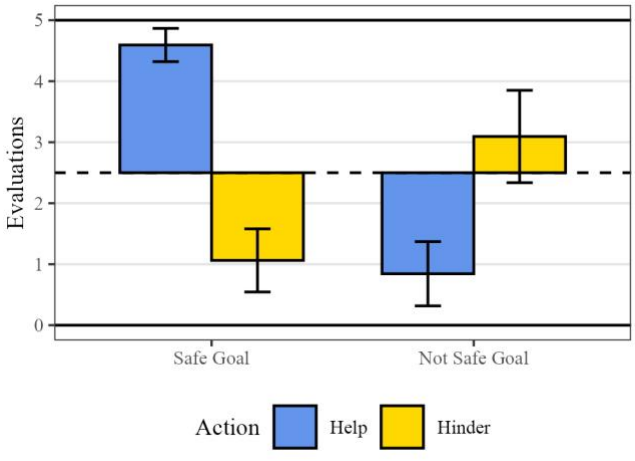

Figure 2: 4-year-olds' evaluations of helpful and hindering caregivers for overtly safe and not safe goals. Bars represent means; error bars represent 95% CI. The dashed line at 2.50 represents chance responding.

characterize the developmental trajectory of evaluations of protective hindering, and to examine if findings replicate across another experimental modality, we conducted another experiment using the same paradigm with participants from a larger age range (4 and 5 years of age) tested online.

Experiment 2

Methods

Participants Forty-eight 4–5-year-old children (M = 5.01, SD = 0.56, range = 4.05–5.92; 20 boys, 27 girls, 1 gender-diverse child) were recruited through Children Helping Science and participated via Zoom from August through September of 2024. Per parent report, participants were of the following racial and/or ethnic backgrounds: 8.33% Asian American or Asian, 2.08% Hispanic, 2.08% South American, 64.58% White, 22.92% two or more races and/or ethnicities.

Procedure Following the same procedure as Experiment 1, participants first evaluated helpers and hinderers in the context of overtly safe and unsafe goals and then evaluated a helper and hinderer for a more ambiguous goal. The only difference was that the ambiguous help and hinder trials were presented in the same order as the preceding block.

Results

Analysis Plan We planned to first test the critical interaction and pattern of results between Action (2: Help vs. Hinder) and Goal (2: Safe vs. Not Safe) across the entire 4–5-year-old sample, separately for the overtly and ambiguously safe or not safe vignettes. Then, to consider the possibility that there may be developmental differences, we included Age (continuous in years; e.g., 4.71) as a predictor and interaction variable to the original models. We planned that if the three-way interaction between Action, Goal, and Age was significant, we would then separately analyze the pattern of results in 4-year-olds and 5-year-olds.

Overtly Safe or Not Safe Goal Vignettes

All participants As predicted, there was a significant interaction between Action and Goal, χ^2 (1, N = 48) = 200.93, BF > 1,000.00, p < .001. Participants negatively evaluated the Safe goal hinderer (M = 1.08, 95% CI [0.67, 1.49]; t(47) = -6.95, p < .001) and the Not Safe goal helper (M = 0.27, 95% CI [0.07, 0.48]; t(47) = -21.85, p < .001); they positively evaluated the Safe goal helper (M = 4.48, 95% CI [4.15, 4.81]; t(47) = 11.94, p < .001) and the Not Safe goal hinderer (M = 3.94, 95% CI [3.43, 4.45]; t(47) = 5.67, p < .001).

Continuous Age Interaction The three-way interaction between Action, Goal, and Age was statistically significant, χ^2 (1, N = 48) = 8.52, BF = 102.42, p = .004, indicating children's relative evaluations of safe and not safe goal helpers and hinderers change with age (see Figure 3). As preregistered, we then separately analyzed 4-year-old and 5-year-old participants' evaluations.

Four-year-old participants The 4-year-old only model revealed a significant Action and Goal interaction, χ^2 (1, N = 24) = 78.15, BF > 1,000.00, p < .001. As in Experiment 1, this age group positively evaluated the Safe goal helper (M = 4.42, 95% CI [3.94, 4.90]; t(23) = 8.24, p < .001), and negatively evaluated the Safe goal hinderer (M = 1.17, 95% CI [0.55, 1.79]; t(23) = -7.78, p < .001) and Not Safe goal helper (M = 0.42, 95% CI [0.04, 0.79]; t(23) = -11.59, p < .001). Unlike Experiment 1, however, these participants positively evaluated the Not Safe goal protective hinderer (M = 3.46, 95% CI [2.61, 4.30]; t(23) = 2.35, p = .028).

Five-year-old participants The 5-year-old only model also yielded a significant interaction between Action and Goal, χ^2 (1, N = 24) = 133.26, BF > 1,000.00, p < .001. These participants positively evaluated the Safe goal helper (M = 4.54, 95% CI [4.04, 5.04]; t(23) = 8.49, p < .001), and negatively evaluated the Safe goal hinderer (M = 1.00, 95% CI [0.42, 1.58]; t(23) = -5.31, p < .001) and Not Safe goal helper (M = 0.13, 95% CI [-0.06, 0.31]; t(23) = -25.95, p < .001). They also gave robust positive evaluations of the Not Safe goal protective hinderer (M = 4.42, 95% CI [3.85, 4.99]; t(23) = 6.96, p < .001).

Continuous Age Effects Separate models of evaluations for each vignette type, with Age as the sole predictor variable, indicated that only evaluations of the protector changed across development: With age, children more positively evaluated the parent who protectively hindered their infant's Not Safe goal (β = 1.26, p = .005; all other ps > .186).

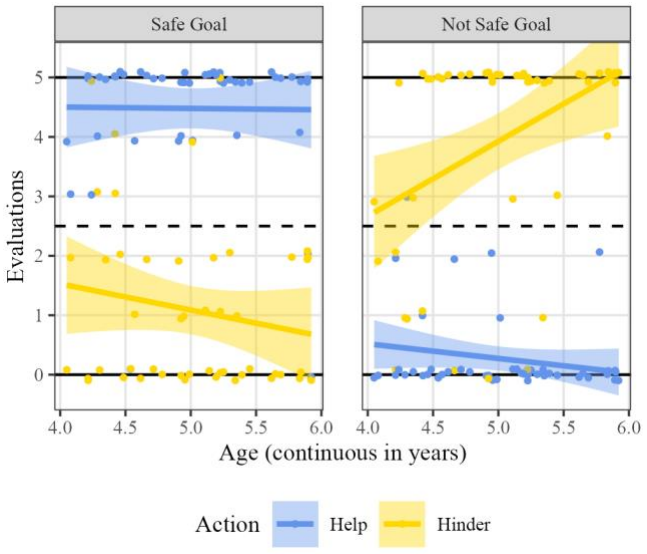

Figure 3: 4–5-year-olds' evaluations of helpful and hindering caregivers for overtly safe and not safe goals. Dots represent individual evaluations, jittered to better illustrate response patterns. Dark lines represent regressions; shaded error bands represent SE. The dashed line at 2.50 represents chance responding.

Ambiguously Safe or Not Safe Goal Vignettes

All participants The interaction between Action (Help vs. Hinder) and Safety Judgment [whether participants judged the balloons as Safe (n = 35) vs. Not Safe (n = 13)] for data from all participants was statistically significant, χ^2 (1, N = 48) = 47.41, BF > 1,000.00, p < .001.

Participants who judged the balloons as safe for babies to play with positively evaluated the helper (M = 4.51, 95% CI [4.20, 4.83]; t(34) = 12.96, p < .001) and negatively evaluated the hinderer (M = 0.89, 95% CI [0.42, 1.36]; t(34) = -6.99, p < .001). Those who judged the balloons as not safe for babies to play with were no different from chance when evaluating the helpful caregiver under the Frequentist framework analysis (M = 1.62, 95% CI [0.44, 2.79]; t(12) = -1.65, p = .126), though the Bayesian framework analysis indicates they made a negative evaluation (i.e., intervals were below and not overlapping with chance). Critically, participants who judged ballons as not safe for a baby to play with were no different from the midpoint in their evaluations of the protective hinderer (M = 3.08, 95% CI [1.78, 4.37]; t(12) = 0.97, p = .350). Notably, few children in Experiment 2 judged the balloons as not safe (n = 13), so this pattern of results should be interpreted with caution.

Continuous Age Interaction There was a statistically significant interaction between Action, Safety Judgment, and Age, χ^2 (1, N = 48) = 8.54, BF = 153.76, p = .003, indicating their evaluations change with development. To further unpack the results, we analyzed the results separately for 4-year-olds and 5-year-olds.

Four-year-old participants The 4-year-old only model yielded a significant interaction between Action and Safety Judgment, χ^2 (1, N = 24) = 7.54, BF = 92.33, p = .006. Participants who judged the balloons as safe (n = 18) positively evaluated the helper (M = 4.50, 95% CI [3.95, 5.05]; t(17) = 7.73, p < .001) and negatively evaluated the hinderer (M = 0.83, 95% CI [0.08, 1.58]; t(17) = -4.70, p < .001). Replicating the results from Experiment 1, participants who judged the balloons as not safe (n = 6) did not demonstrate directional evaluations of the helper (M = 2.50, 95% CI [0.65, 4.35]; t(5) = 0.00, p = 1.00) or the protector (M = 1.50, 95% CI [-0.57, 3.57]; t(5) = -1.24, p = .270).

Five-year-old participants The 5-year-old only model revealed a significant interaction between Action and Safety Judgment, χ^2 (1, N = 24) = 53.65, BF > 1,000.00, p < .001.

Participants who judged the balloons as safe (n = 17) positively evaluated the helper (M = 4.53, 95% CI [4.16, 4.90]; t(16) = 11.66, p < .001) and negatively evaluated the hinderer (M = 0.94, 95% CI [0.30, 1.58]; t(16) = -5.15, p < .001). Those who judged the balloons as not safe (n = 7) demonstrated mixed evidence in their evaluations of the helper. Under the Frequentist analysis, their evaluations were no different from the midpoint (M = 0.86, 95% CI [-0.87, 2.58]; t(6) = -2.33, p = .059); under the Bayesian analysis, their evaluations were below and not overlapping with the midpoint, indicative of a negative evaluation. Critically, however, their evaluations of the protector, or the not safe goal hinderer, were positive (M = 4.43, 95% CI [3.38, 5.48]; t(6) = 4.50, p = .004). However, given the very small sample, these findings should be interpreted with caution.

Continuous Age Effects Simple linear regression models of evaluations for each vignette category (broken down by Action type and Safety Judgment), with Age as the sole predictor variable, did not reveal any statistically significant change across development (ps > .058).

Discussion

Experiment 2 provides further evidence of how young children's evaluations of helping and hindering hinge on the safety of the relevant goal. The results were broadly similar to Experiment 1, providing further validation of online testing modalities (see Chuey et al., 2021, 2024). In the context of overtly dangerous goals, however, the expanded age range revealed that children's evaluations of protective hinderers become substantially more positive between 4 and 5 years. However, participants who judged the ambiguous goal as unsafe again did not demonstrate directional evaluations: neither 4- nor 5-year-old children gave reliably positive or negative evaluations of helpful or hindering caregivers for a goal participants themselves judged as unsafe, though there was some preliminary evidence of 5-year-olds positively evaluating protective hinderers.

General Discussion

Two preregistered experiments provide evidence that 4–5-year-old children's evaluations of helpful and hindering caregivers attend to the safety and danger of the goal in question. Across Experiments 1 and 2, participants positively evaluated a caregiver who helped and negatively evaluated a caregiver who hindered their infant's overtly safe goal, replicating and extending previous findings documenting the early foundations for human social evaluation (Hamlin, 2013; Hamlin et al., 2013; Jara-Ettinger et al., 2016; Marshall et al., 2023; Schlingloff-Nemecz et al., 2023; Szarek et al., 2023; Van de Vondervoort & Hamlin et al., 2017; Woo & Spelke, 2023a, 2023b; Woo et al., 2017). Switching to a dangerous goal context had a large effect on evaluations of helpers: Participants in both experiments gave strongly negative evaluations of caregivers who helped their infants reach dangerous objects, with no evidence of age-related change. In contrast, evaluations of a caregiver who protectively hindered their infant's dangerous goal were less consistent and showed more substantial developmental change: There was conflicting evidence as to whether 4-year-old children endorsed the protective caregiver, but these evaluations became more positive with age, and 5-year-old children gave robustly positive evaluations of the protective caregiver's actions.

When a goal's safety or danger was more ambiguous, and the caregiver's belief about its safety was unknown, there were intriguing findings: Children who judged the

ambiguous goal as "safe" for infants to enact positively evaluated the helpful caregiver and negatively evaluated the hindering caregiver; children who judged the ambiguous goal as "not safe", however, did not demonstrate consistent evaluations of the helpful and hindering caregivers.

Overall, these findings provide an axis in which children base their social evaluations of helping and hindering. Past experiments have included trials in which helpers and hinderers act intentionally or accidentally, finding general patterns of praise for actual and intended helping and condemnation of actual and intended hindering in the context of safe, standard goals. Here, we demonstrate that children's reasoning about helping and hindering is more nuanced. Similar to findings indicating that 6–10-year-old children understand it is better to abstain from helping in contexts involving disease (Marshall et al., 2023), and that 5-year-old children negatively evaluate helping antisocial goals (Szarek et al., 2023), here we show that children recognize dangerous helping is not good: Even the youngest children tested condemned explicitly not safe helpers.

Young children did not always endorse protective hinderers, however. What could be driving this pattern of evaluation? Children's counterfactual reasoning is still forming over the tested age range (see Byrne, 2016), so one factor may have been an inability to recognize what *could* have happened had the caregiver *not* protectively hindered their child's unsafe goal. Vignettes featuring risky helping more clearly depicted the dangerous outcome for participants. Future investigations can ask children to describe what could happen in a dangerous situation, before asking for their evaluations, to understand if they can imagine a bad outcome and if doing so affects evaluations. Studies should also investigate if children's own helping and hindering shows the same asymmetry evident here: Does resistance to helping with dangerous goals emerge before the tendency to hinder them? Choosing not to aid an unsafe goal seems to depend on understanding the possible harm in advance. However, some evidence suggests young children do successfully incorporate multiple possibilities in their action planning (Turan-Küçük & Kibbe, 2025).

The negative infant emotions depicted after protective hindering also may have affected children's judgments. Even if participants understood the parent's intention was to keep the infant from harm, not to make him cry, young children, at least those raised in Western cultures, are notorious for weighing outcome over intention, and only with age do they come to reason that unintended harms are not necessarily blameworthy (e.g., Cushman et al., 2013; see Barrett & Saxe, 2021). The salience of the hindered infants' distress could thus have driven younger participants' mixed evaluations.

What could be driving discrepancies between the explicitly and ambiguously safe and not safe trials? That is, when told that certain objects would be not safe for infants to play with, children robustly condemned the helpful caregiver and with age came to robustly appreciate the protectively hindering caregiver. However, for more ambiguous objects, ones that participants themselves deemed dangerous, there were not consistent patterns of results. In fact, their reasoning was largely no different from chance. We hypothesize that this may stem from children's self-awareness of their incomplete knowledge of what is safe and dangerous. If the caregiver's actions make them doubt their own judgment, they may no longer know how to evaluate those actions.

Finally, these findings suggest that representations of both specific goals and broader welfare play a role in social evaluation. Even if children do recognize helping as a second-order goal to bring about someone else's goal (Hamlin et al., 2013; Schlingoff-Nemecz et al., 2023, 2025), this is not the sole basis for their judgments about which actions are good or bad. They clearly also consider the overall consequences of the action for both the target and those around them (see Szarek et al., 2023) and positively evaluate those whose actions indicate concern for others' broader welfare (Powell, 2022).

Historically, research on caring for young children has primarily focused on adults' concepts, enactment, and thoughts. Less research, however, has addressed how young humans themselves conceptualize and think about the care they receive from others (see Gopnik, 2023; Thomas et al., 2025). This is surprising, given the fact that such questions can provide key insights as to why younger humans and those providing care to them experience conflict over seemingly cut-and-dry scenarios. The present findings provide evidence to suggest one reason why young children and their caregivers experience problems over enacted protection: Even in third-party scenarios involving no personal costs, 4-year-old children did not reliably positively evaluate a caregiver who knowingly hinders their infant from playing with overtly dangerous objects, like sharp scissors and tools. With age, however, children's evaluations become robustly positive, indicating that with further life experience and cognitive development children come to revere caregiver protection. Further research is needed to uncover whether and how these judgments translate to real-world behaviors and cognitive abilities. For instance: What kinds of interventions can help young children understand their caregiver is prioritizing more important goals (e.g., safety and wellbeing) over more immediate, albeit safe or dangerous, goals (e.g., interacting with a novel object like a knife that could result in bad outcomes)?

Conclusion

The present investigation provides novel insights regarding children's evaluations of a ubiquitous behavior they themselves likely experience in their everyday lives: caregiver protection. With age and only for overtly dangerous scenarios, children positively evaluated caregivers who hindered their infants' goal pursuit and achievement. Future research is necessary to understand how children experience and think about caregiving to support healthy and strong parent-child relationships. Moreover, by providing a developmental basis of a concept of protection, we can better understand and support children in detecting and affiliating with protectors who can help them safely navigate the world.

Acknowledgments

We thank the following for making this research possible: our participating families; coordinators at Fleet Science Center and Children Helping Science for supporting participant recruitment and data collection; Eliana Enriquez, Wesley Ge, Ethan Gurevich, Saminyasar Islam, and Katherine Kim for research assistance; Adena Schachner, Alexis Smith-Flores, Katie Vasquez, Zoe Liberman, and Salih Özdemir for discussion; members of the Social Cognition and Learning Laboratory and many others of the Departments of Psychology and Cognitive Science at UC San Diego. This research was funded by a National Science Foundation grant DGE-2038238 awarded to RT and a Hellman Foundation Fellowship awarded to LJP.

References

Agran, P. F., Anderson, C., Winn, D., Trent, R., Walton-Haynes, L., & Thayer, S. (2003). Rates of pediatric injuries by 3-month intervals for children 0 to 3 years of age. *Pediatrics*, *111*(6), e683–692.

Barrett, H. C., & Saxe, R. (2021). Are some cultures more mind-minded in their moral judgments than others? *Transactions of the Royal Society B: Biological Sciences*, *376*(1838), Article 20200288.

Behne, T., Carpenter, M., Call, J., & Tomasello, M. (2005). Unwilling versus unable: Infants' understanding of intentional action. *Developmental Psychology*, *41*(2), 328–337.

Dunfield, K. A., & Kuhlmeier, V. A. (2010). Intention-mediated selective helping in infancy. *Psychological Science*, *21*(4), 523–527.

Byrne, R. M. J. (2016). Counterfactual thought. *Annual Review of Psychology*, *67*, 135–157.

Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using stan. *Journal of Statistical Software*, *80*(1), 1–28.

Chuey, A., Asaba, M., Bridgers, S., Carrillo, B., Dietz, G., Garcia, T., Leonard, J. A., Liu, S., Merrick, M., Radwan, S., Stegall, J., Velez, N., Woo, B. M., Wu, Y., Zhou, X. J., Frank, M. C., & Gweon, H. (2021). Moderated online data-collection for developmental research: Methods and replications. *Frontiers in Psychology*, *12*, Article 734398.

Chuey, A., Boyce, V., Cao, A., & Frank, M. C. (2024). Conducting developmental research online vs. in-person: A meta-analysis. *Open Mind*, *8*, 795–808.

Crittenden, P. M. (1999). Danger and development: The organization of self-protective strategies. *Monographs of the Society for Research in Child Development*, *64*(3), 145–171.

Cushman, F., Sheketoff, R., Wharton, S., & Carey, S. (2013). The development of intent-based moral judgment. *Cognition*, *127*(1), 6–21.

Dahl, A. (2016). Mothers' insistence when prohibiting infants from harming others in everyday interactions. *Frontiers in Psychology*, *7*, Article 1448.

Dahl, A., Gross, R. L., & Siefert, C. (2020). Young children's judgments and reasoning about prosocial acts: Impermissible, suberogatory, obligatory, or supererogatory? *Cognitive Development*, *55*, Article 100908.

Gjata, N. N., Ullman, T. D., Spelke, E. S., & Liu, S. (2022). What could go wrong: Adults and children calibrate predictions and explanations of others' actions based on relative reward and danger. *Cognitive Science*, *46*(7), Article e13163.

Gopnik, A. (2023). Caregiving in philosophy, biology and political economy. *Daedalus*, *152*(1), 58–69.

Green, P., & MacLeod, C. J. (2016). simr: An R package for power analysis of generalised linear mixed models by simulation. *Methods in Ecology and Evolution*, *7*(4), 493–498.

Grieve, R. & Williams, A. (1985). Young children's perception of danger. *British Journal of Developmental Psychology*, *3*(4), 385–392.

Hamlin, J. K. (2013). Failed attempts to help and harm: Intention versus outcome in preverbal infants' social evaluations. *Cognition*, *128*(3), 451–474.

Hamlin, J. K., Ullman, T., Tenenbaum, J. B., Goodman, N., & Baker, C. (2013). The mentalistic basis of core social cognition: Experiments in preverbal infants and a computational model. *Developmental Science*, *16*(2), 209–226.

Hamlin, J. K., Wynn, K., Bloom, P., & Mahajan, N. (2011). How infants and toddlers react to antisocial others. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(50), 19931–19936.

Hill, R., Lewis, V., & Dunbar, G. (2000). Young children's concepts of danger. *British Journal of Developmental Psychology*, *18*(1), 103–119.

Jara-Ettinger, J., Gweon, H., Schulz, L. E., & Tenenbaum, J. B. (2016). The naïve utility calculus: Computational principles underlying commonsense psychology. *Trends in Cognitive Sciences*, *20*(8), 589–604.

Liu, S., Pepe, B., Kumar, M. G., Ullman, T. D., Tenenbaum, J. B., & Spelke, E. S. (2022). Dangerous ground: One-year-old infants are sensitive to peril in other agents' action plans. *Open Mind*, *6*, 211–231.

Lucca, K., Yuen, F., Wang, Y., Alessandroni, N., Allison, O., Alvarez, M., Axelsson, E. L., Baumer, J., Baumgartner, H. A., Bertels, J., Bhavsar, M., Byers-Heinlein, K., Capelier-Mourguy, A., Chijiiwa, H., Chin, C. S. S., Christner, N., Cirelli, L. K., Corbit, J., Daum, M. M., …, Hamlin, J. K. (2025). Infants' social evaluation of helpers and hinderers: A large-scale, multi-lab, coordinated replication study. *Developmental Science*, *28*(1), Article e13581.

Lüdecke, D. (2018). ggeffects: Tidy data frames of marginal effects from regression models. *The Journal of Open Source Software*, *3*(26), Article 772.

Marshall, J., Lee, Y., Deutchman, P., Wang, Z., Horsey, C. D., Warneken, F., & McAuliffe, K. (2023). When not helping is nice: Children's changing evaluations of helping during COVID-19. *Developmental Psychology*, *59*(5), 953–962.

Martin, A. & Olson, K. R. (2013). When kids know better: Paternalistic helping in 3-year-old children. *Developmental Psychology*, *49*(11), 2071–2081.

Martin, A., Lin, K., & Olson, K. R. (2016). What you want versus what's good for you: Paternalistic motivation in children's helping behavior. *Child Development*, *87*(6), 1739–1746.

Peterson, L., Mori, L., & Scissors, C. (1986). Mom or Dad says I shouldn't: Supervised and unsupervised children's knowledge of their parents' rules for home safety. *Journal of Pediatric Psychology*, *11*(2), 177–188.

Pfeffer, K. (1989). Children's awareness and understanding of dangers at home. *Current Psychology*, *8*(4), 307–315.

Powell, L. J. (2022). Adopted utility calculus: Origins of a concept of social affiliation. *Perspectives on Psychological Science*, *17*(5), 1215–1233.

R Core Team. (2021). *A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing.

Schlingloff-Nemecz, L., Pomiechowska, B., Tatone, D., Revencu, B., Mészégető, D., & Csibra, G. (2025). Young children's understanding of helping as increasing another agent's utility. *Open Mind*, *9*, 169–188.

Schlingloff-Nemecz, L., Tatone, D., & Csibra, G. (2023). The representation of third-party helping interactions in infancy. *Annual Review of Developmental Psychology*, *5*, 67–88.

Szarek, K. M., Baryla, W., & Wojciszke, B. (2023). Is helping always morally good? Study with toddlers and preschool children. *Developmental Psychology*, *59*(5), 918–927.

Thomas, A. J., Steele, C., Saxe, R., & Gopnik, A. (2025). How do infants experience caregiving? *Daedalus*, *154*(1), 14–35.

Turan-Küçük, E. N., & Kibbe, M. M. (2025). Three-and four-year-old children represent mutually exclusive possible identities. *Journal of Experimental Child Psychology*, *249*, Article 106078.

Van de Vondervoort, J. W. & Hamlin, J. K. (2017). Preschoolers' social and moral judgments of third-party helpers and hinderers align with infants' social evaluations. *Journal of Experimental Child Psychology*, *164*, 136–151.

Woo, B. M. & Spelke, E. S. (2023a). Infants and toddlers leverage their understanding of action goals to evaluate agents who help others. *Child Development*, *94*(3), 734–751.

Woo, B. M. & Spelke, E. S. (2023b). Toddlers' social evaluations of agents who act on false beliefs. *Developmental Science*, *26*(2), Article e13314.

Woo, B. M., Steckler, C. M., Le, D. T., & Hamlin, J. K. (2017). Social evaluation of intentional, truly accidental, and negligently accidental helpers and harmers by 10-month-old infants. *Cognition*, *168*, 154–163.