Article:
Anderson, RA orcid.org/0000-0002-6285-8358, Heck, IA, Young, K et al. (1 more author) (Cover date: September 2023) Development of beliefs about censorship. Cognition, 238. 105500. ISSN 0010-0277

https://doi.org/10.1016/j.cognition.2023.105500

Development of Beliefs about Censorship

Rajen A. Anderson

University of Leeds

Isobel A. Heck

University of Chicago

Kayla Young

Cornell University

Katherine D. Kinzler

University of Chicago

Author note: Correspondence concerning this manuscript can be addressed to Katherine D. Kinzler, kinzler@uchicago.edu.

Acknowledgements: We would like to thank Alex Shaw, Radhika Santhanagopalan, Rachel A. King, Mikaela Spruill, Kacie Armstrong, Molly C. Gibian, and Benjamin Morris for helpful feedback on previous versions of this manuscript. We would like to thank Julianna Meely for drawing the pictures used to represent the movies. We would like to thank Lara Akande and Sophie Klopfenstein for assistance with coding the free response explanations.

Open Practices: All preregistration documentation, materials, data, and analysis scripts for these studies are available on the Open Science Framework at https://osf.io/wzjdx

Abstract

Across four studies (total $N = 431$), we examined 5- to 10-year-old children's choices to censor depictions of harm. In all studies, children learned about (fictional) movies that depicted harmful behaviors and decided whether specific audiences should be allowed to watch those movies. In Study 1, children often censored depictions of harms and did so similarly when considering both themselves and another hypothetical child as the viewer. At the same time, children did not censor indiscriminately: Children censored depictions of intentional harms more than accidental harms and, in Study 2, children (and adults; $N = 101$) censored harms (especially intentional ones) more from younger versus older audiences. In Studies 3 and 4, we more directly tested children's motivations for censoring harms, examining dual potential motivations of 1) preventing viewers from feeling sad; and 2) preventing viewers from being inspired to engage in harmful behaviors. We found that children who were motivated to avoid inspiring harmful behaviors were especially likely to censor depictions of harmful intentions. Together, our results indicate that children make sophisticated decisions regarding censorship and underscore an early emerging motivation to disrupt cascades of harmful behavior. These findings hold implications for children's thinking about the psychological and behavioral consequences of harm and for children's thinking about the potential effects of media on themselves and others.

Keywords: morality, censorship, intentionality, social cognition, social cognitive development

Development of beliefs about censorship

"No one pretends that actions should be as free as opinions. On the contrary, even opinions lose their immunity, when the circumstances in which they are expressed are such as to constitute their expression a positive instigation to some mischievous act."

- John Stuart Mill, *On Liberty* (1869)

Young children are evaluators of harm beginning early in life (for reviews, see Dahl & Freda, 2017; Killen & Smetana, 2015). By the preschool years, children judge harms to others' bodies, feelings, and property to be wrong (e.g., Goldstein et al., 2002; Helwig et al., 2001; Heck et al., 2021a; Mulvey et al., 2016; Smetana et al., 2014; Tisak, 1993) and children intervene to stop harms from taking place (Vaish et al., 2011; see also Kanakogi et al., 2017 for evidence that infants prefer those who intervene), in part due to an awareness of the consequences that harms can hold (e.g., Helwig et al., 2001; Mulvey et al., 2016). When harms do occur, children seek to punish those who are responsible (for a review, see Marshall & McAuliffe, 2022), even when doing so comes at a cost to themselves (Yudkin et al., 2019; Leshin et al., 2022).

Despite extensive evidence that young children hold sophisticated understandings of harm, much of this research has focused on children's reasoning about harms at the time that they occur. Yet, instances of harm can hold consequences that extend far into the future (e.g., lasting psychological impacts; the reoccurrence of harmful behavior) and that can have impacts beyond the original victim of the harm (e.g., observers of a harm), raising open questions about whether young children represent the broader and longer-term consequences that harms can hold. Studies in recent years provide hints that children do: For example, between ages 5 and 10, children increasingly recognize that the memory of an emotional harm may linger (Heck et al., 2021a). Moreover, children engage in third-party punishment based not only on retributive motives (e.g., making the perpetrator feel bad or ensuring that they receive their "just deserts")

but also on consequentialist ones (e.g., efforts to deter the perpetrator from engaging in future harmful behaviors; Marshall et al., 2020).

Here, we build on emerging evidence that children can represent the broader and longer-lasting psychological and behavioral consequences of harm—beyond the initial time and specific individuals involved—by examining the context of censorship. In its broadest form, censorship is the prevention of information transfer, often because the censor deems the information immoral, harmful, or otherwise inappropriate for its audience (Coetzee, 1996; Purdy, 2009). Censorship offers a novel domain to understand how children reason about the longer-term consequences that harms can generate. This includes impacts beyond the initial individuals depicted (e.g., the audience viewing the harm) and for a duration that extends beyond when the initial harm has occurred (e.g., in the creation of subsequent negative feelings or behaviors on the part of the viewer). We examined whether children intervene on their own and others' viewing of depictions of harm, and if so, what motivations might underlie children's decisions to do so. We predicted that children would prevent others from viewing depictions of harm beginning early in life, and that children's decisions to censor might reflect motivations both (1) to protect viewers from feeling negative emotions (i.e., *emotional* consequences) and (2) to disrupt potential cascades of future harmful behaviors (i.e., preventing viewers from being inspired to engage in the depicted behaviors; *behavioral* consequences).

Despite censorship being a reoccurring issue in human social life, there is surprisingly little empirical work on the psychological mechanisms motivating censorship and how people decide if something is too harmful for an audience to view (for exceptions, see Anderson & Masicampo, 2017; Chung & Moon, 2016; Davison, 1983; McLeod et al.,1997). The present work thus also aims to help fill this broader empirical and theoretical gap. We posit that a developmental angle on questions about the psychology of censorship can provide particularly

valuable insights. There has been a recent burst of work using a developmental lens to understand intuitive thinking about societal laws and systems (e.g., Bregant et al., 2019; Dunlea & Heiphetz, 2021; Patterson et al., 2019; Heck et al., 2021b; Reifen-Tagar & Cimpian, 2022). Embracing this approach, we examine intuitive thinking about censorship before children are likely to have knowledge of formalized versions of censorship (e.g., governmental regulations). This approach thus allows for understanding whether people's motivations to censor require extensive socialization, or whether motivations to censor may be rooted in early developing socio-cognitive processes and understandings of harm. In the present paper, we focus particularly on children's decisions about whether to censor movies and ask three main questions: (1) Do 5- to 10-year-old children censor depictions of harm *at all*? (2) If so, when in life might this tendency emerge? (3) If children censor, why might children do so?

With respect to whether children intervene on others' viewing of harm, one possibility is that children might simply allow audiences to watch whatever they want, without censoring anything. This would suggest that the motivation to censor depictions of harm might not be rooted in early thinking about harm, or that young children to do not yet understand the consequences that viewing depictions of harm can hold for an audience. Indeed, depictions of harm are not themselves harmful, at least not in a direct physical sense: Representing the potential consequences of viewing depictions of harm may thus be a complex task, especially for younger children. Children may also feel that they themselves do not enjoy being told what to do or to watch, and so might wish to avoid infringing on others' viewing behaviors. In addition, surveys indicate that sizable proportions of U.S. adolescents and children have already seen violent, "R-rated" movies (Worth et al., 2008), suggesting that children may consider watching harmful content to be acceptable. A related possibility is that children may restrict others' viewing of harms but not their own: Notably, children often endorse fairness norms for others,

yet do not always hold themselves to the same standards (see Blake et al., 2014; Smith et al., 2013). To test this possibility in the context of censorship, we varied in Study 1 whether children were asked to make censorship decisions for themselves or for another child.

Despite these possibilities, we predicted that even the 5–6-year-olds in our sample (i.e., 5-year-olds) would censor depictions of harm. By at least age 5, children view emotional harms as wrong (Heck et al., 2021a; Helwig et al., 2001; Yoo & Smetana, 2019) and understand that harms can have consequences that extend beyond the immediate moment the harm has occurred (see Marshall et al., 2020; Heck et al., 2021a). In addition, prior research suggests that at least in the U.S., children consider the moral implications of a piece of speech when considering whether to regulate said speech (Helwig, 1995, 1998; Helwig & Prencipe, 1999). We thus predicted that even the 5–6-year-olds in our sample would be capable of representing the consequences that depictions of harm can hold for viewers of the depiction.

If children do indeed choose to censor depictions of harm from others, then it is possible that children may engage in indiscriminate censorship of any depictions of harm. Even preschool-age children are highly motivated to protect others from feeling sad (Chernyak & Kushnir, 2014), and both children and adults are averse to causing harm (Killen & Smetana, 2015; Cushman et al., 2012; Grossman, 1995). Moreover, both children and adults have a bias to attend to threatening and negative stimuli over nonthreatening and positive stimuli (e.g., Hansen & Hansen, 1988; Kinzler & Shutts, 2008; Schwartz et al., 1985; for a review, see Vaish et al., 2008). Thus, one possibility is that children may restrict any depictions of harmful outcomes simply because viewing those outcomes could make others "feel sad."

We predicted, though, that children might censor depictions of harm in more nuanced ways. In particular, we thought that children's censorship choices may relate to their underlying psychological motivations. Past research on censorship from across a range of domains shows

that adults' motivations to censor are predicted by the extent to which they believe others will be impacted by the material in question (e.g., Anderson & Masicampo, 2017; Chung & Moon, 2016; Davison, 1983; McLeod et al., 1997). These predicted consequences on the audience may take at least two forms (which are not necessarily mutually exclusive). One potential consequence involves the emotional consequences of viewing depictions of harm—viewing depictions of harm may result in a negative emotional response in the audience. Censors—including children in this role—may thus be motivated to reduce the experiences of distress on the audience (Feinberg, 1985; Mill, 1869).

A second potential consequence of viewing depictions of harm involves potential behavioral consequences on the part of the viewer—viewing depictions of harm may inspire cascades of future harms by inspiring viewers to themselves engage in harmful behaviors. Suggestive evidence of this psychological motivation for censorship decisions among adults comes from work by Anderson and Masicampo (2017), who examined what values drive adults' attitudes toward movie censorship from children. They found that adults' intentions to censor were not driven by moral care and condemnation of harm, or from how immoral people found the harms to be, but instead by how much adults valued the moral innocence and sanctity of the audience. This suggests that censorship derives its psychological motivation at least in part from a desire to protect the moral character of the audience. Both children and adults may intuit that watching depictions of harm may lead viewers to commit their own harms in the future: harm perpetuates harm.

We predicted that young children's decisions to censor would reflect motivations both to avoid potential emotional consequences and potential behavioral consequences. We make this prediction in line with recent work (e.g., Marshall et al., 2020), which finds that children punish perpetrators of harm not only based on retributive motives (e.g., making the perpetrator feel bad

or ensuring that they receive their "just deserts") but also based on consequentialist ones (e.g., efforts to deter the perpetrator from engaging in future harmful behaviors). Further, we predicted that, if children censor harms in part to avoid potential behavioral consequences (i.e., the possibility of inspiring viewers to themselves engage in harmful behaviors), then children might be particularly likely to censor depictions of *intended* (vs. accidental) harms. Even toddlers are more likely to mimic intended (vs. accidental) acts (Gergely et al., 2002), and prior work suggests children expect intended emotional harms—but not accidental ones—to reoccur in the future (Heck et al., 2021a). Following this logic, we thought that children might be especially focused on censoring depictions of the desire to do harm when motivated to avoid inspiring potential behavioral consequences.

Finally, we predicted that children's rates of censorship—and their motivations for censoring—may depend on the specific audience whose viewing they were considering censoring. Previous work on censorship across a variety of contexts has found that, among adults, concerns for the audience are important to decisions to censor, with adults being more likely to censor information from more vulnerable, impressionable audiences than from less vulnerable audiences (Anderson & Masicampo, 2017; Chung & Moon, 2016; Davison, 1983; McLeod et al., 1997). Children may learn that different audiences are given different affordances in viewing media and then apply those considerations when making decisions about what can be watched. Children may also think that certain audiences (e.g., younger audiences) are more susceptible to a "corrupting" influence from harmful media and thus require greater or different oversight. Further, if children view younger audiences as particularly impressionable, then they may be especially likely to prevent younger audiences from viewing depictions of intentional harms.

Present Research

In the present research, we examined young children's decisions to censor depictions of harm from others as a lens into the psychological underpinnings of people's decisions to censor in the real world. Across four studies, we presented 5- to 10-year-old children with vignettes that described different fictitious movie scenes depicting harms (e.g., physical harms, property harms). For each depicted movie scene (represented by a static image), children decided whether a particular audience should be allowed to watch that scene. Although this domain certainly does not capture all cases of censorship, we chose to focus on children's decisions to censor movies in particular because this case of censorship is one with which we expected young children to be familiar. Across studies, we examined whether children censor depictions of harm when given the opportunity, what sorts of harms they censor, whom they censor for, and what motivations children have for censoring. In Study 1, we examined whether children would allow or prevent themselves and others from viewing depictions of harm. We included depictions of intentional and accidental harms both as a test of whether children censor indiscriminately (or more selectively), and to begin investigating the motivations underlying children's decisions to censor. In Study 2, we examined effects of audience age on children's censorship decisions and tested whether children may differentially consider the intentionality of depicted harms when considering audiences of different ages. In Studies 3 and 4, we more directly manipulated children's motivations (i.e., to avoid potential emotional consequences or to avoid potential behavioral consequences) and examined the impact of doing so on children's rates of censoring accidental versus intentional harms. Across studies, we predicted that children, when given the option, would censor depictions of harmful acts, and that children would be particularly likely to censor intentional harms when motivated to avoid propagating future harmful behavior. Throughout the studies, we probed children's censorship decisions as well as themes that arose

in children's explicit justifications of their choices. All data, analysis scripts, preregistration documentation, materials, and supplemental study details can be found at https://osf.io/wzjdx/.

Study 1

Study 1 served as an initial examination of whether children censor at all, and if so, of the psychological motivations that may underlie children's doing so (e.g., to avoid potential emotional or behavioral consequences). We presented children with static images representing different fictional movie scenes depicting harmful actions and asked children whether these movies were acceptable to watch. Between-subjects, we asked children to make these decisions for either themselves (*Self* condition) or another same-age, same-gender child (*Other* condition). To examine whether children's motivations to censor might vary across different types of harm, we included a range of potential moral infractions (e.g., Graham, Haidt, & Nosek, 2009; Rai & Fiske, 2011): physical harm (to a person and to an animal); property violation; and purity violation[1].

We additionally varied the intentionality of the depicted harms (within-subjects), such that half of the movie scenes depicted harms done on purpose and half of the movie scenes depicted harms done by accident. We varied the intentionality of the depicted harms for two main reasons: First, given that children's moral judgments and behaviors are sensitive to the intentionality of acts from a young age (e.g., Ames & Fiske, 2015; Chernyak & Sobel; Cushman, 2008; Cushman et al., 2013; Killen et al., 2011; Shultz et al., 1986; Vaish et al., 2009, 2010, 2011), doing so allowed us to examine whether children censor all depictions of harmful outcomes indiscriminately, or whether children may censor more selectively. Second (and perhaps more interestingly), we manipulated the intentionality of the depicted harms to begin

[1] The purity violations we used involved not simply being unhygienic or dirty oneself but negatively acting on another person and their property (i.e., being unhygienic toward them), which could lead to psychological and property harm.

examining children's underlying motivations. We predicted that, if censorship is only motivated by a desire to prevent viewers from feeling sad, then children might focus on censoring depictions of any negative outcomes (regardless of intentions). In contrast, if censorship is also driven by a desire to avoid viewers from witnessing and potentially being inspired to engage in harmful behaviors, then children might be especially concerned with censoring depictions of the intent to harm (thus censoring intentional harms more than accidental ones). To examine children's explicit motivations for censoring harms, we additionally asked participants, after making all decisions, to explain how they decided which movies were acceptable to watch.

Method

Participants

Participants included 96 5- to 10-year-old children (60% girls, 40% boys; M_{age} = 93.12 months; *SD* = 20.32) tested in Central New York, with a roughly even number of children within each age tested (16 5-year-olds, 16 6-year-olds, 16 7-year-olds, 17 8-year-olds, 15 9-year-olds, 16 10-year-olds). We chose this age range for our samples because: (1) children by age 5 can reliably represent others' mental states, which would be necessary to think about their emotional states and future potential behavior (Chobhthaigh & Wilson, 2015; Heck et al., 2021a; Leslie et al., 2006); (2) we wanted to cover a wide age range to examine possible developmental effects; and (3) it is a typical age range used for studying children's thinking about harm (e.g., Bregant et al., 2019; Heck et al., 2021a; Jambon & Smetana, 2014; Shaw et al., 2014).

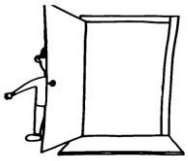
Children participated in a lab or a children's science museum. Based on an *a priori* stopping rule and sample sizes from related research (e.g., Bregant et al., 2019; Heck et al., 2021a), we aimed to recruit at least 40 participants per experimental condition in both this and subsequent studies. To maximize statistical power, additional participants were recruited based on availability during recruitment visits to the museum. Between-subjects, we randomly assigned

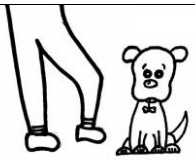
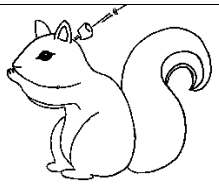
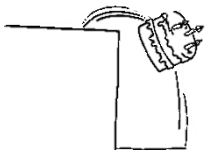
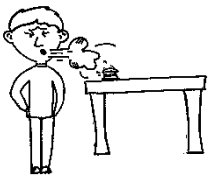
participants to make censorship decisions for themselves (*Self* condition; $n = 50$) or for another fictional same-age, same-gender child (*Other* condition; $n = 46$). Of parents who provided demographic information, 53% identified their children as White; 16% as Black or African American; 13% as Biracial; 8% as Asian or Asian American; 8% as Other; and 2% as Native American. Of parents reporting annual family income, annual income ranged from < $15,000 to > $150,000, with a median family income of $50,000–$75,000. After participating in this study, children answered questions for an unrelated project about free speech.

Materials

We used eight hand-drawn pictures to represent the movies about which children were asked to make decisions (see Table 1). Thus, the stimuli were static images without any movement or sound. We presented each picture one at a time using Microsoft PowerPoint software on a laptop. The images representing the movies depicted four harms with which we expected children to be familiar: physical harm against a person (e.g., this is a movie where a kid opens a door and hits another kid); physical harm against an animal (e.g., this is a movie where a kid kicks a dog); property violation (e.g., this is a movie where a kid rips up another kid's drawing); and purity violation (e.g., this is a movie where a kid leaves moldy food in another kid's room, and it starts to smell). For each movie image description, we created two versions: in one, we described the harm as happening "on purpose" (the *intentional* version), and in the other, as happening "by accident" (the *accidental* version).

Table 1

Harm Type	Image	Stimuli Set	*Accidental* Text	*Intentional* Text
Physical Harm Against a Person		A	*This is a movie where a kid opens a door and hits another kid by accident.*	*This is a movie where a kid opens a door and hits another kid on purpose.*

		B	*This is a movie where a kid hits another kid in the back by accident.*	*This is a movie where a kid hits another kid in the back on purpose.*
Physical Harm Against an Animal		A	*This is a movie where a kid kicks a dog by accident.*	*This is a movie where a kid kicks a dog on purpose.*
		B	*This is a movie where a kid throws a rock and hits a squirrel with the rock by accident.*	*This is a movie where a kid throws a rock and hits a squirrel with the rock on purpose.*
Property Violation		A	*This is a movie where a kid rips up another kid's drawing by accident.*	*This is a movie where a kid rips up another kid's drawing on purpose.*
		B	*This is a movie where a kid pushes another kid's birthday cake to the ground by accident.*	*This is a movie where a kid pushes another kid's birthday cake to the ground on purpose.*
Purity Violation		A	*This is a movie where a kid sneezes on another kid's food by accident.*	*This is a movie where a kid sneezes on another kid's food on purpose.*
		B	*This is a movie where a kid leaves moldy food in another kid's room by accident, and it starts to smell.*	*This is a movie where a kid leaves moldy food in another kid's room on purpose, and it starts to smell.*

Stimuli used across all studies. Children were presented with static images representing eight fictional movie scenes in total. For each Harm Type, participants received the *accidental* text of one movie scene and the *intentional* text of the other movie scene. Children either saw the *accidental* versions for all images from Stimuli Set A and the *intentional* versions for all images from Stimuli Set B, or vice versa.

Design

For each of the four harm types described above, each participant made decisions about two movie scenes (represented by static images; see Table 1). Within each harm type, each participant heard about one movie scene depicting an *intentional* harm and another movie scene depicting an *accidental* harm. For logistical reasons, one movie image from each harm type was designated as Stimuli Set A and the other as Stimuli Set B. Participants either received the *accidental* versions for all images from Stimuli Set A and the *intentional* versions for all images from Stimuli Set B, or vice versa. Each participant was randomly assigned to one of eight pseudo-randomized presentation orders; across participants, each image was presented first an equal number of times. Thus, each participant rated eight total movie images, in a pseudo-randomized order, with four movie images depicting *intentional* harms (one from each harm type) and four movie images depicting *accidental* harms (also one from each harm type).

Procedure

Movie Censorship. To introduce the task, the experimenter read one of two prompts corresponding to the assigned condition (see OSF link for full text). In both conditions, the experimenter first told children about how "sometimes it's OK for a kid to watch a movie, but sometimes someone will say that a movie isn't OK for kids to watch." Next, the experimenter asked children to imagine either that they themselves (in the *Self* condition) or another kid of the same age and gender (in the *Other* condition) had a list of movies that they wanted to watch. Children were told that their job was to say for each movie whether or not that target (themselves or another child) should be allowed to watch that movie. Following these instructions, the experimenter presented participants with the movie images one at a time, with each image representing a different movie scene. For each movie image, the experimenter read the accompanying verbal description of the scene and then asked participants: "Should [you/the

other kid] be allowed to watch this or not allowed to watch this?" After viewing and responding to the eight movie images, the experimenter asked children "Why did you choose the way you did?" and recorded children's responses verbatim.

Explanation Coding. To examine children's explicit justifications for deciding which items to allow, we first filtered out children who did not provide any meaningful or coherent information ("uncodable"; e.g., "I don't know" or "I used my brain"). Here and in all studies, reported analyses and percentages include only the "codable" responses. We coded this subset of responses in three ways (Table 2). First, we coded children's responses based on their focus, coding responses as focusing only on the depicted harms (e.g., "Because all of these movies are horrible"), only on the audience (e.g., "My mom lets me watch as long as I don't repeat anything"), or both. Second, we coded children's responses for whether (or not) they mentioned the intentionality of the depicted acts (e.g., "On purpose. They mean it, so not okay"). Third, we coded children's responses for references to potential consequences of viewing the depicted harms. We categorized such references as pertaining to potential impacts on the audience's behavior (e.g., "He might get bad ideas about being mean"), the audience's emotions (e.g., "Because they could hurt feelings"), both, or neither. Two research assistants unaware of conditions or hypotheses coded children's responses across all studies. Overall agreement between the coders across studies was high (91% mean agreement; mean κ(Cohen) = .78, range from 0.50 to 1.00); any disputes were resolved by the first author.

Table 2
Summary of Explanation Coding Across Experiments

Question	Codes	Rates Per Study			
		Study 1	Study 2	Study 3	Study 4
Focus of explanation	The movies/harms: *"Cause some of them may have violence"*	75%	26.5%	67%	86%
	The audience: *"No teaching kids how to be mean"*	5%	29.5%	5%	5%
	Both: *"If the other kid does something in the movie, she'll start doing the stuff in the movies"*	20%	44%	27%	9%
References to intentions	Yes: *"On purpose they mean it, so not okay"*	46%	25%	43%	68%
	No	54%	75%	57%	32%
References to potential consequences	Negative influence on behavior: *"He might get bad ideas about being mean"*	25%	34%	17%	9%
	Negative influence on emotions: *"Because they could hurt feelings"*	2%	10%	7%	5%
	Both	0%	2%	1%	0
	Neither	73%	54%	75%	86%
References to target age (only in Study 2)	Yes: *"Because I know what little kids and big kids and adults should watch."*		73.5%		
	No		26.5%		

The percentage of codable responses across studies that focused on the depicted harms, the audience, or both and that spontaneously mentioned the intentionality of the acts; potential consequences of viewing the movie scenes on viewers' behaviors and emotions; and the age of the targets (Study 2 only). Quoted statements are examples from participants that qualified for each code. Full open response data for all studies is available at https://osf.io/wzjdx/.

Results

Movie Censorship

We first examined the overall rate at which children censored the depicted harms. On average, across conditions, harm types, and harm intentionality, participants allowed less than

half of the fictional movies[2] (i.e., censored more than half of the movies; M = 2.73/8 movies allowed; $t(93) = 5.63$, $p < .001$).

Next, to examine the potential influences of condition (*Self* vs. *Other*), harm type, and harm intentionality on children's decisions to censor, we conducted a binomial mixed-effects model predicting whether participants allowed each movie. As fixed effects, we included condition (0 = *Self*, 1 = *Other*), harm type (Physical Harm-Person, Physical Harm-Animal, Property Violation, Purity Violation), harm intentionality (0 = *accidental*, 1 = *intentional*), the intentionality X harm type interaction, the intentionality X condition interaction, the specific movie being decided on (to control for idiosyncratic differences in the severity of depicted harms), participant age (in all studies, treated as a continuous variable), and participant gender (0 = girl, 1 = boy). We included random effects for participant and for the order in which movie scenes were presented.

There was no significant effect of audience condition (*Self* vs. *Other*): Participants allowed movies at similar rates when making decisions for themselves (M = 2.74/8) and for another child (M = 2.72/8; OR = 0.75, β = -0.28, SE = 0.53, p = .59). In addition, the rate of allowing movies did not significantly differ based on the harm type depicted (OR = 1.07, β = 0.06, SE = 0.12, p = .63). There was a significant main effect of intentionality, such that participants allowed more movies described as depicting *accidental* harms (M = 1.97/4) than movies described as depicting *intentional* harms (M = .76/4; OR = 0.18, β = -1.69, SE = 0.57, p = .003). Additionally, there was a significant main effect of participant age on the number of movies children allowed, such that older participants allowed more movies on average than

[2] Participants did not view actual movies but instead responded to static images representing fictional movie scenes and made decisions about whether the corresponding movie was allowed or not allowed to be watched. We use the terms "movie scene" and "movie image" to refer to the visual stimuli and the term "movie" to refer to the content about which children made censorship decisions.

younger participants (OR = 1.45, β = 0.37, SE = 0.14, p = .008). There were no significant interactions, including with participant age, and no other significant main effects (all ps > .16).

Explanation Coding

To further examine *why* children censor depictions of harms, we identified themes in children's verbal responses (see Table 2; open response data for this and the following studies are available at https://osf.io/wzjdx/). After excluding 16 children who did not provide codable responses (N = 80 included in explanation coding analyses), we first examined the focus of children's responses. Overall, 75% of children who provided codable responses focused on the depicted harms, 5% focused only on the audience, and 20% of children focused on both the depicted harms and the audience. We examined whether these groups of children responded differently when asked which movies were okay to watch. To do so, we conducted a binomial mixed-effects model, predicting whether children allowed each movie by the focus of children's responses. We additionally included as fixed effects participant age, the intentionality of the depicted harm, and the interaction between harm intentionality and the focus of children's responses. We included a random effect for each participant and for presentation order. This model revealed no significant main or interactive effects of the focus of children's responses on children's rates of censoring harms (all ps > .19).

We additionally examined whether condition (*Self* vs. *Other*) or participant age influenced the focus of children's responses. There was a significant effect of condition (χ^2 (2, N = 80) = 16.69, p < .001): When deciding for themselves, children overwhelmingly focused only on the depicted harms (87% of codable responses); in contrast, when deciding for another child, children more often focused on both the depicted harms and the audience (36% of codable responses). To examine effects of age, we binned children into the youngest participants (ages 5 to 6 years), middle participants (ages 7 to 8), and oldest participants (ages 9 to 10). The focus of

children's responses shifted across the age range tested (χ^2 (4, N = 80) = 13.82, p = .008); with age, children were increasingly likely to focus on both the depicted harms *and* the audience (4% of the 5- to 6-year-olds, 15% of the 7- to 8-year-olds, and 38% of the 9- to 10-year-olds did so). This finding points to the idea that children's reasoning about censorship may become more nuanced over the early childhood years, increasingly incorporating a focus on both the audience and the content involved, as well as the relation between them.

Second, we examined whether children explicitly mentioned the intentionality of the different depicted harms. Overall, 46% of children who provided a codable response spontaneously mentioned intentions and 54% of children did not. Using the same analytic strategy as for the focus of children's responses, we examined whether these groups of children responded differently when asked which movies were okay to watch. There was a significant interaction between references to intentions and the intentionality of the depicted harms (*OR* = 0.05, β = -3.11, *SE* = 0.54, $p < .001$): Participants who mentioned intentionality, compared to those who did not mention intentionality, allowed fewer movies depicting *intentional* harms (M = 0.57/4 vs. M = 0.84/4) and more movies depicting *accidental* harms (M = 2.97/4 vs. M = 1.30/4). As above, we additionally examined whether references to intentions differed across conditions or shifted with age. References to intentions did not differ across conditions (χ^2 (1, N = 80) = 1.34, p = .25) but did shift with age (χ^2 (2, N = 80) = 12.41, p = .002), such that the 5- to 6-year-olds were considerably less likely to reference intentions (17% of codable responses) compared to both the 7- to 8-year-olds (63% of codable responses) and the 9- to 10-year-olds (55% of codable responses).

Third, we examined whether children spontaneously mentioned potential emotional or behavioral consequences of viewing the depicted harms. Overall, 25% of children who provided a codable response mentioned a potential impact on behavior, 2% mentioned a potential impact

on emotions, and 73% mentioned neither potential consequence. Using the same analytic strategy as above, we examined whether these groups of children responded differently when asked which movies were okay to watch. This model revealed no significant effects of children's references to potential consequences on children's rates of censoring harms (all *p*s > .29). We further examined whether children's references to potential consequences varied by condition or changed with age. References to consequences did not differ across conditions (χ^2 (2, N = 80) = 5.08, p = .08) but did shift with age (χ^2 (4, N = 80) = 12.90, p = .01): References to behavioral consequences were increasingly likely with age (provided by 4% of the 5- to 6-year-olds; 22% of the 7- to 8-year-olds; and 45% of the 9- to 10-year-olds).

Discussion

As early as age 5, children in our sample censored depictions of harm when given the opportunity. Moreover, children did so at a relatively high rate, censoring more than half of the fictional movie scenes. In addition, children's rates of censoring depictions of harm did not depend on the type of harm depicted or on whether children were making censorship decisions for themselves or another child. This latter finding is particularly notable given past research showing that children often exhibit a self-serving bias (e.g., Baer & Odic, 2022; Smith et al., 2013; Tasimi & Johnson, 2015) and a general interest in watching violent movies (Worth et al., 2008). Our results from Study 1 thus suggest that motivations to intervene on the potential consequences of viewing depictions of harm may be rooted early in life. Moreover, our finding that children censor depictions of harm builds on recent evidence (e.g., Heck et al., 2021a; Marshall et al., 2020) that children are capable of representing the potential emotional and behavioral consequences that harms—and even witnessed depictions of harm—can hold.

One possibility is that children censored at high rates merely because they like censoring (rather than that they are driven to censor depictions of harm *specifically*). To address this

possibility, we conducted a follow-up study (Supplemental Study 1; reported in detail on our OSF page) in which we asked a new sample of participants to make decisions about content depicting intentional harms in addition to neutral acts (e.g., This is a movie where a kid opens a door) and positive acts (e.g., This is a movie where a kid opens a door for another kid carrying a lot of books). Supporting our prediction that children would not simply censor any movies indiscriminately, we found that participants again allowed relatively few movies that depicted intentional harms (M = 0.64/4) but allowed nearly *all* of the movies that depicted neutral (M = 3.84/4) and positive acts (M = 3.69/4). Therefore, children's decisions to censor appear to be strategic, based on a motivation to intervene on the potential consequences of viewing depictions of *harm*—rather than on a baseline inclination to censor any information presented in our task.

Further, even for depictions of harm, children did not censor indiscriminately. Instead, children appeared to consider the intentionality of the harms when deciding which movies to allow. Beyond demonstrating that children engaged in selective rather than indiscriminate patterns of censorship, this intentionality effect also provides initial evidence that children's censorship decisions might be based not only on concerns about potential emotional consequences but also on potential behavioral consequences. Put another way, children appear to be especially concerned with censoring depictions of the desire to do harm, rather than only depictions of harmful outcomes. This point is further underscored by our finding that nearly a quarter of children spontaneously mentioned motivations to avoid potential behavioral consequences (e.g., inspiring the viewer to themselves engage in harmful behaviors). In fact, references to behavioral consequences were more common than references to impacts on viewers' emotions and children's likelihood of mentioning behavioral consequences increased with age, such that nearly half of the 9- to 10-year-olds who provided a codable response spontaneously mentioned potential behavioral impacts of viewing depictions of harm. A central

goal of Studies 2 to 4 was to further examine this focus on behavioral consequences and to unpack this observed relation between censoring depictions of harmful *intentions* and a motivation to disrupt potential cascades of future harmful behaviors.

Last, Study 1 revealed a shift across ages 5 to 10 in children's rates of censoring depictions of harm. Specifically, younger children were the *most* likely to censor, and with age, children were increasingly less likely to censor. One possibility is that this age effect results simply from older children imagining older audiences, particularly given that we explicitly prompted children in the *Other* condition to think about a same-age child. In Study 2, we examined potential effects of audience age on children's rates of censoring depictions of harm, as well as whether the motivations that underlie children's decisions to censor might vary depending on the specific audience involved.

Study 2

In Study 2, we directly examined the influence of audience age on children's rates of censoring depictions of harm and their motivations for doing so. To further contextualize our findings with children, we also included a comparison sample of adults. We asked participants to make censorship decisions for three targets of different ages: a little kid, a big kid, and a grown-up. We examined two main questions: First, we investigated whether children's rates of censorship may vary depending on the audience involved. Prior research has found that adults are more likely to censor information from vulnerable, impressionable audiences (Anderson & Masicampo, 2017; Chung & Moon, 2016; Davison, 1983; McLeod et al., 1997). Children may similarly think that different audiences may have different reactions to viewing depictions of harm, or that adults have already experienced and been exposed to more than young children (e.g., Taylor et al, 1991). We thus predicted that children would allow fewer movies for the younger targets (e.g., children) than for the older targets (e.g., adults).

Second, we examined whether censorship motivations may vary depending on the audience's age. Specifically, we thought that if people view younger audiences as more "corruptible," then they may be especially motivated to avoid potential behavioral consequences (i.e., inspiring future harmful behaviors) when making censorship decisions for younger audiences (e.g., children). If this is the case, then participants may be especially likely to censor depictions of intentional harms (that depict the *desire* to do harm) from younger vs. older audiences, rather than just censoring information from younger children more generally across the board without regard to intentionality.

To further examine children's motivations—and how they may differ when considering audiences of different ages—we additionally asked child participants a series of more structured questions regarding their censorship motivations, inquiring directly about the presence of motivations to avoid potential emotional consequences and motivations to avoid potential behavioral consequences. We asked children about the presence of each motivation overall, as well as for each audience (i.e., little kid, big kid, and grown-up). Last, we asked children to resolve a censorship dilemma between two fictional kids: One kid who is always good and never gets in trouble and one kid who is sometimes good and sometimes bad and often gets in trouble. We predicted that, if children are more motivated to avoid potential emotional consequences, then children may be more likely to prevent the kid who never gets in trouble from watching the depictions of harm, or perhaps decide between the two kids at even rates. However, if children are more motivated to avoid potential behavioral consequences of viewing depictions of harm, then children may be more likely to prevent the kid who often gets in trouble from watching the violent movies, due to a concern that this kid may be more likely to mimic the harmful behaviors depicted.

Method

Child Participants

Participants included 103 5- to 10-year-old children (49% girls, 51% boys; M_{age} = 93.32 months; *SD* = 19.73) tested in Central New York (19 5-year-olds, 21 6-year-olds, 21 7-year-olds, 20 8-year-olds, 13 9-year-olds, 9 10-year-olds). The sample size for this and the two subsequent studies was based on the effect size of the intentionality effect observed in Study 1 and provides > 90% power. Children participated in a lab, children's science museum, or elementary school. Of parents reporting their children's race/ethnicity, 77% identified their children as White; 12% as Other; 9% Asian or Asian American; and 2% as Black or African American. Of parents reporting family income, annual income ranged from < $15,000 to > $150,000, with a median annual family income of $75,000 – $100,000.

Adult Participants

We additionally recruited 101 adult participants (49% women, 49% men, 2% other; M_{age} = 32.6 years; range 18-85 years; *SD* = 11.6 years; 76% White, 15% Asian, 13% Hispanic or Latin(o/a/e), 7% Black, 2% Other; median gross family income = $50,000 – $74,999; median education level = Bachelor's Degree or equivalent 4-year undergraduate degree). We recruited participants via Prolific.co, an online data collection service (Palan & Schitter, 2018), at a later point in time than when we collected the child sample.

Materials and Design

The materials and design closely matched those from Study 1. We used the same drawings from Study 1 to represent the eight movie scenes. Additional materials included a laminated sheet of paper depicting three stick figures increasing in size from small (i.e., the little kid) to medium (i.e., the big kid) to large (i.e., the grown-up). Adult participants viewed all materials as part of an online survey constructed using Qualtrics.

Procedure

Child Participants.

Movie Censorship. The experimenter first told participants about how "sometimes it's OK for someone to watch a movie, but sometimes someone will say that a movie isn't OK for people to watch." Next, the experimenter asked children to imagine three people—a little kid, a big kid, and a grown-up—and told children that each of these people had a list of movies they wanted to watch. While telling children about these three targets, the experimenter brought out a laminated sheet of paper and pointed to stick figures varying in size from small (i.e., the little kid) to medium (i.e., the big kid) to large (i.e., the grown-up). Children were told that for each movie scene, their job would be to say whether each target should be allowed (or not allowed) to watch that movie. Children were told that, for each movie, they could pick none of the targets, one of the targets, two of the targets, or all three targets. Following these instructions, the experimenter presented children with the eight movie pictures (Table 1). For each movie, participants were asked: "Who should be allowed to watch this movie?" As in Study 1, four of the vignettes depicted accidental harms and four depicted intentional harms. After presenting children with all eight movie scenes, the experimenter asked children "Why did you choose the way you did?" and recorded children's responses verbatim. We coded children's responses in the same way as in Study 1. In addition, we coded whether (or not) children mentioned the age of the targets in their responses (e.g., "The little kid shouldn't watch movies that hurt people or make people sad").

Censorship Motivations. Next, we asked children a series of questions that tested the presence of motivations (1) to avoid potential *emotional* consequences of viewing depictions of harm; and (2) to avoid potential *behavioral* consequences of viewing depictions of harm. Children were told that people sometimes have different reasons for deciding why someone cannot watch a movie, with people sometimes wanting to avoid making the other person feel sad

and sometimes wanting to avoid making the other person act bad. Children were asked what they were thinking about: (1) people feeling sad; (2) people acting bad; (3) or both. After children provided an initial answer, we additionally asked children to think about each target individually: "When picking for the [grown-up/big kid/little kid], were you thinking about that person feeling sad, acting bad, or both?"

Last, to further examine children's motivations for censoring, we introduced children to two fictional characters (depicted as equally sized stick figures): One kid who is always good at school and never gets in trouble, and one kid who is sometimes good and sometimes bad at school and gets in trouble a lot (order counterbalanced across participants). Children heard that both these kids want to watch all the movies previously talked about. Children were asked to decide which kid should be allowed to watch the movies, and to explain their choice (i.e., "Why?"). We identified consistent themes and coded responses to the explanation question along three dimensions. First, we coded whether children mentioned the movies as having potential *behavioral* consequences on the viewer, such that the viewer might commit the same harms depicted in the movies. Second, we coded whether children mentioned watching the movies as a reward. Third, we coded responses for a general positivity/negativity evaluation of the two kids.

Adult Participants. We presented adults with the same eight movie images (four depicting intentional harms, four depicting accidental harms; Table 1) in a randomized order. For each movie scene, we asked adults to rate how much each of three targets—a little kid, a big kid, and a grown-up—should be allowed to watch each movie (on a scale from 1: *Definitely should not be allowed* to 5: *Definitely should be allowed*).

Results

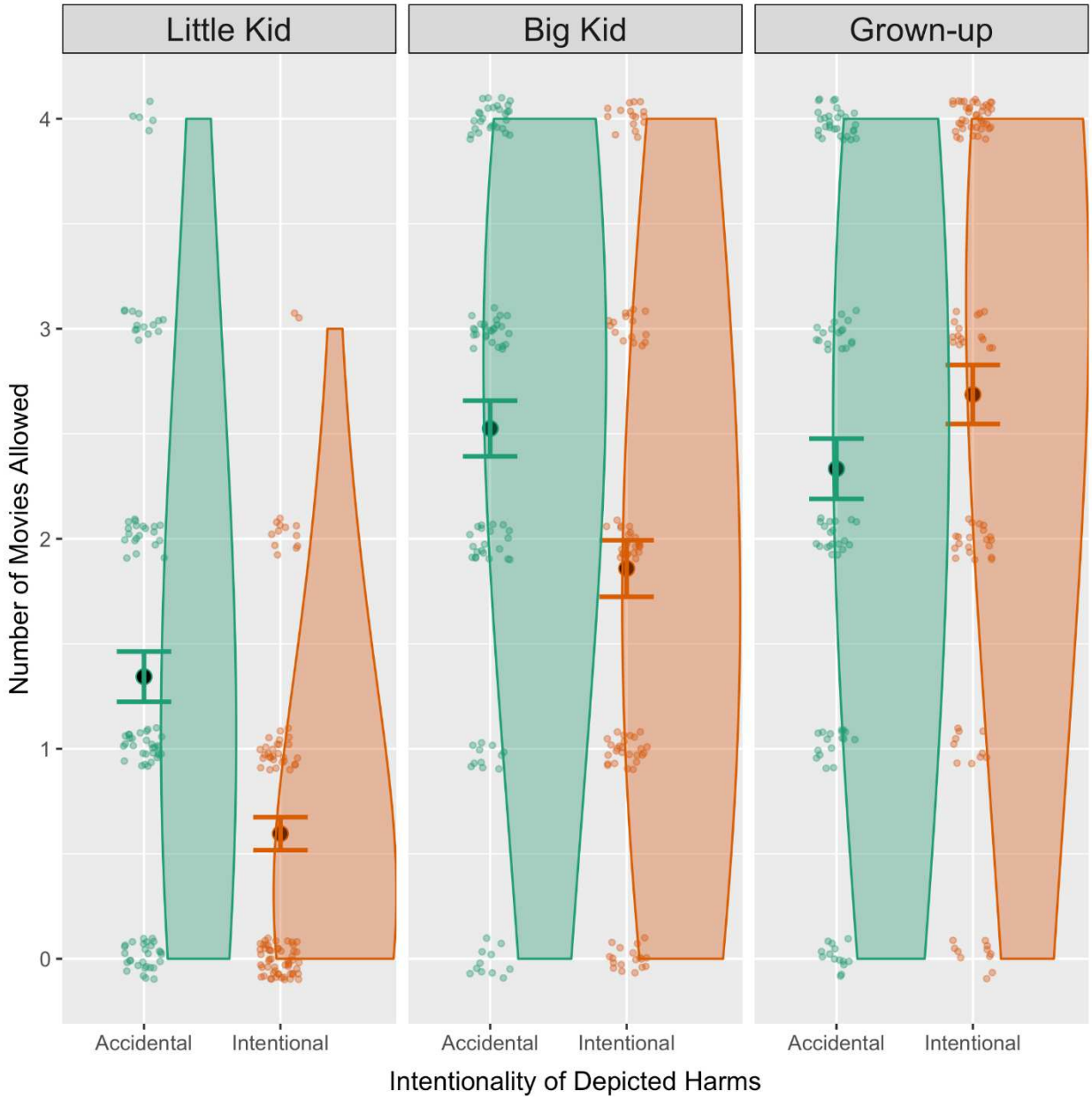
Child Participants

Movie Censorship. First, we examined the effect of movie intentionality (*intentional*, *accidental*) and target (little kid, big kid, grown-up) on the number of movies allowed. To do so, we conducted a binomial mixed-effects model predicting whether each movie was allowed. We entered fixed effects for target (little kid, big kid, grown-up), intentionality (0 = accidental, 1 = intentional), harm type (Physical Harm-Person, Physical Harm-Animal, Property Violation, Purity Violation), the intentionality X target interaction, the specific movie being decided on, participant age, and participant gender (0 = girl, 1 = boy).

Replicating our Study 1 results, we found a significant main effect of intentionality, such that participants allowed fewer movies featuring *intentional* harms than movies featuring *accidental* harms ($OR = 0.11$, $\beta = -2.17$, $SE = 0.27$, $p < .001$). As predicted, we also found a significant effect of target ($OR = 1.83$, $\beta = 0.60$, $SE = 0.08$, $p < .001$), such that participants allowed more movies for grown-ups than for big kids and more movies for big kids than for little kids (see Figure 1).

These effects were further qualified by a significant target X intentionality interaction ($OR = 2.27$, $\beta = 0.82$, $SE = 0.12$, $p < .001$): Participants allowed significantly fewer *intentional* harm movies than *accidental* harm movies for the little kids ($M_{intentional} = 0.58/4$ vs. $M_{accidental} = 1.38/4$) and the big kids ($M_{intentional} = 1.85/4$ vs. $M_{accidental} = 2.53/4$) but allowed more *intentional* harm movies than *accidental* harm movies for the grown-ups ($M_{intentional} = 2.74/4$ vs. $M_{accidental} = 2.39/4$). In addition, we again found that older participants allowed more movies than younger participants ($OR = 1.34$, $\beta = 0.30$, $SE = 0.06$, $p < .001$).

Figure 1



Distribution of the number of movies allowed for each target audience by the child participants in Study 2, based on whether the depicted harm was accidental or intentional (out of a total of 4 movies), along with means and SEs.

Explanation Coding. We coded children's responses based on the same codes used in Study 1. We excluded 35 children who did not provide codable explanations, leaving 68 codable responses. Overall, 26.5% of children providing a codable explanation focused only on the depicted harms, 29.5% focused only on the audience, and 44% focused on both the depicted

harm and the audience. As in Study 1, we examined whether the focus of children's responses changed with age. We found that they did ($\chi 2$ (4, N = 64) = 12.20, p = .02): With age, children became increasingly likely to focus on both the depicted harms and the audience (for children who gave a codable response: 26% of the 5- to 6-year-olds, 39% of the 7- to 8-year-olds, and 71% of the 9- to 10-year-olds).

Overall, 25% of codable responses mentioned intentions. This rate is considerably lower than in Study 1 (39%). One possibility is that this reduced rate resulted from participants being asked to make decisions for multiple targets, a potentially more salient feature of the task than the intentionality of the depicted harms. As in Study 1, references to intentions again increased with age ($\chi 2$ (2, N = 64) = 9.56, p = .008): Whereas the 5- to 6-year-olds never mentioned intentions (0% of codable responses), the 7- to 8-year-olds (39%) and 9- to 10-year-olds (29%) mentioned intentions at higher rates.

We next coded children's responses for whether they mentioned potential behavioral or emotional consequences. Even more than in Study 1, 34% of children providing a codable response mentioned a potential behavioral consequence, 10% mentioned a potential emotional consequence, 2% mentioned both influences, and 54% mentioned neither consequence. Children's references to potential consequences did not change with age ($\chi 2$ (6, N = 64) = 11.16, p = .08). As in Study 1, we examined whether the content of children's responses predicted their rates of censorship, finding no significant effects of different targets or mentioning intentions or these potential consequences (all ps > .35).

Finally, in just this study, we additionally coded children's responses for whether they mentioned the age of the audience. Overall, 73.5% of children providing a codable response mentioned audience age in their responses. We examined whether children who explicitly mentioned audience age responded differently when asked which movies were okay to watch

compared to children who did not mention audience age. To do so, we constructed a binomial mixed-effects model, predicting whether children allowed each movie by whether children mentioned audience age (0 = no, 1 = yes). As fixed effects, we additionally included participant age, question target (little kid, big kid, grown-up), and the interaction between question target and whether children mentioned audience age. We included a random effect for each participant and for order of movie scene introduction. There were no significant main or interactive effects with whether participants mentioned target age on censorship decisions (all *p*s > .48). There was also no difference in referring to the age of the audience based on the age of the participants ($\chi 2$ (2, N = 64) = 3.42, p = .18).

Censorship Motivations. When asked whether they were concerned about viewers feeling sad, acting bad, or both, 19% of participants said feeling sad, 16% said acting bad, 56% said both feeling sad and acting bad, and 9% said neither (although this was not provided as an option explicitly, some children spontaneously generated this answer). We followed-up on this question to ask children about each target (grown-up, big kid, little kid) individually. When asked what they were thinking about for the little kid, 24% of children said feeling sad, 23% said acting bad, 47% said both, and 6% said neither. When asked what they were thinking about for the big kid, 28% of participants said feeling sad, 29% said acting bad, 33% said both, and 10% said neither. When asked what they were thinking about for the grown-up, 35% of children said feeling sad, 10% said acting bad, 33% said both, and 22% said neither. Perhaps most notably, child participants were more likely to endorse a motivation to avoid behavioral consequences (i.e., acting bad) for child audiences compared to adult ones.

To further examine children's motivations regarding censorship and their thinking about who gets to watch movies depicting harms, we analyzed participants' choice for who gets to watch the movies—a kid who never gets in trouble or a kid who gets in trouble a lot—and their

explanation for their choice. Participants overwhelmingly selected the "good" kid over the "bad" kid (87% vs. 13%), $p < .001$. Looking at children's explanations of their choices on this question, 25% of responses mentioned a potential consequence of watching the movies on the viewer (75% did not); 13% of responses mentioned watching the movies as a reward (87% did not); and 43% of responses included a basic mention of the moral valence of the two targets (57% did not).

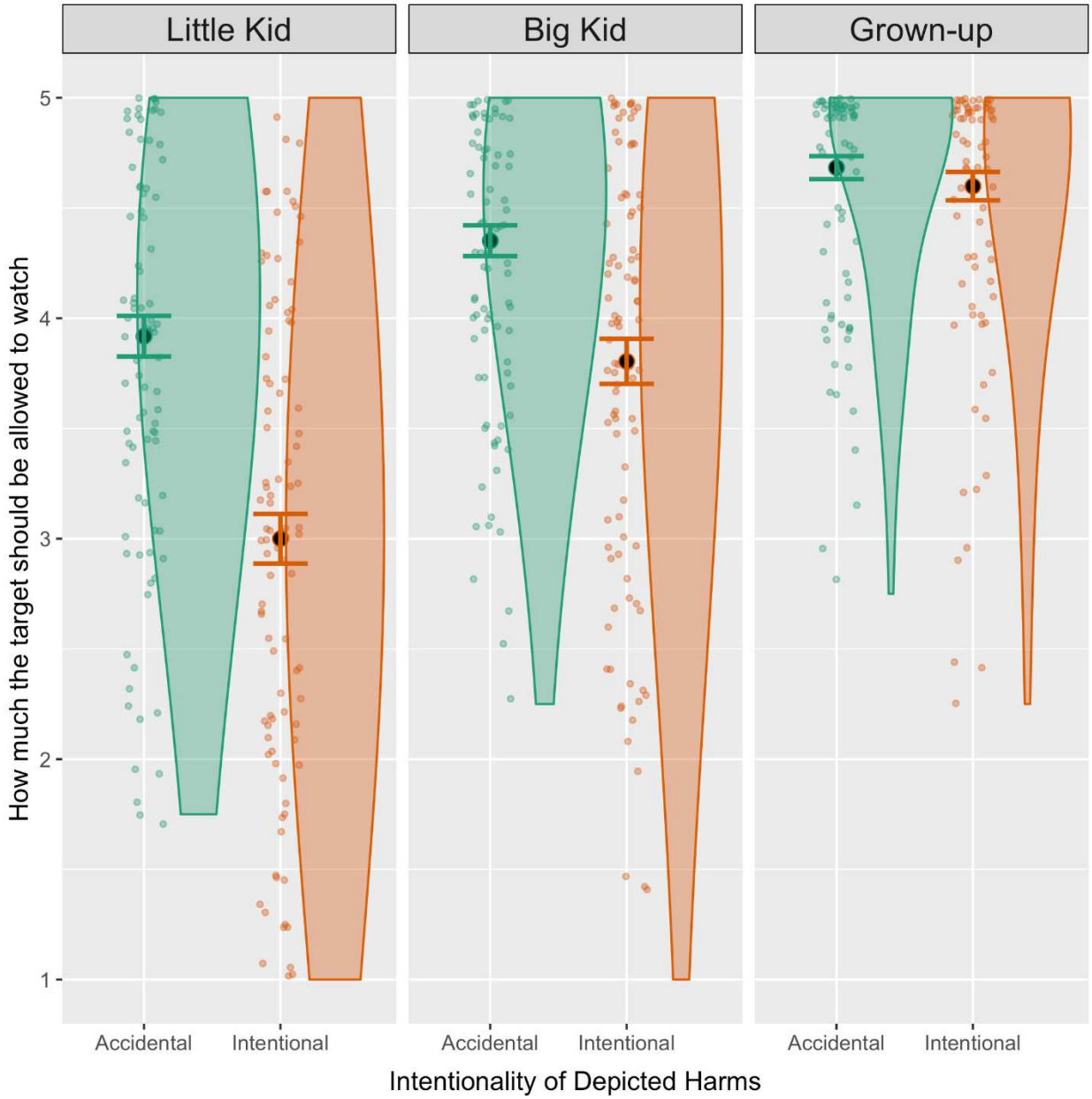
Adult Participants

Adults' ratings largely mirrored the pattern of results observed with children (see Figure 2). Specifically, we observed a significant main effect of target ($F(2, 99) = 132.16$, $p < .001$, $\eta_p^2 = .73$) a significant main effect of intentionality ($F(1, 100) = 114.11$, $p < .001$, $\eta_p^2 = .53$) and a significant interaction between target age and intentionality ($F(2, 99) = 64.42$, $p < .001$, $\eta_p^2 = .58$). As predicted—and consistent with the results from children—there was a larger difference in ratings of allowability of accidental versus intentional movies when adults were thinking about the kids compared to the grown-up.

Discussion

In Study 2, we both replicated and extended our results from Study 1. As before, children were quite willing to censor movies depicting harmful content. We further observed that children's rates of censoring depended on the audience involved: Children across ages 5 to 10 allowed fewer movies for younger targets than for older targets, suggesting that already by age 5, children recognize that depictions of harm may have different impacts on people of different ages. This finding highlights the importance of the target in censorship: Potentially vulnerable audiences (e.g., young children) receive greater regulation in what they can watch compared to less vulnerable audiences (e.g., grown-ups).

Figure 2



Distribution of responses for the adult participants of Study 2, with mean ratings (and SEs) for how allowed each target should be to watch the different movies (on a scale from 1: *Definitely should not be allowed* to 5: *Definitely should be allowed*), based on whether the harm depicted in the movie was *accidental* or *intentional*.

In addition to an effect of audience age, we again also observed an effect of participant age: As in Study 1, older children censored fewer depictions of harm compared to younger children. Thus, it is unlikely that the age effect observed in Study 1 (and here in Study 2) resulted only from younger and older children bringing to mind audiences of different ages. One

possibility is that despite being told to imagine a "little kid" and a "big kid" in Study 2, children of different ages were still imagining audiences of slightly different ages. An alternative possibility is that older children may have different—or more—experience with themselves being censored. Relatedly, older children in our age range may be particularly at an age where they are seeking independence and perhaps resisting parents' and others' regulations around their viewing behaviors (thus resulting in older children being potentially more reluctant to censor). A final (and complementary) possibility is that older (vs. younger) children viewed the depicted harms as less severe (e.g., due to greater experience with or sensitization to harm). Our current data cannot differentiate between these possibilities, but better understanding the developmental trajectories of censorship decisions represents an exciting line of potential future work. In asking these questions, future research may also consider whether similar age effects persist when children actually find themselves in situations where they have opportunities to watch movies (where younger children may find it more challenging to follow the abstract principles they endorse; see also Smith et al., 2013).

Our findings in Study 2 further speak to the motivations underlying children's decisions to censor. Particularly, we observed that children were more likely to censor depictions of intentional harms from younger audiences. This finding extends past research on intentionality and the development of moral judgment (e.g., Ames & Fiske, 2014; Chernyak & Sobel, 2016; Cushman, 2008; Cushman et al., 2013; Darley et al., 1978; Killen et al., 2011; Shultz et al., 1986) by suggesting that children incorporate situational and social considerations (e.g., who is the target?) into moral judgments involving intentionality. That is, the importance of intentionality in moral judgments depends on who is being considered. At the same time, this finding suggests that children's tendency to censor intentional harms more than accidental ones reflects more than simply a basic attention to the intentionality of the depicted acts (which may

be more similar across audiences of different ages)—instead, children's decisions to censor depictions of intentional harms appear to reflect a particular concern with avoiding potential behavioral consequences. This point is underscored by our finding that children were also more likely to endorse a motivation to avoid audiences from "acting bad" when thinking about younger versus older audiences. Our central goal in Studies 3 and 4 was to build on these findings and to more directly examine a potential link between a motivation to avoid cascades of future harmful behaviors and the decision to censor depictions of intentional harms.

Study 3

In Studies 1 and 2, we observed results consistent with the idea that children consider both the potential emotional and behavioral consequences of viewing depictions of harm. Specifically, we found evidence that children censor not only to prevent viewers from "feeling sad," but also to avoid inspiring potential cascades of future harmful behaviors. In Study 3, we sought to further explore dual motivations of avoiding potential emotional and behavioral consequences of viewing harm and to directly examine how each of these motivations informs children's rates of censoring accidental vs. intentional harms.

Using a similar design to the *Other* condition in Study 1, we told all participants that they would be deciding what movies another same-age, same-gender kid could watch. Then, between-subjects, we randomly assigned children to hear either a motivation to avoid making the other kid feel sad (*emotional* consequence) or to avoid making the other kid do something bad (*behavioral* consequence). We hypothesized that participants would censor more depictions of the intention to harm when motivated to avoid the other kid doing something bad relative to feeling sad, as children might infer that intentional harms are more likely to be copied than accidental ones. That is, intentional actions (harmful or not) are thought to have reasons motivating them (e.g., Cushman, 2008; Cushman et al., 2013). For depictions of intentional

harms, it is this underlying intention to harm that may particularly be what children want to censor. Although we favored this hypothesis (and observed indirect evidence for this idea in Studies 1 and 2), it was also possible that children might attend to intentionality equally for both motivations. Children might infer that intentional (vs. accidental) harms are more likely to make another kid feel sad because of the more negative, harmful intentions involved.

Method

Participants

Participants included 116 5- to 10-year-old children (39% girls, 61% boys; M_{age} = 93.34 months; *SD* = 20.12; 20 5-year-olds, 22 6-year-olds, 21 7-year-olds, 23 8-year-olds, 12 9-year-olds, 18 10-year-olds) tested in Central New York. Children participated in a lab or a children's science museum. Between-subjects, we randomly assigned children to the *Avoid Feeling Sad* condition (*n* = 60) or the *Avoid Doing Bad* condition (*n* = 56). Of parents reporting children's race/ethnicity, 67% identified their children as White; 12% as Biracial; 8% as Other; 6% as Hispanic or Latin(o/a/e); 3% as Asian or Asian American; 2% as Black or African American; and 2% as Native American. Of parents reporting family income, income ranged from < $15,000 to > $150,000, with a median annual family income of $75,000 – $100,000.

Materials and Design

The materials and design were identical to those used in Study 1.

Procedure

Movie Censorship. The experimenter first told children about how "sometimes it's okay for someone to watch a movie, but sometimes someone will say that a movie isn't okay for people to watch." Next, the experimenter asked participants to imagine that another kid who was their same age and gender had a list of movies that they wanted to watch. Then, the experimenter told participants, based on their assigned condition: "However, I'm a bit worried that some of

these movies might make the kid [feel sad/do something bad].” After explaining to children that their job was to decide for each movie whether the other kid could watch it, the experimenter reiterated the motivational condition by saying that: “I want to make sure the other kid doesn’t [feel sad/do something bad] after watching the movies.” The procedure for the movie decision task was otherwise identical to that in Study 1. We again additionally asked children “Why did you choose the way you did?” and coded children’s responses using the coding scheme from Study 1.

Results

Movie Censorship

Overall, across condition, harm type, and intentionality, children allowed fewer than half of the movies to be watched (M = 1.97 movies; $t(115) = 12.82$, $p < .001$). It is worth noting that this mean rate of overall censorship is significantly below that observed in Study 1 (M = 2.73 movies allowed; $t(210) = 2.72$, $p = .007$). This difference in the number of movies allowed between Studies 1 and 3 makes sense given that children in Study 3 received a direct warning about potential negative consequences that children in Study 1 did not receive.

We conducted a binomial mixed-effects model predicting whether participants allowed each movie. As fixed effects, we included condition (0 = *Avoid Feeling Sad*, 1 = *Avoid Doing Bad*), harm type, intentionality, the intentionality X condition interaction, the specific movie being decided on, participant age, and participant gender. We included random effects for participant and for the viewing order of the movies.

Contrary to our prediction, there was no significant interaction between condition and intentionality, suggesting that there was no difference in the effect of different motivations on the tendency to censor depictions of *intentional* versus *accidental* harms (OR = 0.50, β = -0.69, SE = 0.47, $p = .14$). However, replicating our results from Studies 1 and 2, there was a significant

main effect of intentionality on how many movies children allowed overall, with participants allowing more movies depicting *accidental* harms (M_{sad} = 1.70/4 movies, M_{bad} = 1.55/4 movies) than movies depicting *intentional* harms (M_{sad} = 0.45/4 movies, M_{bad} = 0.21/4 movies; *OR* = 0.09, β = -2.36, *SE* = 0.30, $p < .001$). There was no main effect of condition on number of movies allowed (*OR* = 0.75, β = -0.29, *SE* = 0.35, $p = .41$). As in Study 1, there was a significant association between participant age and movies allowed such that older participants allowed more movies on average than younger participants (*OR* = 1.38, β = 0.32, *SE* = 0.10, $p = .002$).

Explanation Coding

We coded children's responses in the same way as in Study 1. We excluded 22 children who did not provide codable responses, leaving 94 codable responses. Children's codable responses most often focused on the depicted harms (67%), followed by both the depicted harms and the audience (27%), and then only the audience (5%). As before, we examined whether the focus of children's responses related to their rates of censorship; we found no significant main or interactive effects of the focus of children's responses on their rates of censorship, all *p*s > .99. We examined whether the focus of children's responses varied by condition or changed with age. There were no significant effects of condition ($\chi2$ (3, N = 94) = 4.32, $p = .23$) or participant age ($\chi2$ (2, N = 94) = 8.04, $p = .09$).

Overall, 43% of children who provided codable responses mentioned harm intentionality (comparable to the rate found in Study 1). As in Study 1, whether children mentioned intentions significantly interacted with the intentionality of the depicted harm in predicting their censorship decisions (*OR* = .20, β = -1.60, *SE* = 0.52, $p = .002$). Specifically, participants who mentioned intentionality, compared to those who did not mention intentionality, allowed more movies with *accidental* harms (*M* = 2.20/4 vs. *M* = 1.36/4) and fewer movies with *intentional* harms (*M* = 0.24/4 vs. *M* = 0.42/4). We again examined whether children's likelihood of mentioning harm

intentionality varied by condition or changed with age. Children in the *Avoid Doing Bad* condition (59%) mentioned intentions at a higher rate than participants in the *Avoid Feeling Sad* condition (40%), $\chi 2$ (1, N = 94) = 3.31, p = .069, although this finding was not significant. Similar to in Study 1, we found a significant effect of age (χ^2 (2, N = 94) = 13.97, p < .001) such that the 5- to 6-year-olds were considerably less likely to mention intentions (14% of children) compared to both the 7- to 8-year-olds (45%) and the 9- to 10-year-olds (43%).

Finally, 17% of children who provided codable responses mentioned potential behavioral consequences of viewing the depictions of harm, 7% mentioned potential emotional consequences, 1% mentioned both consequences, and 75% mentioned neither consequence. Using the same analytic strategy as above, we examined whether children who mentioned consequences censored differently from those who did not. There was a significant interaction between whether children mentioned emotional consequences and the intentionality of the depicted harms (OR = 1.02, β = 2.32, SE = 0.70, p < .001). Children who mentioned emotional consequences censored depictions of accidental and intentional movies at more similar rates (M = 2.00/4 accidental movies vs. M = 1.57/4 intentional movies) compared to children who mentioned behavioral consequences (M = 1.59/4 accidental movies vs. M = 0.05/4 intentional movies) and children who made no reference to potential consequences (M = 1.77/4 accidental movies vs. M = 0.31/4 intentional movies). We again also examined whether children's likelihood of mentioning behavioral and emotional consequences varied by condition or changed with age. There was no significant effect of condition (p = .72); indeed, it is worth noting that children spontaneously mentioned behavioral consequences (at rates higher than emotional consequences) in both the *Avoid Doing Bad* condition (31%) *and* the *Avoid Feeling Sad* condition (16%), the latter of which contained no prompt regarding consequences on behavior, but an explicit prompt regarding consequences on emotions. There was a significant effect of age

(χ^2 (6, N = 94) = 15.10, p = .02), such that references to behavioral consequences increased with age (7% of the 5- to 6-year-olds providing a codable response, 29% of the 7- to 8-year-olds, and 32% of the 9- to 10-year-olds).

Discussion

In Study 3, we aimed to more directly examine the presence and influence of motivations to prevent viewers from feeling sad (i.e., emotional consequences) and to avoid viewers from acting bad (i.e., behavioral consequences). We predicted that priming different motivations would have differential effects on children's rates of censoring intentional harms. However, contrary to our hypothesis, we observed no difference in participants' censorship of intentional harms between conditions. At the same time, children's responses provide some supportive evidence: Children who were given instructions to prevent negative consequences on behavior were more likely to spontaneously mention intentions in their open-ended explanations.

One possibility is that the motivational instructions we provided were too strong, such that children were overly cautious in both conditions. In support of this conclusion, the majority of children allowed zero movies depicting intentional harms (78% of participants). This distribution thus presents a limited range issue: Given the very low rates of allowing intentional harm movies, it would be very difficult to detect any differences between conditions if such differences really did exist. In Study 4, we thus altered our method such that children were prompted to select a set of four movies that are most acceptable to watch.

Study 4

In Study 4, we continued our examination of how dual motivations to avoid potential emotional and behavioral consequences of viewing depictions of harm may differentially influence children's rates of censoring depictions of intentional harms. Given children's low rates of allowance in Study 3, we altered our method so that children were asked to choose

several movies for each target to watch. Framed another way, we asked children to select the movies that they deemed *most* acceptable for the target to watch. This allowed us to get a sense of children's relative rankings of the different movies. By holding constant the overall number of movies allowed, our goal was to overcome the limited range issue in Study 3.

As in Study 3, we provided participants in Study 4 with either a motivation to avoid making another child feel sad (i.e., *emotional* consequences) or to avoid making another child do something bad after watching the movies (i.e., *behavioral* consequences). Then, we instructed children to pick four of the eight total movies that they deemed most acceptable for the other child to watch. We predicted that participants with a motivation to avoid inspiring future harmful behaviors (compared to the motivation to avoid making another child feel sad) would allow fewer depictions of intentional harms.

Method

Participants

Participants included 116 5- to 10-year-old children (43% girls, 57% boys; M_{age} = 87.71 months; *SD* = 18.59; 18 5-year-olds, 26 6-year-olds, 22 7-year-olds, 18 8-year-olds, 18 9-year-olds, 10 10-year-olds, 5 missing exact age data) tested in Central New York. Children participated in a lab, a children's science museum, or an elementary school. Between-subjects, we randomly assigned children to the *Avoid Feeling Sad* condition (*n* = 57) or the *Avoid Doing Bad* condition (*n* = 59). Of parents reporting children's race/ethnicity, 83% identified their children as White; 7% as Hispanic; 3% as Asian or Asian American; 3% as Black; and 3% as Biracial. Of parents reporting family income, income ranged from < $15,000 to > $150,000, with a median annual family income of $75,000–$100,000.

Materials

We used the same eight movie images from the previous studies but presented them to participants on individually laminated cards. We used two sets of cards, with one movie scene within each harm type being done "by accident" and the other movie scene done "on purpose" within each set. We counterbalanced the presentation of the two sets between participants.

Procedure

We used the same procedure from Study 3, with a few changes. Instead of asking children to make a decision about each individual movie, we instead asked children to identify four of the eight total movies as ones that the other kid should be allowed to watch. The experimenter first shuffled the eight cards depicting movie scenes, and then read through all the cards, arranging them on a table in front of children. After introducing all the movie scenes, the experimenter asked children to pick four movies that the other kid should be allowed to watch. Afterwards, the experimenter asked children why they picked the way they did. We coded responses to this explanation question in the same way as in Studies 1 and 3.

Results and Discussion

Across analyses, we consider the four movies children chose as most allowable. To examine the effect of motivational framings on children's movie choices, we compared the relative rates of picking movies depicting *intentional* harms across conditions. Because the data were non-normally distributed, we compared rates of selecting movies with *intentional* harms using a Mann-Whitney *U* test. Consistent with our predictions, we found that participants picked significantly fewer movies depicting *intentional* harms when prompted to avoid making the other kid do something bad ($M = 0.51/4$) than when prompted to avoid making the other kid feel sad ($M = 0.97/4$), $Z = 2.86$, $p = .004$. When children were in the frame of mind of wanting to avoid negative consequences on behavior—relative to negative consequences on emotions—they were more likely to avoid showing content that features *intentional* harms.

We coded children's open-ended explanations in the same way as in Studies 1 and 3. We excluded 30 children who did not provide codable responses, leaving 87 codable responses. Similar to Studies 1 and 3, the majority of codable responses (86%) focused on the depicted harms. There was no effect of age on the focus of participants' responses (χ^2 (4, N = 86) = 6.77, p = .15). Overall, 68% of children providing codable responses mentioned intentionality, suggesting that intentionality was critical to how children approached their decisions. In addition, references to intentions increased with age (χ^2 (2, N = 86) = 16.51, p < .001): of children providing codable responses, 40% of the 5- to 6-year-olds mentioned intentions versus 77% of the 7- to 8-year-olds and 88% of the 9- to 10-year-olds. With respect to references to potential consequences of viewing the depicted harms, 9% of children providing codable responses mentioned potential behavioral consequences, 5% mentioned potential emotional consequences, and the majority of participants mentioned neither consequence. There was no effect of participant age on children's likelihood of mentioning potential consequences (χ^2 (4, N = 86) = 7.22, p = .12). Last, there were no significant differences by condition on any of the three codes (all ps > .31).

General Discussion

Across four studies, we found robust, consistent evidence that young children censored depictions of harm when considering both themselves and a variety of other audiences. Our findings suggest that children thus represent the potential ripple effects that harms can hold, anticipating potential emotional and behavioral consequences of exposure to even depictions of harm. These findings contribute to an emerging literature suggesting that children represent harms within a wider "landscape" of harm—that is, that harms exist and hold consequences outside of the immediate context of a perpetrator harming a victim (see also Marshall et al., 2020; Heck et al., 2021a).

Notably, the youngest in our sample were the most likely to censor, and even these younger children censored depictions of harm in nuanced ways. Our findings thus suggest that the motivation to censor depictions of harm may be rooted in early socio-cognitive processes and understandings of harm. At the same time, it is likely that socialization and experience continue shaping beliefs about censorship, during childhood and into adulthood. Indeed, one possibility is that older children were less likely to censor due to greater experience with being censored themselves, or with viewing depictions of harm (which may then result in greater sensitization to depictions of harm). Future research may further consider the potential role of socialization in thinking about censorship. For example, open questions concern how parents' attitudes and values (e.g., political orientation, religion) may inform children's thinking about censorship and when it is appropriate.

Our results suggest that beginning early in life, motivations to censor reflect not only a concern with potential emotional consequences of viewing depictions of harm (e.g., making the audience feel sad) but also potential behavioral consequences of viewing depictions of harm (e.g., inspiring the audience to themselves engage in harmful behaviors). We saw initial evidence for this idea in Studies 1 and 2, such that children often spontaneously mentioned potential behavioral consequences in their explanations (and did so increasingly with age), and that children censored depictions of intentional harms more than accidental ones, especially when considering younger audiences. We posited that if children are motivated only to protect viewers' emotions, then they may censor all depictions of harmful outcomes at similar rates. However, if children are particularly motivated to avoid the possibility of inspiring future harmful behaviors, then they may be especially concerned with censoring depictions of the intent to harm (which may be more likely to be copied). After addressing the limited range issue of Study 3, we found more direct evidence for this possibility in Study 4, finding that children are

more likely to censor depictions of intentional harms when motivated to avoid potential behavioral versus emotional consequences.

Our finding that children censored depictions of harm both to avoid potential emotional and behavioral consequences aligns with recent work that children punish perpetrators of harm based on both retributive (i.e., making the perpetrator feel bad or ensuring that they receive their "just deserts") and consequentialist motives (i.e., efforts to deter the perpetrator from engaging in future harmful behaviors; Marshall et al., 2020). Together, these findings suggest that a motivation to disrupt cycles of harmful behavior underlies moral cognition beginning early in life.

Our research also connects to a broader discussion regarding the role of intentionality in both children's and adult's moral judgments, whereby intentional moral acts are judged more intensely than accidental moral acts (e.g., Ames & Fiske, 2014; Chernyak & Sobel, 2016; Cushman, 2008; Cushman et al., 2013; Killen et al., 2011). Like other sorts of moral judgments, we consistently found that children censored depictions of harm based on whether the depicted harm was done intentionally or accidentally. Importantly, extending past research, our results demonstrate that the role of intentions in censorship (and perhaps moral judgment broadly) depends on the potential audience and on other situational and social factors. That is, the intentionality of depicted harms influence decisions to a greater degree for more vulnerable audiences (e.g., young children) and based on the presence of certain motivations (e.g., wanting to prevent others from doing something harmful). Although we focused here on the role of intentionality of depicted harms in decisions to censor, future work may also examine other features of moral judgments, like the role of severity (i.e., the overall "wrongness" or amount of harm of the depicted act) or the particular people involved in the act (e.g., the same act might be evaluated differently depending on if the hero or the villain of the story is doing the act).

Our findings also hold a range of implications for children's broader moral thinking and behaviors. First, given that part of children's motivation to censor was to prevent subsequent harmful behavior committed by the audience, our findings hold implications for how both adults and children predict moral character (e.g., Critcher et al., 2020; Liberman, Howard, Vasquez, & Woodward, 2018; Uhlmann, Pizarro, & Diermeier, 2015), as well as social behavior, more broadly (e.g, Carpendale & Lewis, 2004; Tamir & Thornton, 2018; Tomasello, Carpenter, Call, Behne, & Moll, 2005). Children appear to consider depictions of harm as potentially having negative consequences on an audience's future behavior, suggesting that censorship is a means of regulating and managing the behavior of others to prevent harms from occurring. In making decisions to restrict access to certain content, censors assume the potential vulnerability of the audience's moral character, predicting that such character will *become* more immoral through exposure to depictions of harm.

Second, our findings on censorship may also provide insight into the development and maintenance of moral taboos—behaviors that are considered morally restricted such that no trade-off is worth engaging in the behavior (Tetlock, 2003; Tetlock, et al., 2000). Cultures differ in what actions are considered morally taboo or permissible (Awad et al., 2020; Miller et al., 1990; Shweder et al., 1987). Taboos can exist across a wide variety of concerns (Graham et al., 2009), including which foods are considered eligible for consumption (Meyer-Rochow, 2009), practices relating to sexuality (Lieberman & Smith, 2012), and—aligned with censorship—what words and language people may use (Allan & Burridge, 2006). Censorship may work to facilitate moral taboos, such that actions considered taboo will be more likely to be censored, as well as that censoring particular content may make that content seem more taboo. In this way, censorship may work as a moral signal, designating what behaviors, ideas, and images are taboo by the very act of censoring them. In addition, our results highlight a potential mechanism for

this relation, whereby taboo acts are censored out of a concern that viewing such taboo acts may inspire the audience to perform them.

Third, our findings hold implications for how children navigate their own and others' consumption of media. Children today are exposed to vast amounts of media and are more able to share media with others than perhaps ever before (e.g., through social media; Rideout et al., 2022). This present-day context thus makes understanding children's thinking about the effects of media—on both themselves and others—particularly relevant. Prior research suggests that exposure to depictions of harm can indeed result in both emotional and behavioral consequences for children (e.g., Anderson et al., 2003; Gentile et al., 2007). Our finding that children understand these downstream consequences from an early age suggests that media literacy interventions may focus on both these potential consequences, and particularly on the potential behavioral consequences of depictions of harm (see also Pfaff-Rüdiger & Riesmeyer, 2015; Potter, 2010). Future research may also build on our findings to think about children's specific understanding of *sharing* depictions of harm (e.g., via social media) and to study children's thinking about the effects of other forms of media.

Limitations and Broader Future Directions

This paper presents an initial examination of the development of beliefs about censorship, and many open questions remain for future research. First, although we operationalized depictions of harm specifically as movie scenes depicting harmful acts, additional research may investigate how the nature of the media may impact censorship. For example, there may be important differences in whether the media has a visual component or not (e.g., movies vs. books), or whether the media allows for active participation from the audience (e.g., video games vs. movies). For instance, audience participation may influence both relative rates of censorship and the importance of certain motivations (e.g., the motivation to avoid modeling harmful

behavior may be stronger for video games than for movies because of the viewer's active participation in video games). In addition, people censor not only depictions of harmful acts but also profanity, nudity, explicit sexual content, and even simply opinions, facts, and people they disagree with. Although we did not find any differences in censorship based on harm type (for a similar finding, see Anderson & Masicampo, 2017), this does not mean that other types of content perceived as objectionable would be treated similarly.

Second, additional work may also investigate not only children's thinking regarding censorship but also related topics, such as children's broader understanding of free speech (see Helwig, 1998; Helwig & Prencipe, 1999). In the present research, we have focused on considerations regarding the effect of media on particular audiences, but another way to think about censorship is about dictating what media, information, and ideas are generated in the first place. That is, in this case, the target of censorship is not the audience but the creator. The relation between censorship and content generation has received extensive investigation in the philosophical and legal literature, often centering on the rights and liberties of creators (e.g., Coetzee, 1996; Dworkin, 1985; Dyzenhaus, 1992; Hall, 1906; West, 2018). In some cases, people may object to a particular audience (e.g., children) consuming certain media (e.g., a violent movie) but not to the creation of that media itself (i.e., people may deem it okay for some audiences, like adults, to consume that media). In other cases, people may object to the creation of media, regardless of who the audience is. In this way, censoring for creators and censoring for audiences may be independent processes, each worthy of their own investigation.

Third, the present findings are limited in that our samples were drawn from a single location, and a relatively homogenous population. However, different contexts—both at the broader societal level, in terms of regulatory organizations, and at the local household level, in terms of individual restrictions and instructions—are likely to shape how children (and adults)

approach censorship. It will be important in future work to examine how attitudes about censorship develop across a wider array of cultural and sociological contexts, as well as to consider the potential emergence and development of individual differences in children's beliefs about censorship. Doing so will both improve the generalizability of our findings and also allow for consideration of how attitudes about censorship may develop differently within and across cultures. For example, countries vary in the legal restrictions placed on media and speech (Warf, 2009, 2011; Warf & Vincent, 2007), and such differences may influence how children learn the cultural norms regarding what sorts of things are allowed in media and speech. As a related concern, additional research could understand how children learn about and potentially internalize media rules and expectations as articulated and demonstrated by parents/guardians, teachers, authority figures, and society more broadly. Open questions include how children develop an understanding of what their parents and others expect of them in terms of what media is acceptable to consume, and how this understanding changes with age.

Fourth, in our studies, children acted as censors for specific individuals, who were mostly anonymous strangers. However, censorship in the real world can take many forms and can be done by and for many different people, often by individuals who directly know the potential audiences. For example, parents who censor for their own children may be more directly motivated to ensure that their children are not inspired by the violent media they consume. Additional research may aim to understand whether particular individuals have greater perceived authority for determining what an audience can watch. For example, one likely possibility is that parents are seen as having a great deal of authority over what their children are allowed to consume. In addition, our studies largely sidestep the issue of censorship as enacted by organizations, in addition to other formalized versions of censorship. Both governments and industry regulation bodies exert an influence on how information and media is created and

distributed (e.g., King et al., 2013; Warf, 2009). Future research may examine whether children think that political leaders and governments have the authority to censor, or whether that belief emerges later in life. Children's thinking about whether governments and organizations have such authority may also relate to individual differences in children's emerging sociopolitical values (e.g., authoritarian values or political orientation; see also Reifen-Tagar et al., 2014; Reifen-Tagar & Cimpian, 2022; Heck et al., 2021b). Related open questions concern how beliefs about censorship may vary based on who the censor is and what their relationship is to the audience.

Last, in our studies, we used a broad open response question to examine children's explicit justifications of their censorship choices. Although this measure provides initial insight into children's reasoning and motivations, a sizeable proportion of children in our studies did not provide a "codable" response, perhaps in part due to the broad and open-ended nature of the question. The results from our explanation coding are thus limited in that they reflect only a subset of the total number of children and that our coding themes depend on the points that children spontaneously raised. Future research might consider asking children questions that target more specific aspects of their reasoning about censorship, for example by explicitly asking children to reason about the potential consequences that viewing a harm could hold. In addition, future work might ask children to explain their decision-making after each censorship decision, rather than only at the end of the study.

Conclusion

In sum, the present studies reveal that children as young as age 5 demonstrate a propensity to censor depictions of harmful content. These decisions seem driven not simply by a blanket desire to protect audiences' emotions, but especially by a motivation to disrupt potential cascades of future harmful behavior that may be inspired by others viewing depictions of harm.

Together, our work advances an exciting new paradigm for examining both the development of moral cognition and the development of beliefs regarding censorship.

References

Allan, K., & Burridge, K. (2006). *Forbidden words: Taboo and the censoring of language*. Cambridge University Press.

Ames, D., & Fiske, S. T. (2014). A potential path to integration of blame judgments. *Psychological Inquiry, 25*(2), 193-196. https://doi.org/10.1080/1047840X.2014.902724

Anderson, C. A., Berkowitz, L., Donnerstein, E., Huesmann, L. R., Johnson, J. D., Linz, D., Malamuth, N. M., & Wartella, E. (2003). The influence of media violence on youth. *Psychological Science in the Public Interest, 4*(3), 81–110. https://doi.org/10.1111/j.1529-1006.2003.pspi_1433.x

Anderson, R. A., & Masicampo, E. J. (2017). Protecting the innocence of youth: Moral sanctity values underlie censorship from young children. *Personality and Social Psychology Bulletin*, *43*(11), 1503-1518. https://doi.org/10.1177/0146167217722557

Awad, E., Dsouza, S., Shariff, A., Rahwan, I., & Bonnefon, J. F. (2020). Universals and variations in moral decisions made in 42 countries by 70,000 participants. *Proceedings of the National Academy of Sciences*, *117*(5), 2332-2337. https://doi.org/10.1073/pnas.1911517117

Baer, C., & Odic, D. (2022). Mini managers: Children strategically divide cognitive labor among collaborators, but with a self-serving bias. *Child Development*, *93*(2), 437-450. https://doi.org/10.1111/cdev.13692

Bender, P. K., Plante, C., & Gentile, D. A. (2018). The effects of violent media content on aggression. *Current Opinion in Psychology, 19,* 104-108. https://doi.org/10.1016/j.copsyc.2017.04.003

Blake, P. R., McAuliffe, K., & Warneken, F. (2014). The developmental origins of fairness: The knowledge–behavior gap. *Trends in Cognitive Sciences, 18*(11), 559–561. https://doi.org/10.1016/j.tics.2014.08.003

Bregant, J., Wellbery, I., & Shaw, A. (2019). Crime but not punishment? Children are more lenient toward rule-breaking when the "spirit of the law" is unbroken. *Journal of Experimental Child Psychology, 178,* 266–282. https://doi.org/10.1016/j.jecp.2018.09.019

Carpendale, J. I., & Lewis, C. (2004). Constructing an understanding of mind: The development of children's social understanding within social interaction. *Behavioral and Brain Sciences*, *27*(1), 79-96. https://doi.org/10.1017/S0140525X04000032

Chernyak, N., & Kushnir, T. (2014). The self as a moral agent: Preschoolers behave morally but believe in the freedom to do otherwise. *Journal of Cognition and Development, 15*(3), 453–464. https://doi.org/10.1080/15248372.2013.777843

Chernyak, N., & Sobel, D. M. (2016). "But he didn't mean to do it": Preschoolers correct punishments imposed on accidental transgressors. *Cognitive Development*, *39*, 13-20. https://doi.org/10.1016/j.cogdev.2016.03.002

Chobhthaigh, S. N., & Wilson, C. (2015). Children's understanding of embarrassment: Integrating mental time travel and mental state information. *British Journal of Developmental Psychology*, *33*(3), 324–339. https://doi.org/10.1111/bjdp.12094

Chung, S., & Moon, S. (2016). Is the third-person effect real? A critical examination of rationales, testing methods, and previous findings of the third-person effect on censorship attitudes. *Human Communication Research*, *42*(2), 312-337. https://doi.org/10.1111/hcre.12078

Coetzee, J. M. (1996). *Giving offense: Essays on censorship*. The University of Chicago Press.

Critcher, C. R., Helzer, E. G., & Tannenbaum, D. (2020). Moral character evaluation: Testing another's moral-cognitive machinery. *Journal of Experimental Social Psychology, 87,* Article 103906. https://doi.org/10.1016/j.jesp.2019.103906

Cushman, F. (2008). Crime and punishment: Distinguishing the roles of causal and intentional analyses in moral judgment. *Cognition*, *108*(2), 353–380. https://doi.org/10.1016/j.cognition.2008.03.006

Cushman, F., Gray, K., Gaffey, A., & Mendes, W. B. (2012). Simulating murder: The aversion to harmful action. *Emotion, 12*(1), 2–7. https://doi.org/10.1037/a0025071

Cushman, F., Sheketoff, R., Wharton, S., & Carey, S. (2013). The development of intent-based moral judgment. *Cognition*, *127*(1), 6-21. https://doi.org/10.1016/j.cognition.2012.11.008

Dahl, A., & Freda, G. F. (2017). How young children come to view harming others as wrong: A developmental analysis. In J. A. Sommerville & J. Decety (Eds.), *Social cognition: Development across the life span* (pp. 151–184). Routledge/Taylor & Francis Group.

Darley, J. M., Klosson, E. C., & Zanna, M. P. (1978). Intentions and their contexts in the moral judgments of children and adults. *Child Development*, *49*(1), 66-74. https://doi.org/10.2307/1128594

Davison, W. P. (1983). The third-person effect in communication. *Public Opinion Quarterly*, *47*(1), 1-15. https://doi.org/10.1086/268763

Dworkin, R. (1985). *A matter of principle*. Harvard University Press.

Dunlea, J. P., & Heiphetz, L. (2021). Moral psychology as a necessary bridge between social cognition and law. *Social Cognition*, *39*(1), 183-199. https://doi.org/10.1521/soco.2021.39.1.183

Dyzenhaus, D. (1992). John Stuart Mill and the harm of pornography. *Ethics*, *102*(3), 534-551.

Feinberg, J. (1985). *Offense to others*. Oxford University Press.

Gentile, D. A., Saleem, M., & Anderson, C. A. (2007). Public policy and the effects of media violence on children. *Social Issues and Policy Review, 1*(1), 15–61. https://doi.org/10.1111/j.1751-2409.2007.00003.x

Gergely, G., Bekkering, H., & Király, I. (2002). Rational imitation in preverbal infants. *Nature, 415*(6873), 755. https://doi.org/10.1038/415755a

Goldstein, S. E., Tisak, M. S., & Boxer, P. (2002). Preschoolers' normative and prescriptive judgements about relational and overt aggression. *Early Education and Development*, *13*(1), 41–58. https://doi.org/10.1207/s15566935eed1301_2

Graham, J., Haidt, J., & Nosek, B. A. (2009). Liberals and conservatives rely on different sets of moral foundations. *Journal of Personality and Social Psychology*, *96*(5), 1029-1046. https://doi.org/10.1037/a0015141

Grossman, D. (1995). *On killing: The psychological cost of learning to kill in war and society*. Back Bay Books.

Hall, E. B. (1906). *The friends of Voltaire*. Smith, Elder, & Co.

Hamlin, J. K., Wynn, K., & Bloom, P. (2007). Social evaluation in preverbal infants. *Nature, 450*(7169), 557–559. https://doi.org/10.1038/nature06288

Hansen, C. H., & Hansen, R. D. (1988). Finding the face in the crowd: An anger superiority effect. *Journal of Personality and Social Psychology*, *54*(6), 917–924. https://doi.org/10.1037/0022-3514.54.6.917

Heck, I. A., Bregant, J., & Kinzler, K. D. (2021a). "There are no band-aids for emotions": The development of thinking about emotional harm. *Developmental Psychology, 57*(6), 913–926. https://doi.org/10.1037/dev0001187

Heck, I. A., Santhanagopalan, R., Cimpian, A., & Kinzler, K. D. (2021b). Understanding the developmental roots of gender gaps in politics. *Psychological Inquiry, 32*(2), 53–71. https://doi.org/10.1080/1047840X.2021.1930741

Helwig, C. C. (1995). Adolescents' and young adults' conceptions of civil liberties: Freedom of speech and religion. *Child Development*, *66*(1), 152-166. https://doi.org/10.2307/1131197

Helwig, C. C. (1998). Children's conceptions of fair government and freedom of speech. *Child Development*, *69*(2), 518-531. https://doi.org/10.2307/1132181

Helwig, C. C., & Prencipe, A. (1999). Children's judgments of flags and flag-burning. *Child Development*, *70*(1), 132-143. https://doi.org/10.1111/1467-8624.00010

Helwig, C. C., Zelazo, P. D., & Wilson, M. (2001). Children's judgments of psychological harm in normal and noncanonical situations. *Child Development*, *72*(1), 66-81. https://doi.org/10.1111/1467-8624.00266

Jambon, M., & Smetana, J. G. (2014). Moral complexity in middle childhood: Children's evaluations of necessary harm. *Developmental Psychology, 50*(1), 22–33. https://doi.org/10.1037/a0032992

Kalb, L. M., & Loeber, R. (2003). Child disobedience and noncompliance: A review. *Pediatrics*, *111*(3), 641-652. https://doi.org/10.1542/peds.111.3.641

Kanakogi, Y., Miyazaki, M., Takahashi, H., Yamamoto, H., Kobayashi, T., & Hiraki, K. (2022). Third-party punishment by preverbal infants. *Nature Human Behaviour*, *6*(9), 1234–1242. https://doi.org/10.1038/s41562-022-01354-2

Killen, M., Mulvey, K. L., Richardson, C., Jampol, N., & Woodward, A. (2011). The accidental transgressor: Morally-relevant theory of mind. *Cognition, 119*(2), 197-215. https://doi.org/10.1016/j.cognition.2011.01.006

Killen, M., & Smetana, J. G. (2015). *Origins and development of morality.* In M. E. Lamb & R. M. Lerner (Eds.), *Handbook of child psychology and developmental science: Socioemotional processes* (p. 701–749). John Wiley & Sons Inc.

King, G., Pan, J., & Roberts, M. E. (2013). How censorship in China allows government criticism but silences collective expression. *American Political Science Review*, 326-343. https://doi.org/10.1017/s0003055413000014

Kinzler, D.K., & Shutts, K. (2008). Memory for "mean" over "nice": The influence of threat on children's face memory. *Cognition, 107,* 775-783. https://doi.org/10.1016/j.cognition.2007.09.005

Kohlberg, L. (1969). Stage and sequence: The cognitive-developmental approach to socialization. In D. A. Goslin (Ed.), *Handbook of socialization theory and research*. (pp. 151–235). New York: Academic Press.

Kohlberg, L., & Kramer, R. (1969). Continuities and discontinuities in childhood and adult moral development. *Human Development*, *12*(2), 93–120. https://doi.org/10.1159/000270857

Leshin, R. A., Yudkin, D. A., Van Bavel, J. J., Kunkel, L., & Rhodes, M. (2022). Parents' political ideology predicts how their children punish. *Psychological Science*, 0(0). https://doi.org/10.1177/09567976221117154

Leslie, A. M., Knobe, J., & Cohen, A. (2006). Acting Intentionally and the Side-Effect Effect: Theory of Mind and Moral Judgment. *Psychological Science, 17*(5), 421–427. https://doi.org/10.1111/j.1467-9280.2006.01722.x

Liberman, Z., Howard, L. H., Vasquez, N. M., & Woodward, A. L. (2018). Children's expectations about conventional and moral behaviors of ingroup and outgroup members.

Journal of Experimental Child Psychology, 165, 7-18. https://doi.org/10.1016/j.jecp.2017.03.003

Lieberman, D., & Smith, A. (2012). It's all relative: Sexual aversions and moral judgments regarding sex among siblings. *Current Directions in Psychological Science, 21*(4), 243–247. https://doi.org/10.1177/0963721412447620

Marshall, J., & McAuliffe, K. (2022). Children as assessors and agents of third-party punishment. *Nature Reviews Psychology*, *1*(6), 334-344. https://doi.org/10.1038/s44159-022-00046-y

Marshall, J., Yudkin, D. A., & Crockett, M. J. (2020) Children punish third parties to satisfy both consequentialist and retributive motives. *Nature Human Behaviour*. https://doi.org/10.1038/s41562-020-00975-9

Mill, J. S. (1869). *On liberty*. Longman, Roberts & Green.

McLeod, D. M., Eveland Jr, W. P., & Nathanson, A. I. (1997). Support for censorship of violent and misogynic rap lyrics: An analysis of the third-person effect. *Communication Research*, *24*(2), 153-174. https://doi.org/10.1177/009365097024002003

Meyer-Rochow, V. B. (2009). Food taboos: their origins and purposes. *Journal of Ethnobiology and Ethnomedicine*, *5*(1), 18. https://doi.org/10.1186/1746-4269-5-18

Miller, J. G., Bersoff, D. M., & Harwood, R. L. (1990). Perceptions of social responsibilities in India and in the United States: Moral imperatives or personal decisions? *Journal of Personality and Social Psychology, 58*(1), 33–47. https://doi.org/10.1037/0022-3514.58.1.33

Mulvey, K. L., Palmer, S. B., & Abrams, D. (2016). Race-based humor and peer group dynamics in adolescence: Bystander intervention and social exclusion. *Child Development, 87*(5), 1379–1391. https://doi.org/10.1111/cdev.12600

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84*(3), 231–259. https://doi.org/10.1037/0033-295X.84.3.231

Nucci, L. P. (2001). *Education in the moral domain.* Cambridge University Press. https://doi.org/10.1017/CBO9780511605987

Patterson, M., Bigler, R.S., Pahlke, E., Brown, C.S., Hayes, A.R., Ramirez, M.C., & Nelson, A. (2019). Toward a developmental science of politics. *Monographs of the Society for Research in Child Development*, *84*(3).

Piaget, J. (1997). *The Moral Judgment of the Child*. New York, NY: Free Press Paperbacks. (Original work published 1932)

Palan, S., & Schitter, C. (2018). Prolific.ac—A subject pool for online experiments. *Journal of Behavioral and Experimental Finance*, *17*, 22-27. https://doi.org/10.1016/j.jbef.2017.12.004

Pfaff-Rüdiger, S., & Riesmeyer, C. (2016). Moved into action. Media literacy as social process. *Journal of Children and Media*, *10*(2), 164-172. https://doi.org/10.1080/17482798.2015.1127838

Potter, W. J. (2010). The state of media literacy. *Journal of Broadcasting & Electronic Media*, *54*(4), 675-696. https://doi.org/10.1080/08838151.2011.521462

Rai, T. S., & Fiske, A. P. (2011). Moral psychology is relationship regulation: Moral motives for unity, hierarchy, equality, and proportionality. *Psychological Review*, *118*(1), 57-75. https://doi.org/10.1037/a0021867

Reifen-Tagar, M., & Cimpian, A. (2022), Political ideology in early childhood: Making the case for studying young children in political psychology. *Political Psychology*. https://doi.org/10.1111/pops.12853

Reifen Tagar, M., Federico, C. M., Lyons, K. E., Ludeke, S., & Koenig, M. A. (2014). Heralding the authoritarian? Orientation toward authority in early childhood. *Psychological Science, 25*(4), 883–892. https://doi.org/10.1177/0956797613516470

Rideout, V., Peebles, A., Mann, S., & Robb, M. B. (2022). *Common Sense census: Media use by tweens and teens, 2021*. San Francisco, CA: Common Sense.

Riedl, K., Jensen, K., Call, J., & Tomasello, M. (2015). Restorative justice in children. *Current Biology*, *25*(13), 1731–1735. https://doi.org/10.1016/j.cub.2015.05.014

Schwartz, G. M., Izard, C. E., & Ansul, S. E. (1985). The 5-month-old's ability to discriminate facial expressions of emotion. *Infant Behavior & Development*, *8*(1), 65–77. https://doi.org/10.1016/S0163-6383(85)80017-5

Shaw, A., Montinari, N., Piovesan, M., Olson, K. R., Gino, F., & Norton, M. I. (2014). Children develop a veil of fairness. *Journal of Experimental Psychology: General, 143*(1), 363–375. https://doi.org/10.1037/a0031247

Shepperd, J., Malone, W., & Sweeny, K. (2008). Exploring causes of the self-serving bias. *Social and Personality Psychology Compass, 2*(2), 895–908. https://doi.org/10.1111/j.1751-9004.2008.00078.x

Shweder, R. A., Mahapatra, M., & Miller, J. G. (1987). *Culture and moral development.* In J. Kagan & S. Lamb (Eds.), *The emergence of morality in young children* (p. 1–83). University of Chicago Press.

Shultz, T. R., Wright, K., & Schleifer, M. (1986). Assignment of moral responsibility and punishment. *Child Development*, *57*(1), https://doi.org/10.2307/1130649

Smetana, J. G. (2013). Young children's moral and social-conventional understanding. In M. R. Banaji & S. A. Gelman (Eds.), *Navigating the social world: What infants, children, and other species can teach us* (pp. 352–355). New York: Oxford University Press.

Smetana, J. G., Jambon, M., & Ball, C. (2014). The social domain approach to children's moral and social judgments. In M. Killen & J. G. Smetana (Eds.), *Handbook of moral development* (2nd ed., pp. 23–45). Psychology Press.

Smith, C. E., Blake, P.R., Harris, P.L. (2013) I should but I won't: Why young children endorse norms of fair sharing but do not follow them. *PLOS ONE*, 8(3): e59510. https://doi.org/10.1371/journal.pone.0059510

Tamir, D. I., & Thornton, M. A. (2018). Modeling the predictive social mind. *Trends in Cognitive Sciences*, *22*(3), 201-212. https://doi.org/10.1016/j.tics.2017.12.005

Tasimi, A., & Johnson, M. K. (2015). A self-serving bias in children's memories? *Journal of Experimental Psychology: General, 144*(3), 528–533. https://doi.org/10.1037/xge0000072

Taylor, M., Cartwright, B. S., & Bowden, T. (1991). Perspective taking and theory of mind: Do children predict interpretive diversity as a function of differences in observers' knowledge? *Child Development, 62*(6), 1334–1351. https://doi.org/10.2307/1130810

Tetlock, P. E. (2003). Thinking the unthinkable: Sacred values and taboo cognitions. *Trends in Cognitive Sciences, 7*(7), 320–324. https://doi.org/10.1016/S1364-6613(03)00135-9

Tetlock, P. E., Kristel, O. V., Elson, S. B., Green, M. C., & Lerner, J. S. (2000). The psychology of the unthinkable: Taboo trade-offs, forbidden base rates, and heretical counterfactuals. *Journal of Personality and Social Psychology, 78*(5), 853–870. https://doi.org/10.1037/0022-3514.78.5.853

Tisak, M. S. (1993). Preschool children's judgments of moral and personal events involving physical harm and property damage. *Merrill-Palmer Quarterly*, *39*(3), 375–390.

Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences*, *28*(5), 675-

691. https://doi.org/10.1017/S0140525X05000129

Uhlmann, E. L., Pizarro, D. A., & Diermeier, D. (2015). A person-centered approach to moral judgment. *Perspectives on Psychological Science, 10*(1), 72-81. https://doi.org/10.1177/1745691614556679

Vaish, A., Carpenter, M., & Tomasello, M. (2009). Sympathy through affective perspective taking and its relation to prosocial behavior in toddlers. *Developmental Psychology*, *45*(2), 534–543. https://doi.org/10.1037/a0014322

Vaish, A., Carpenter, M., & Tomasello, M. (2010). Young children selectively avoid helping people with harmful intentions. *Child Development*, *81*(6), 1661-1669. https://doi.org/10.1111/j.1467-8624.2010.01500.x

Vaish, A., Carpenter, M., & Tomasello, M. (2010). Young children selectively avoid helping people with harmful intentions. *Child Development*, *81*(6), 1661–1669. https://doi.org/10.1111/j.1467-8624.2010.01500.x

Vaish, A., Grossmann, T., & Woodward, A. (2008). Not all emotions are created equal: the negativity bias in social-emotional development. *Psychological Bulletin*, *134*(3), 383-403. https://doi.org/10.1037/0033-2909.134.3.383

Vaish, A., Missana, M., & Tomasello, M. (2011). Three-year-old children intervene in third-party moral transgressions. *The British Journal of Developmental Psychology*, *29*(Pt 1), 124–130. https://doi.org/10.1348/026151010X532888

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. https://doi.org/10.1016/j.jecp.2017.07.004

Warf, B. (2009). Diverse spatialities of the Latin American and Caribbean internet. *Journal of*

Latin American Geography, *8*(2), 125–146. https://doi.org/10.1353/lag.0.0063

Warf, B. (2011). Geographies of global Internet censorship. *GeoJournal*, *76*(1), 1-23. https://doi.org/10.1007/s10708-010-9393-3

Warf, B., & Vincent, P. (2007). Multiple geographies of the Arab internet. *Area*, *39*(1), 83–96. https://doi.org/10.1111/j.1475-4762.2007.00717.x

West, C. (2018). Pornography and censorship. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Fall 2018 ed.). Stanford University. https://plato.stanford.edu/archives/fall2018/entries/pornography-censorship/

Woodward, A. L. (1998). Infants selectively encode the goal object of an actor's reach. *Cognition, 69*(1), 1–34. https://doi.org/10.1016/S0010-0277(98)00058-4

Worth, K. A., Gibson Chambers, J., Nassau, D. H., Rakhra, B. K., & Sargent, J. D. (2008). Exposure of US adolescents to extremely violent movies. *Pediatrics*, *122*(2), 306–312. https://doi.org/10.1542/peds.2007-1096

Yudkin, D. A., Van Bavel, J. J., & Rhodes, M. (2019). Young children police group members at personal cost. *Journal of Experimental Psychology: General*, *149*(1), 182–191. https://doi.org/10.1002/yea.1312