When Children Treat Condemnation as a Signal: The Costs and Benefits of Condemnation

4 authors, including:

Some of the authors of this publication are also working on these related projects:

Meaning of Punishment View project

Fairness View project

When Children Treat Condemnation as a Signal: The Costs and Benefits of Condemnation

Hannah Hok ⓘD
University of Chicago

Alia Martin
Victoria University of Wellington

Zachary Trail and Alex Shaw
University of Chicago

Condemnation is ubiquitous in the social world and adults treat condemnation as a costly signal. We explore when children begin to treat condemnation as a signal by presenting 4- to 9-year-old children ($N = 435$) with stories involving a condemner of stealing and a noncondemner. Children were asked to predict who would be more likely to steal as well as who should be punished more harshly for stealing. In five studies, we found that 7- to 9-year-old children treat condemnation as a signal—thinking that a condemner is less likely to steal and should be punished more harshly if caught hypocritically stealing later. We discuss the implications of these results for children's emerging understanding of signaling and moral condemnation.

People frequently make judgments about others and their actions as morally good or bad. Condemnation, defined specifically as the expression of disapproval toward an action, is ubiquitous in the day-to-day social world (Haidt, 2012; Rai & Fiske, 2011; Shweder, Mahapatra, & Miller, 1987). Adults not only condemn others regularly, but can use others' condemnation to make predictions about those individuals' future moral behavior (e.g., Alicke, Gordon, & Rose, 2013; Barden, Rucker, & Petty, 2005; DeScioli & Kurzban, 2009, 2013). For example, imagine two people, Andy and Ben. Andy strongly condemns promise-breaking and thinks people who break promises should be ostracized. Ben does not express an opinion about promise-breaking. Intuitively, who do you think would be less likely to break a promise? Furthermore, imagine that both Andy and Ben later go on to break a promise. Who do you think deserves more scorn? Recent research by Jordan, Sommers, Bloom, and Rand (2017) suggests that many adults would answer Andy to both questions: they predict that Andy will be less likely to break a promise, and should be punished more harshly if he does break a promise. The authors take this as evidence that adults treat condemnation as a signal of moral character and dislike those who falsely signal.

What does it mean to call condemnation a signal? In biology, a signal is a trait or behavior by an agent that conveys information to a receiver, providing a potential benefit to the signaler (e.g., Bird & Smith, 2005). Importantly, many signals also have costs or constraints associated with them that make it difficult for agents to signal dishonestly. For example, peacocks with large and brilliant plumage are more attractive to potential mates. However, this plumage can be costly to maintain because they make the peacock more conspicuous, and may impede its ability to escape from predators. The plumage may be considered an honest signal, because only high-quality mates can don them and attract females (Zahavi, 1975). Without associated costs, signals could be easily faked and thus would quickly be ignored by receivers (Bird & Smith, 2005; Zahavi, 1975).

If condemnation functions as a signal of moral character, it should be the case that condemnation is associated with both a benefit to the signaler and a cost that keeps the signal honest. Jordan et al. (2017) provide evidence in line with this view: that is, condemnation seems to provide the condemner with the reputational benefit of being perceived by others as less likely to commit the condemned moral transgression, but also can be costly to the

condemner by leading to harsher punishment in the event of a hypocritical transgression. However, there has been no research that we are aware of directly examining how children form expectations about others based on their condemnation of immoral actions. Understanding how and when children use condemnation to predict others' future behavior may help constrain hypotheses about the kinds of cognitive capacities or experiences that are necessary to make such inferences, and provide insight into children's developing ability to use social information to make predictions about others' behavior. The current work examines whether children treat condemnation of an immoral action (stealing) as a social signal, and compares this signal to other potential signals of moral goodness (e.g., someone who praises sharing or denies that they steal).

Social and moral condemnations serve as commonplace in children's daily lives, providing them with many opportunities to observe condemnation, form expectations about others based on condemnation, and make inferences based on condemnation. In the first 2 years of life, infants already make negative social evaluations of others, showing a preference for characters who are helpful or fair over those who are harmful or unfair (Geraci & Surian, 2011; Hamlin, Wynn, & Bloom, 2007). As children grow older, their condemnation becomes increasingly sophisticated and they inflict costs on others for their wrongdoing (Hamlin, Wynn, Bloom, & Mahajan, 2011; Tasimi & Wynn, 2016). For example, by age 3 or 4, children clearly differentiate conventional violations (i.e., things that are only wrong because someone says they are bad) from moral violations (i.e., things that are wrong no matter what others say; Helwig & Jasiobedzka, 2001; Schmidt, Rakoczy, & Tomasello, 2012; Smetana & Braeges, 1990; Turiel, 1998). Their evaluations of moral transgressions are also influenced by group membership (Dunham, Baron, & Banaji, 2008; Rhodes & Chalik, 2013; Rizzo, Cooley, Elenbaas, & Killen, 2018), fairness (Grocke, Rossano, & Tomasello, 2015; Schmidt, Svetlova, Johe, & Tomasello, 2016; Shaw & Olson, 2014), the intentions of a perpetrator (Cushman, Sheketoff, Wharton, & Carey, 2013; Darley & Shultz, 1990; Yuill & Perner, 1988), and whether the transgression violates the letter or spirit of a rule (Bregant, Wellbery, & Shaw, 2019). Furthermore, although there are cross-cultural differences in terms of what counts as a moral violation, children across cultures around the world condemn the actions of others (Wainryb & Recchia, 2013). Given how adept children are at engaging in

moral condemnation and how ubiquitous it is in their daily lives, it seems plausible that even preschool-age children could use others' condemnation to make predictions about their future behavior.

However, engaging in condemnation may be considerably easier than knowing to treat condemnation as a signal of an agent's future moral behavior. While the former only requires children to judge that an action is wrong from their own perspective, the latter requires children to engage in a more abstract process of considering how the current behavior of a third party might relate to that individual's behavior in the future. Additionally, responding negatively to hypocrisy may require an inference that another person is trying to dishonestly manage their reputation (i.e., modify what others think of them). Seven- to 8-year-old children may find making such inferences difficult because doing so may rely on second-order theory of mind abilities (Banerjee & Yuill, 1999) which continue to improve during middle childhood (Miller, 2009). Relatedly, children under age 8 fail to appreciate that public generosity may reflect a disingenuous attempt to appear generous. In particular, 6- and 7-year-olds do not judge someone who acts generously in private to be a nicer person than someone who acts generously in public, but 8- to 10-year-olds do (Heyman, Barner, Heumann, & Schenck, 2014). Furthermore, although even children as young as age 3 dislike liars (Lane, Wellman, & Gelman, 2013), it is not until around age 7 that children fully appreciate the differences between lies that benefit others and lies that benefit the self (Fu, Heyman, Chen, Liu, & Lee, 2015). Indeed, previous research has suggested that although children have some reputational concerns as early as age 5 (Engelmann, Herrmann, & Tomasello, 2012; Engelmann, Over, Herrmann, & Tomasello, 2013; Leimgruber, Shaw, Santos, & Olson, 2012), they become more adept at reputation management, and particularly at recognizing others' deceptive signaling behavior, between the ages of 7 and 9 (Aloise-Young, 1993; Apfelbaum, Pauker, Ambady, Sommers, & Norton, 2008; Fu et al., 2015; Heyman & Legare, 2005; Heyman et al., 2014; Mills & Keil, 2005; Shaw & Olson, 2015; Shaw et al., 2014; Silver & Shaw, 2018; Tasimi, Dominguez, & Wynn, 2015). If children have difficulty understanding deceptive acts of signaling before the age of 6 or 7, then they might not begin to treat condemnation as a signal until they are older than 7. On the other hand, it is possible that children's experiences observing and engaging in moral condemnation from a very young age (for review, see Bloom, 2013) could allow them to

appreciate its signaling value earlier in development.

While there is no specific work to suggest that children treat condemnation as a social signal, there is some work suggesting that children understand the deterrence value of punishment and positively evaluate those who punish wrongdoers. By the time children are 5 or 6 years old, they recognize that those who are punished for their immoral actions are less likely to transgress again than those who are not punished (Bregant, Shaw, & Kinzler, 2016). Most relevantly, Vaish, Herrmann, Markmann, and Tomasello (2016) found that 5-year-old children prefer those who enforce norms by sanctioning transgressors more than those who do not. These studies demonstrate that children treat others' condemnation of negative actions (via punishment) positively. However, this work does not address whether children use others' condemnation to make predictions about their future behavior, or whether children consider hypocrisy when deciding who should be punished. Indeed, if children simply view condemnation and punishment of bad behavior as positive (i.e., if they just want to see bad consequences for bad actions), then we would not necessarily expect harsher punishments for hypocritical transgressions than nonhypocritical ones.

In the current experiments, we explore if and when children treat moral condemnation as a signal of future moral behavior. To explore this question, we presented 4- to 9-year-old participants with a story in which they were told about two children, one child who condemned stealing and one who did not. Participants were then told that someone had stolen something in the classroom and were asked to predict which of the two children had stolen. Participants were specifically asked to make this prediction before receiving any information about the two children's actual behavior. After participants made their prediction, they were told that both children had stolen and were asked which of the two should be punished more harshly for stealing. If children treat condemnation as a signal, then they should think that the condemner is less likely to transgress (in this example, steal), but also that the condemner should be punished more harshly than the noncondemner for being caught transgressing (stealing).

We predicted that older children would be more likely than younger children to treat condemnation as a costly signal, and split the children into two age groups, 4- to 6-year-olds (henceforth "younger children") and 7- to 9-year-olds (henceforth "older children"). We chose these age ranges because, as

we reviewed above, previous research suggests that it is around the age of 7 or 8 that children begin to develop a more sophisticated understanding of the way others try to manage their reputations through signaling (for review, see Banerjee, 2002; Heyman et al., 2014; Shaw, Li, & Olson, 2013). We predicted that older children would treat condemnation as a signal, inferring that condemners are less likely to engage in the behavior that they condemn (i.e., bestowing the benefit of believing that the person will act morally in this regard), and judging that the condemner should be punished more harshly if she is caught hypocritically engaging in the behavior that she condemned (i.e., levying a cost for signaling dishonestly). We made no specific predictions about younger children except that they should be less likely to treat condemnation as a signal than the older children.

Study 1

Study 1 explored if and when children begin to treat moral condemnation as a costly signal of future moral behavior. To investigate this question, 4- to 9-year old participants were told about two children: one who condemned stealing as wrong (condemner), and the other who made a negative but nonmoral statement about finding broccoli gross (noncondemner). Next, they were told that these two children were suspected of stealing something, and asked to predict which child they thought had stolen (prediction measure). After they answered, participants were told that, actually, both children had stolen something, and were asked which of the two children should be punished more harshly for stealing (punishment measure). If participants treat moral condemnation as a signal, then they should both predict that the condemner was less likely to steal and judge that the condemner should be punished more harshly for stealing.

Method

Participants

One hundred and twenty-one 4- to 9-year-olds (M_{age} = 88.10 months, SD = 26.16 months, 56 female) were tested in Study 1. For analysis, data were broken down into two age groups, sixty 7- to 9-year-olds (M_{age} = 101.76 months, SD = 10.39 months, 25 female) and sixty-one 4- to 6-year-olds (M_{age} = 67.15 months, SD = 10.44 months, 28 female).

We begin by outlining some general methods for the reported studies. For all studies, participants were recruited and tested at a local science museum in the Midwest. Written informed consent was given by the parents and then verbal assent of the child was secured by researchers before experiments. The Institutional Review Boards approved all procedures. Participants were compensated with stickers for their participation in the study.

Due to variability in the daily number of visitors, we set our stopping criterion for data collection at the end of the day on which there were at least 60 participants per age group per condition for Studies 1 and 3. Studies 1 and 3 were done as part of one of the authors' senior thesis and because we did not know the nature of the effect size, we aimed to collect 60 participants per age group. For all subsequent studies (except Study 3, which we explain below), we set our stopping criterion for data collection at the end of the day on which there were at least 30 participants per age group per condition. For all studies, sample sizes were determined ahead of time. We designed these studies to be run quickly at the museum and, as such, did not ask parents to fill out detailed demographic forms. We collected information only about children's birthdates and gender. We found no effect of gender in any of our studies, so we collapsed across gender.

Procedure

Participants were told a story about two children, represented by cardboard dolls. In the story, participants were introduced to the two children described as classmates, and were told that one of the children "always talks about how much (s)he dislikes stealing and tells other people that stealing is really, really bad" (condemner) and that the other "always talks about how much (s)he dislikes broccoli and tells other people that broccoli is really, really gross" (noncondemner). Participants were then told that after class, something was stolen. They were reminded of what each character said, and were asked: "Who do you think stole?" (prediction measure). After making their predictions, participants were informed that, actually, both of the characters independently stole money from their respective parent after school. Specifically, children were told: "Actually both of these boys stole after school. This boy [L] went home and stole money from his mom's purse and his mom caught him, and this boy [R] went home and stole money from his dad's wallet and his dad caught him. So they both got caught stealing." Participants were

then asked which character "should be punished more for stealing?" (punishment measure). The names of the children in the story were gender-neutral and dolls were gender matched to participants. The order of presentation (condemner presented first or second), character positioning (condemner on the right or left), and who stole from which parent (mother or father) were counterbalanced.

After giving their responses, participants were given two comprehension checks. They were asked, "Which character said stealing was 'really, really bad'?" and, "Which character said broccoli was 'really, really gross'?". No children failed comprehension checks.

Results

Prediction Measure

A logistic regression analysis was conducted to reveal whether children's predictions of which character was more likely to steal changed with age. Age was set as the continuous factor. The analysis revealed an effect of age on choice, Wald $\chi^2(1, N = 121) = 34.37$, $p < .001$. With age, children were less likely to predict that the condemner was the one who stole. We then examined children's choices by our preset age groups using binomial sign tests. These tests revealed that older children predicted that the condemner would steal at levels significantly below chance (13 out of 60, 22%, $p < .001$), whereas younger children predicted that the condemner would steal at levels significantly above chance (39 of 61, 64%, $p = .035$). That is, older children suspected that the condemner was less likely to steal than the noncondemner, whereas younger children thought the condemner was more likely to steal.

Punishment Measure

A logistic regression analysis was conducted to reveal whether children's judgments of which character should receive harsher punishment changed with age. Age was again set as the continuous factor. The analysis revealed an effect of age on choice, Wald $\chi^2(1, N = 121) = 4.47$, $p = .035$. With age, children were more likely to say that the condemner should receive harsher punishment than the noncondemner. We then examined children's choices by our preset age groups using binomial sign tests. These tests revealed that both older children (51 out of 60, 85%, $p < .001$) and younger children (42 out of 61, 69%, $p = .035$) desired harsher punishment

for the condemner at levels significantly above chance. That is, both older and younger children wanted condemners to be punished more harshly than noncondemners (see Figure 1).

Discussion

Results from Study 1 demonstrate that by the time children are 7 years old, they can treat condemnation as a costly signal. Seven- to 9-year-olds predicted that a condemner was less likely to steal than a noncondemner, and judged that the condemner should be punished more harshly than the noncondemner when caught stealing. Younger children showed a different pattern and did not predict that the condemner was less likely to steal than the noncondemner. Interestingly, not only did younger children not give the condemner the benefit of the doubt; they actually thought the condemner was more likely to steal than the noncondemner. We had no specific predictions about younger children's responses, and from this result alone it is unclear why younger children responded the way they did. However, it could be that younger children were merely latching onto the word "stealing" and selecting the character who was associated with stealing (we return to this question in Studies 3 and 4). Younger children, like older children, thought that the condemner should be punished more harshly than the noncondemner when caught stealing. However, given that they predicted that the condemner was more likely to steal, it seems that they are likely making the punishment judgment for different reasons than older children (we provide additional evidence for this possibility in Study 2b).
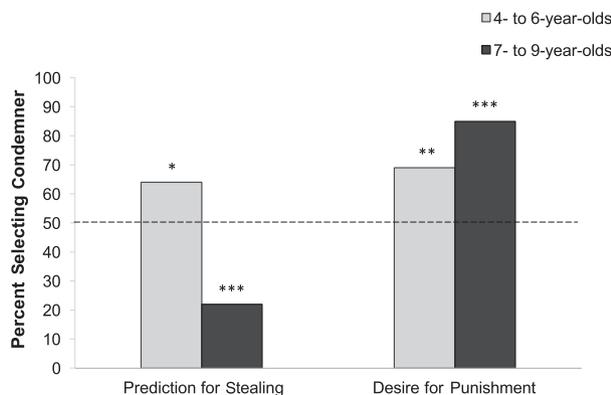
Although the response from younger children here is not completely clear, the results of Study 1 provide preliminary evidence that, by around age 7, children recognize that condemnation functions as a beneficial social signal that should have an accompanying cost for those who signal dishonestly.

Study 2a

Study 1 provided some support for the notion that older children treat condemnation as a signal. However, comparing someone who makes a statement about moral content to someone who makes a neutral, nonmoral statement was not a very stringent test of the signaling hypothesis. Study 1 may merely reveal that 7- to 9-year-olds expect people who make statements with moral content to be more likely to engage in moral behavior. Thus, in Study 2a we had children compare a condemner to a noncondemner who made a different moral statement. Specifically, the noncondemner was someone who praises sharing (a positive moral action) and thinks it is very good to share. Participants were again asked the prediction (Who stole?) and punishment (Who should be punished more harshly if they are caught stealing?) questions from Study 1. If children simply use any statement about morality to make predictions about others' moral behavior, then they might now predict that both are equally (un)likely to steal. Although both characters made statements consistent with being a moral person, we predicted that older children would again take the condemnation of stealing to be a stronger signal than the praising of sharing that one would refrain from stealing.

Method

Participants

Again, participants were recruited and run at a local science museum. As noted above, we stopped collecting data after a museum shift ended with 30 participants in each age group. Sixty 4- to 9-year-olds (M_{age} = 77.93 months, SD = 21.61, 35 female) were tested. Again, data were broken down into two age groups: twenty-five 7- to 9-year-olds (M_{age} = 99.84 months, SD = 9.93 months, 10 female) and thirty-five 4- to 6-year-olds (M_{age} = 62.16, SD = 11.48 months, 25 female).

Figure 1. Study 1. Mean percent of children in each age group selecting the condemner of stealing over the non-condemner who dislikes broccoli (*p < .05. **p < .01. ***p < .001).

Procedure

Participants were told stories similar in structure to those in Study 1. The same dolls were used to represent the characters, and were again gender matched to participants with the same counterbalancing. Participants were again told of two classmates, one who condemns stealing and one who does not, except that here the noncondemner praises sharing ("S(he) always talks about how sharing is really important and when (s)he sees someone share, (s)he says 'Sharing is really, really nice'"). Participants were asked to predict which classmate (the condemner or the praiser) stole and who should be punished more harshly for stealing. Again, participants were given two comprehension checks. They were asked to remind the experimenter which child said stealing was "really, really bad" and which child said sharing was "really, really good." Four children answered both comprehension checks incorrectly, but the pattern of results does not change if the responses of these four children are excluded from analysis.

Results

Prediction Measure

A logistic regression analysis was conducted to reveal whether children's predictions of which character was more likely to steal changed with age. Age was set as the continuous factor and character prediction (between condemner and praiser) as the binary dependent variable. The analysis revealed an effect of age on choice, Wald $\chi^2(1, N = 60) = 17.32$, $p < .001$. With age, children were less likely to predict that the condemner was the one who stole. We then examined children's choices in each age group using binomial sign tests. These tests revealed that older children suspected the condemner of stealing at levels significantly below chance (7 of 25, 28%, $p = .043$) whereas younger children did so at levels significantly above chance (30 of 35, 86%, $p < .001$). That is, older children suspected that the condemner was less likely to steal than the praiser, whereas younger children thought the condemner was more likely to steal than the praiser.

Punishment Measure

A logistic regression analysis was conducted to reveal whether children's preferences of which character should be more punished changed with age. Age was again set as the continuous factor and
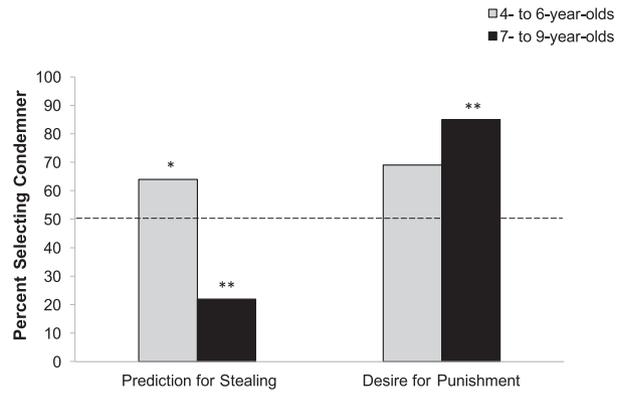

Figure 2. Study 2a results. Mean percent of children in each age group selecting the condemner of stealing over the praiser of sharing (*$p < .05$. **$p < .01$. ***$p < .001$).

character prediction (between condemner and praiser) as the binary dependent variable. The analysis revealed that age did not affect choice, Wald $\chi^2(1, N = 60) = 1.55$, $p = .213$. That is, children's judgments of who should be punished more for stealing was not significantly affected by age. We then examined children's choices in each age group using binomial sign tests. These binomial sign tests revealed that older children judged that the condemner should be punished more harshly for stealing than the praiser (21 out of 25, 84%, $p < .001$). Younger children marginally also judged that the condemner should be punished more harshly for stealing than the praiser (22 out of 35, 63%, $p = .088$; see Figure 2).

Discussion

We again found that 7- to 9-year-old children treated moral condemnation as a signal, whereas younger children did not. That is, even when pitted against a noncondemner who made a positive moral statement about the importance of sharing, older children still predicted that the condemner was less likely to steal. Furthermore, they judged that the condemner should be punished more harshly than the noncondemner when later caught stealing. We again found that younger children made the opposite prediction of older children, predicting that the condemner of stealing was more likely to steal, but were consistent with older children in preferring harsher punishment for the condemner.

Study 2b

The results of Study 2a suggest that older children treat moral condemnation as a signal, predicting

that someone who condemned stealing would be less likely to steal than someone who praised sharing. However, do children always treat condemnation as a better signal of future morally good behavior than expressing a positive moral attitude (e.g., praising sharing)? One possibility is that children see condemnation of bad behavior as particularly morally good, and therefore treat this condemnation as a better predictor of future good behavior than the act of expressing a positive moral attitude. If so, then condemnation should not only serve as a signal that one refrains from engaging in the specific bad behavior one condemns (in this case, stealing), but also as a stronger signal of engaging in other positive moral behaviors like sharing.

In Study 2b, we explored this possibility. Specifically, we presented participants with a similar story to Study 2a, with one character who condemned stealing and one who praised sharing. However, this time, participants were asked to predict who would share. Participants were then told that both characters failed to share, and were asked who should be punished more harshly for not sharing. If condemnation of a bad behavior like stealing is generally viewed as a more moral act than praising a good behavior like sharing, then older children should again respond as they did in our previous study, predicting that the condemner is more likely to share and perhaps that they should also be punished more harshly for failing to share. If, instead, older children use condemnation as a more specific signal, then here we should find a different pattern for children's prediction and punishment judgments.

Method

Participants

Again, participants were recruited and run at a local science museum. Sixty-eight 4- to 9-year-olds (M_{age} = 78.72 months, SD = 19.8 months, 42 female) were tested, Again, data were broken down into two age groups: thirty 7- to 9-year-olds (M_{age} = 97.62 months, SD = 11.52 months, 18 female) and thirty-eight 4- to 6-year-olds (M_{age} = 63.72, SD = 9.06 months, 24 female).

Procedure

Participants were told stories nearly identical to those in Study 2a. The same dolls were used to represent the characters and again gender matched to participants with the same counterbalancing.

Participants were again told of the same two classmates, one who condemns stealing and one who praises sharing. However, instead of being told that someone had stolen something, participants were told that someone had failed to share. Participants were then asked to predict which classmate (the condemner or the praiser) had shared. Then they were told that, in fact, both had failed to share, and were asked who should be punished more harshly for it. Again, participants were given two comprehension checks. They were asked to remind the experimenter which child said stealing was "really, really bad" and which child said sharing was "really, really good." Two children failed both comprehension checks but the pattern of results does not change if the responses of these two children are excluded from analysis.

Results

Prediction Measure

A logistic regression analysis was conducted to reveal whether children's predictions of which character was more likely to share changed with age continuously. Age was set as the continuous factor and character prediction (between Condemner and Praiser) as the binary dependent variable. The analysis revealed no effect of age on choice, Wald $\chi^2(1, N = 68)$ = 1.39, p = .239. We then examined children's choices by our preset age group using binomial sign tests. Binomial sign tests revealed that older children predicted the praiser of sharing at levels significantly above chance (26 out of 30, 86%, p < .001) as did younger children (29 of 38, 76%, p = .002). That is, older children and younger children both suspected that the praiser of sharing was more likely to share than the condemner of stealing.

Punishment Measure

A logistic regression analysis was conducted to reveal whether children's preferences of which character should be more punished changed with age. Age was again set as the continuous factor and character prediction (between condemner and praiser) as the binary dependent variable. The analysis revealed that age affected choice, Wald $\chi^2(1, N = 68)$ = 12.24, p < .001. Specifically, children were more likely to pick the praiser for harsher punishment with age. Binomial sign tests revealed that older children thought the praiser of sharing should be punished more harshly for failing to share than the condemner (20 of 30, 69%; p = .049). Younger

children, however, thought the condemner should be punished more harshly for failing to share than the praiser significantly above chance (12 out of 38, 32%, p = .034; see Figure 3).

Discussion

In line with our predictions, we found that 7- to 9-year-olds did not treat condemnation as a generally better signal of positive moral behavior than the expression of a positive moral attitude. When a moral condemner was pitted against a praiser, older children made a specific prediction based on condemnation: they predicted that a condemner of stealing was less likely to steal than a praiser of a good behavior like sharing (Study 2a), but did not predict that the condemner of stealing was more likely to share than the praiser of sharing (Study 2b). This indicates some specificity to condemnation as a signal for older children. Older children were also more likely to desire harsher punishment for the praiser of sharing when both characters had failed to share. This pattern of punishment suggests that older children are paying attention specifically to hypocrisy and false signaling: they are more likely to punish the character that they predict will do the right moral action (being less likely to steal in Studies 1 and 2a; being more likely to share in Study 2b) once the character fails to do so.

Study 2b additionally provides some clarification about young children's judgments. In the previous two studies, young children chose the same character in their predictions of who stole and their judgments of punishment. This pattern of responding could have arisen from a desire to be consistent. However, here in Study 2b the younger children

predicted that the praiser of sharing was more likely to share, but that the condemner of stealing should be punished more harshly for failing to share. These results show that younger children's prediction responses and punishment responses are not always linked. Instead, it seems possible that the 4- to 6-year-olds simply heard the word "stealing" associated with one of the characters and therefore judged that character to be morally bad; this could explain why they predicted that this character would be more likely to steal (Studies 1 and 2a), would be less likely to share (Study 2b), and deserves harsher punishment (all studies). We explore this possibility further in Studies 3 and 4.

Study 3

Studies 1 and 2a provide support for the idea that older children treat moral condemnation as a signal; however, it is unclear if the judgments that children made in these studies were specific to condemnation itself, or merely the expression of a negative attitude about stealing. To explore this possibility, Study 3 asked children to make predictions about the behavior of a condemner and a noncondemner who both expressed attitudes that would imply that they do not steal (modeled after similar vignettes in adults used by Jordan et al., 2017). One of them was the condemner of stealing (a person who said "stealing is really, really bad") and the other was someone who denied that they stole (a person who said "I never steal"). In this design, children do not just learn that both characters make statements about stealing, but that both make statements that might lead others to believe that they are unlikely to steal. If older children, like adults, treat condemnation as a particularly salient signal of one's moral character, then they should say that the condemner is less likely to steal than the denier. Put another way, we predicted that, with a more stringent test, we would replicate the pattern of results from Studies 1 and 2a: older children should predict that the condemner is less likely to steal than the noncondemner (this time a "denier"). However, because the condemner was being pitted against a character who is explicitly lying, we did not make a specific prediction that older children would desire harsher punishment for either character.

This design also allowed us to further explore younger children's inferences. We have suggested that younger children may predict that the condemner is more likely to steal because the condemner is
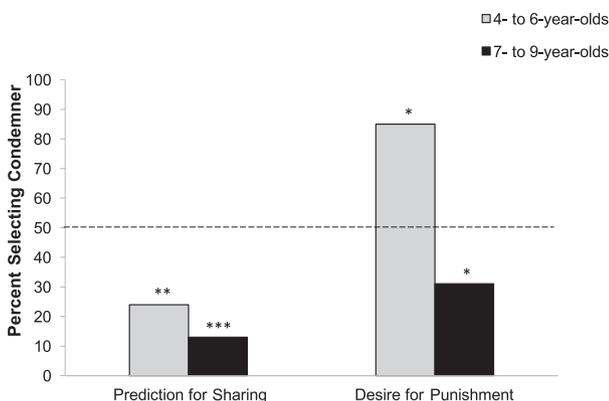

Figure 3. Study 2b results. Mean percent of children in each age group selecting the condemner of stealing over the praiser of sharing (*p < .05. **p < .01. ***p < .001).

the only character associated with the word "stealing." If this is true, then we should expect younger children to select at chance between the characters in this vignette. Alternatively, perhaps younger children really do treat moral condemnation as a signal of immoral behavior, in which case they should again predict that the condemner is more likely to steal than the noncondemner (this time the "denier").

Method

Participants

Again, participants were recruited and run at a local science museum. We stopped collecting data on a shift that resulted in ~60 participants per age group. One hundred and nineteen 4- to 9-year-olds (M_{age} = 80.69 months, SD = 19.32, 47 female) were tested. Again, data were broken down into two age groups: fifty-one 7- to 9-year-olds (M_{age} = 99.83 months, SD = 9.91 months, 20 female) and sixty-eight 4- to 6-year-olds (M_{age} = 66.36 months, SD = 9.79 months, 27 female).

Procedure

Participants were told a story similar in structure to the vignette in Study 1, the same dolls were used to represent the characters, and they again were gender matched to participants with the same counterbalancing. Children were again told of two classmates, one who condemns stealing ("always tells other people that stealing is really, really bad, and when (s)he sees someone steal, (s)he says that person should be punished for stealing") and one who does not, except that here the noncondemner was a denier of stealing ("always tells other people, 'I never steal', and when (s)he sees someone steal, (s)he says 'I would never steal like that'"). After being introduced to these characters, children were asked the same prediction measure ("Who do you think stole?") and punishment measure ("Who should be punished more for stealing?") from the previous studies. Again, participants were given two comprehension checks. They were asked to remind the experimenter which child said stealing was "really, really bad" and which child said "I never steal." Six children answered both comprehension checks incorrectly, but the pattern of results does not change whether or not the responses of these six children are excluded from analysis.

Results

Prediction Measure

A logistic regression analysis was conducted to reveal whether children's predictions of which character was more likely to steal changed with age. Age was set as the continuous factor and character prediction (between condemner and denier) as the binary dependent variable. The analysis revealed an effect of age on choice, Wald χ^2(1, N = 119) = 13.88, p < .001. With age, children were less likely to predict that the condemner was the one who stole. Binomial sign tests revealed that older children predicted that the condemner would steal at levels significantly below chance (9 of 51, 18%, p < .001), whereas younger children's predictions did not differ from chance (31 of 68, 46%, p = .544). That is, older children expected the condemner would be less likely to steal than the denier, but younger children did not have any particular prediction.

Punishment Measure

A logistic regression analysis was conducted to reveal whether children's judgments of which character should receive harsher punishment changed with age. Age was set as the continuous factor and character prediction (between condemner and denier) as the binary dependent variable. The analysis revealed that there was an effect of age on choice, Wald χ^2(1, N = 119) = 6.56, p = .01. With age, children were less likely to pick the condemner for harsher punishment. Binomial sign tests revealed that older children chose the condemner for harsher punishment at levels significantly below chance (16 of 51, 31%, p = .002), instead desiring harsher punishment for the denier, whereas younger children desired marginally harsher punishment for the condemner (42 of 68, 62%, p = .068) than the denier (see Figure 4).

Discussion

Providing even more robust evidence for the signaling account, we again found that 7- to 9-year-olds predicted that someone who condemned stealing is less likely to steal, this time even when compared to someone who directly said that they do not steal. This speaks to the salience of condemnation as a signal for these children. One might intuitively expect that directly
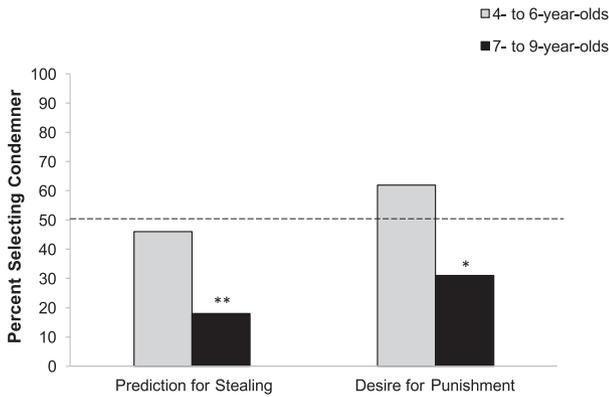
Figure 4. Study 3 results. Mean percent of children in each age group selecting the condemner of stealing over the denier of stealing (*p < .05. **p < .01. ***p < .001).

denying a behavior would be more indicative of future behavior than condemnation which says nothing directly about personal behavior. Additionally, both statements expressed negatively valenced attitudes about the transgression at hand. However, condemnation was still treated as a stronger signal of future behavior than denial, at least by older children. We additionally found that younger children did not predict that the condemner was less likely to steal and, instead, they thought the condemner and noncondemner were equally likely to steal. This suggests that younger children in our previous studies may have just been selecting the character associated with stealing instead of making a prediction based on the condemnation itself. Replicating our previous studies, we again found that older children treat condemnation as a signal of future moral behavior, whereas younger children do not.

Interestingly, in Study 3, older children punished deniers more harshly than condemners, suggesting that saying "I never steal" is a particularly bad strategy. We note that this result was not predicted, and is inconsistent with what Jordan et al. (2017) found in adults. Adults thought the condemner should be punished more harshly, presumably because they take condemnation to a be a stronger signal of moral behavior than mere denial. The older children also took condemnation to be a stronger signal than denial, but then why do older children not believe that person should be punished more harshly, and instead think the denier should be punished more harshly? It is worth noting that both behavioral patterns (denying stealing and then doing it anyway; condemning stealing and then doing it anyway) suggest some level of dishonesty, which children think is

bad (Bussey, 1999). It may merely be easier for the older children to see the dishonesty in someone saying "I do not do X" and then doing X, than it is for them to see the dishonesty in condemning X and then doing X. The former involves an explicit lie, whereas the latter is only tacitly a form of lying. However, we again caution against an interpretation of these particular results that is too rich, as they were not predicted and have not yet been replicated. Importantly, the results from Studies 1, 2a, and 2b suggest that older children do think that hypocritical condemners should be punished more harshly than noncondemners who do not also lie, indicating that 7- to 9-year-olds do attach some cost to falsely signaling one's moral behavior through condemnation.

Study 4

Our first three studies demonstrate that at least by the time children are 7 years old, they treat condemnation as a signal of future behavior. Our studies so far provide more mixed results with younger children. One explanation is that younger children merely did not understand our methods and had trouble following the vignettes, particularly when both actors were associated with the word "stealing" as in Study 3. Alternatively, younger children may specifically struggle to use condemnation to predict behavior. Study 4 attempts to delineate between these two explanations. If younger children are merely having trouble following the vignettes, then they should not be able to predict stealing behavior based on a character's explicitly positive attitude towards stealing. Study 4 uses the same paradigm as Studies 1–3 to ask children about a character who condemns stealing and a character who praises stealing. If younger children are confused by our methods, then they should again make predictions at chance (as seen in Study 3) since both characters are associated with stealing. However, we hypothesize that children should be able to use positive attitudes toward stealing to predict behavior, and will expect the character who praises stealing to be more likely to steal than the condemner. Expressing positive attitudes toward stealing might indicate a preference for that behavior, and we know that younger children can make predictions about others' behavior based on preferences (Heyes & Frith, 2014; Wimmer & Perner, 1983). We had no specific predictions about who children would think should be punished more harshly: a

condemner of stealing (e.g., for being hypocritical) or a praiser of stealing (e.g., for endorsing an immoral behavior that they engage in).

Method

Participants

Again, participants were recruited and run at the local science museum. Sixty-seven 4- to 9-year-olds (M_{age} = 81.30 months, SD = 18.71, 39 female) were tested. Again, data were split and analyzed by age group: twenty-nine 7- to 9-year-olds (M_{age} = 99.24 months, SD = 9.74 months, 16 female) and thirty-eight 4- to 6-year-olds (M_{age} = 67.56 months, SD = 10.22 months, 23 female). No children were excluded from analysis.

Procedure

Participants were told stories similar in structure to the previous studies. The same dolls were used to represent the characters, and again gender matched to participants with the same counterbalancing. Children were again told of two classmates, one who condemns stealing and one who does not, except that here the noncondemner actually praises stealing ("(S)he tells people how much she likes stealing, and that stealing is really, really good"). Participants were asked to predict which character (the condemner or the praiser) stole. It is revealed that both characters stole, and participants were asked who should be punished more harshly for stealing. Comprehension checks were dropped for ease of running since very few participants in our previous studies were excluded based on these checks.

Results

Prediction Measure

A logistic regression analysis was conducted to reveal whether children's predictions of which character was more likely to steal changed with age. Age was set as the continuous factor and character prediction (between condemner and praiser) as the binary dependent variable. The analysis revealed an effect of age on choice, Wald χ^2(1, N = 67) = 4.53, p = .033. With age, children were significantly less likely to predict that the condemner was the one who stole. Binomial sign tests revealed that both older children (5 of 29, 17%, p < .001) and younger children (11 of 38, 29%, p = .014) predicted that the

condemner would steal at levels significantly below chance. That is, older children and younger children both expected the condemner would be less likely to steal when compared to someone who praised stealing.

Punishment Measure

A logistic regression analysis was conducted to reveal whether children's judgments of which character should receive harsher punishment changed with age. Age was set as the continuous factor and character prediction (between condemner and praiser) as the binary dependent variable. The analysis revealed that age did not affect choice, Wald χ^2(1, N = 67) = 2.04, p = .153. Binomial sign tests revealed that older children did not differentially desire punishment for the condemner (16 out of 29, 55%, p = .711) and praiser of stealing, whereas younger children were significantly less likely to select the condemner (8 of 38, 21%, p < .001) than the praiser. In other words, older children's judgments of punishment were at chance, whereas younger children desired harsher punishment for the praiser of stealing than the condemner (see Figure 5).

Discussion

In line with our predictions, both older and younger children predicted that a character who praises stealing would be more likely to steal than a character who condemns stealing. These results suggest that younger children are not merely confused by our paradigm: If younger children were, in fact, only selecting who might steal based on who is associated with the word "stealing," we
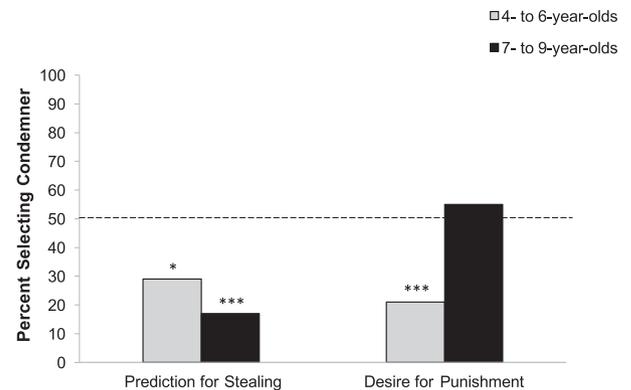
Figure 5. Study 4 results. Mean percent of children in each age group selecting the condemner of stealing over the praiser of stealing (*p < .05. **p < .01. ***p < .001).

ought to have seen chance results as in Study 3. However, even the younger children predicted that someone who praised stealing was more likely to steal than someone who condemned stealing. If children treat the positive attitude toward stealing as a preference, it makes sense that younger children can make these predictions, as previous work has indicated that even younger children can predict future behavior based on preference (Heyes & Frith, 2014; Wimmer & Perner, 1983). Taken together with our previous results, it seems that young children have a specific difficulty with making predictions about one's future behavior based on condemnation.

We did not have any particular predictions about how children would punish the praiser or the condemner; the praiser seems like a more of an ostensibly "bad" person, but the condemner is hypocritical. In our study, we found that older children were at chance in deciding whether the praiser or condemner of stealing should be punished more harshly. Interestingly, younger children desired harsher punishment for the character who praised stealing over the character who condemned stealing. However, older children may have conflicting intuitions about who deserves more punishment— the person who directly praises the immoral action of stealing or the hypocritical condemner—younger children did not have these conflicting intuitions, perhaps because they do not appreciate the hypocrisy of condemning an action and then engaging in it anyway. This result is partially in line with how younger children have responded in our previous studies—they tend to think that the person who they predicted would steal should be punished more harshly. Younger children's judgments of punishment in these situations are interesting, but further investigation of these judgments is beyond the scope of the current article and should be followed up in future research. Broadly, the results of our studies suggest that younger children are able to make predictions based on people's preferences, but that they have specific difficulties with making predictions about future behavior based on condemnation.

General Discussion

These studies demonstrate that by age 7, children begin to treat moral condemnation as a signal of one's future moral behavior. In four studies, we found that 7- to 9-year-old children ("older children") predicted that a condemner of stealing would be less likely to steal than a noncondemner. Older children predicted that a condemner of stealing would be less likely to steal than someone who said broccoli was gross (Study 1). They also predicted that a condemner of stealing would be less likely to steal than someone who made a moral statement praising a prosocial action ("sharing is really, really good," Study 2a) and also someone who explicitly denied ever stealing ("I never steal," Study 3). This is not because older children always think that condemners are more moral than others; they predicted that someone who praises sharing is more likely to share than someone who condemns stealing, suggesting that they do not think that someone who condemns stealing is broadly more moral (2b). These data suggest that older children appear to treat moral condemnation of an action as a signal that one will not do that action. Furthermore, we found that this signal has a cost attached to it: although older children think that a condemner is unlikely to steal, conditional on being caught stealing they think a condemner should be punished more harshly than someone who never made a claim about stealing (Studies 1 and 2a).

Seven- to 9-year-old children in our studies met both criteria of treating condemnation as a costly signal. However, 4- to 6-year-old children did not appear to treat moral condemnation of stealing as a signal that one would not steal. Indeed, they instead were somewhat suspicious of someone for mentioning stealing, and tended to predict that the condemner was more likely to steal (Studies 1 and 2). However, when both characters mentioned the word stealing, children no longer differentiated between the two characters, which suggest that younger children were merely making their selections based on which character was associated with the word stealing (Study 3). Importantly, younger children were not unable to make predictions about the characters' behaviors based on the characters' preferences; they clearly knew that someone who liked and praised stealing was more likely to steal than someone who condemned it (Study 4). Taken together, these studies suggest that older children treat moral condemnation as a signal, whereas younger children do not seem to do so.

Our results join a growing body of research suggesting that by the time children are 7–8 years old, they can make nuanced judgments about others' social signaling behaviors (for review, see Banerjee, 2002) and become increasingly adept at thinking about deceptive reputation management tactics and other forms of self-presentation (Banerjee & Yuill, 1999; Bennett & Matthews, 2000; Heyman et al.,

2014; Silver & Shaw, 2018). Our results demonstrate that, by age 7, children treat moral condemnation as an honest signal: they thought those who condemned an immoral behavior were less likely to immorally engage in that behavior, and that hypocritical condemners should be punished more harshly.

The consistency in the developmental emergence of understanding these types of strategic behaviors suggests that there may be some underlying social cognitive skills that develop or improve around age 7 which allow children to think about moral condemnation as a tactic for improving one's own reputation, and to condemn moral hypocrisy. Research suggests that second-order reasoning about mental states may be related to self-presentation (Banerjee & Yuill, 1999), and is necessary but not sufficient for an understanding of self-conscious concerns (Bennett & Matthews, 2000). The ability to attribute second-order mental states improves around age 6 or 7 (Perner & Wimmer, 1985), and this ability may be important for understanding reputational strategies and, thus, hypocrisy. By age 7, children may understand that an individual who condemns stealing may be manipulating others' mental impressions of herself. Indeed, with age, children show increased skepticism when self-interested motives are likely to be present (e.g., Heyman & Legare, 2005; Heyman et al., 2014; Mills & Keil, 2005; Reyes-Jaquez & Echols, 2015). Furthermore, around this age children are increasingly engaging in reputation management themselves and this may afford children with more and more learning opportunities in which there is discord between their peers' public image and what they do privately (Silver & Shaw, 2018). These learning experiences in social interaction likely play an important role in children's developing understanding of reputation management.

However, it is unclear from our results what, if any, information younger children take away from condemnation. In Studies 1 and 2 younger children, if anything, seemed to expect the condemner to be more likely to steal, which may have been because the character was associated with the word stealing. Indeed, this possibility was explored in Study 3: when both characters mentioned negative attitudes about stealing, younger children did not make a clear prediction about who was more likely to transgress. So, were younger children simply unable to follow our scenarios? If younger children were merely confused by our method, then they should not have been able to make concrete predictions about behavior based on a character's positive

evaluation of a behavior, but in Study 2b even younger children could predict that someone who praised sharing would be more likely to share. Furthermore, in Study 4, younger children were able to predict that someone who expresses a positive attitude about a stealing ("I really, really like stealing") would be more likely to steal than someone who does not. This is consistent with previous work suggesting that even very young children can infer others' preferences (Kushnir, Xu, & Wellman, 2010) and use such preferences to predict others' behavior (Liu, Gelman, & Wellman, 2007; Meltzoff, 1995; Perner & Wimmer, 1985).

Why else then would younger children specifically struggle with making predictions based on negative attitudes about immoral behavior? One possibility is that younger children may have trouble reasoning about people expressing a dispreference—they inferred that someone who expressed a positive attitude about an action would engage in that action, but they did not predict that someone who expressed a negative attitude about an action would avoid that action. In line with this possibility, there is work demonstrating that young toddlers struggle more with inferring dispreference than preference (Feiman, Carey, & Cushman, 2015). However, we believe that this is fairly unlikely for two reasons: first, given that the children tested here were much older than toddlers (i.e., 4- to 6-year-olds) and that, in general, even infants and young children are attentive to others' negative affect and use it to guide their own behavior (for review, see Vaish, Grossmann, & Woodward, 2008), we think it is highly unlikely that children are generally unable to make predictions based on dispreference.

A second possibility is that younger children have a specific difficulty making predictions about moral behaviors based on condemnation. Some research suggests that it is around age 6 that children begin to show an understanding of conflicting mental states. For example, 3- to 5-year-old children seem to believe that it is better to be morally consistent than to be conflicted (Starmans & Bloom, 2016) and until around age 7, children seem to have difficulty understanding the effect of conflicting mental states on behavior (Choe, Keil, & Bloom, 2005). It could be that older children are not only better able to appreciate the motivation to manage one's reputation, but also better able to recognize that condemnation itself is a moral commitment to future actions and mental states. Indeed, some research by Hussar and Harris (2010) suggests that 6- to 10-year-old children negatively judge meat eating

in vegetarians with moral commitments, but not in nonvegetarians who make no such commitments, and other work by Isella, Kanngiesser, and Tomasello (2018) suggests that it is not until around six that children can infer the commitment to future actions of speech acts like promises. Indeed, younger children may have difficulty in understanding that condemnation should be a particularly good signal of consistency between current mental states (a negative attitude toward stealing) and future actions (not stealing).

If children are having specific difficulties with making moral predictions based on dispreference, one could test this idea by using a design similar to Study 1, asking children to make a prediction about who would be more likely to eat broccoli: a child who condemned eating broccoli or one who condemned stealing. If children were able to succeed on this task (predicting that the person who condemned eating broccoli would not eat broccoli), it would suggest that children's difficulties are specifically tied to making predictions about future moral behaviors (not just any behavior), or that they simply think that someone who mentions stealing is more likely to do it. If they failed at this task, then it might suggest that younger children have difficulty making predictions based on condemnation per se. Perhaps younger children find it easier to make predictions about future behavior based on positive attitudes or preferences because they are aware by this point that people's subjective preferences are predictive of their behavior, but may see condemnation as a statement of a rule (e.g., stealing is bad) that anyone could express but that has no predictive force because they assume that everyone understands that stealing is wrong. Indeed, previous work has suggested that children until about age 6 or 7 sometimes struggle to make predictions about immoral actions or understand that others will commit immoral actions at all—thinking that certain immoral actions are almost "impossible" (Shtulman & Phillips, 2018). Future research will be needed to investigate how children's intuitions about the (im)moral behavior of their peers develop alongside their intuitions about reputation and self-presentation (for review, see Silver & Shaw, 2018).

One important question for future research concerns the specificity of condemnation as a costly signal for the older children. Exactly what sort of information do children glean from condemnation? For example, would children assume that someone who condemns stealing is more likely to refrain from other moral transgressions, or that she is more likely to act prosocially than others? That is, does

condemnation of stealing merely convey that the person does not steal, or does it convey that the person is unlikely to engage in related or unrelated immoral behaviors more broadly? Our Study 2b began to answer this question, at least suggesting that condemning stealing is a stronger signal that the condemner will refrain from stealing than it is a signal of other moral actions such as sharing. This result importantly indicates that there is some specificity to the information that children extract; however, this does not mean that children do not extract some general signal of moral goodness from condemnation. For example, if we asked children to predict whether a condemner or a neutral noncondemner of stealing was more likely to share, they may very well pick the condemner. We know that adults form positive impressions of people who condemn or punish immoral behavior (Barclay, 2006; Fessler & Haley, 2003; Jordan et al., 2017) and so it seems likely that children may as well.

A related question concerns the strength of condemnation as a signal. The current studies are limited in that the condemnatory behavior of the characters was the only information participants could draw from in order to predict the characters' behavior. Future work could directly address this question by juxtaposing condemnation information with other relevant social information. By providing children with more information about characters (e.g., their prior behavior, their group status, how others evaluate them, or how they spend their time), future studies could illuminate in which contexts children prioritize a person's condemnation in predicting her subsequent behavior over other relevant pieces of social information. Indeed, our results reveal that children think that others should be punished for being hypocritical, but the extent to which they recognize hypocrisy may be heavily influenced by their personal affiliations. For example, one may be motivated to excuse the hypocrisy of a friend or ingroup member and be particularly harsh about similar hypocrisy from an outgroup member. There are many open questions in this area to pursue.

Finally, we acknowledge a few limitations of the current work. First, these studies all used a forced-choice paradigm. We used a forced-choice paradigm to simplify the decision set and because forced-choice questions do approximate many important decisions that children must make in their daily lives—they must often make choices about whom to trust, affiliate with, or exclude. However, forced-choice paradigms can sometimes obscure meaningful variance, which may lead one to odd conclusions. For example, in our studies

children judged that a condemner should be punished more harshly than a noncondemner for stealing, but obviously this does not mean they think the noncondemner should not be punished. If children had been asked to rate on a scale how much each of them should be punished, they would likely think both characters should be punished. Moving forward it may be valuable to include different measures, like a trust game or other evaluations measures, to further probe children's moral intuitions about each character.

Second, these studies were conducted at a children's museum in the United States and so it is unclear if these results would generalize more broadly as most of the children tested in this paradigm were from WEIRD cultures (Henrich, Heine, & Norenzayan, 2010). Obviously, culture will dramatically influence the acts that children and adults think are morally good, and the acts they will condemn (Haidt, 2012; Rai & Fiske, 2011; Rochat et al., 2009). It seems possible that culture may also influence the extent to which condemnation is treated as signal. Perhaps in other societies, children would not make inferences about behavior based on condemnation or punish others for being hypocritical. However, because condemnation is ubiquitous in a variety of societies, from modern urban environments to small-scale hunter gatherer societies (Boehm, 1993; Guala, 2012; Wainryb & Recchia, 2013), it seems quite possible that children in other cultures would also treat condemnation as a signal. Future research should explore this possibility.

Condemnation is incredibly prevalent in the social world and we have long known that condemnation begins early in ontogeny. The current studies reveal that by age 7, children infer important information from others' condemnation: they make predictions about others' future moral behavior, supporting the notion that children utilize social information to generate predictions as they navigate the social world. They also understand that it is bad to condemn others for stealing if one is a thief, indicating a truth that is commonly understood in adults: no one likes a hypocrite.

References

Alicke, M., Gordon, E., & Rose, D. (2013). Hypocrisy: What counts? *Philosophical Psychology*, *26*, 673–701. https://doi.org/10.1080/09515089.2012.677397

Aloise-Young, P. A. (1993). The development of self-presentation: Self-promotion in 6- to 10-year-old children. *Social Cognition*, *11*, 201–222. https://doi.org/10.1521/soco.1993.11.2.201

Apfelbaum, E. P., Pauker, K., Ambady, N., Sommers, S. R., & Norton, M. I. (2008). Learning (not) to talk about race: When older children underperform in social categorization. *Developmental Psychology*, *44*, 1513–1518. https://doi.org/10.1037/a0012835

Banerjee, R. (2002). Children's understanding of self-presentational behavior: Links with mental-state reasoning and the attribution of embarrassment. *Merrill-Palmer Quarterly: Journal of Developmental Psychology*, *48*, 378–404. https://doi.org/10.1353/mpq.2002.0015

Banerjee, R., & Yuill, N. (1999). Children's understanding of self-presentational display rules: Associations with mental-state understanding. *British Journal of Developmental Psychology*, *17*, 111–124. https://doi.org/10.1348/026151099165186

Barclay, P. (2006). Reputational benefits for altruistic punishment. *Evolution and Human Behavior*, *27*, 325–344. https://doi.org/10.1016/j.evolhumbehav.2006.01.003

Barden, J., Rucker, D. D., & Petty, R. E. (2005). "Saying one thing and doing another:" Examining the impact of event order on hypocrisy judgments of others. *Personality and Social Psychology Bulletin*, *31*, 1463–1474. https://doi.org/10.1177/0146167205276430

Bennett, M., & Matthews, L. (2000). The role of second-order belief-understanding and social context in children's self-attribution of social emotions. *Social Development*, *9*, 126–130. https://doi.org/10.1111/1467-9507.t01-1-00115

Bird, B. R., & Smith, E. A. (2005). Signaling theory, strategic interaction, and symbolic capital. *Current Anthropology*, *46*, 221–248. https://doi.org/10.1086/427115

Bloom, P. (2013). *Just babies: The origins of good and evil*. New York, NY: Broadway Books.

Boehm, C. (1993). Egalitarian society and reverse dominance hierarchy. *Current Anthropology*, *34*, 227–254. https://doi.org/10.1086/204166

Bregant, J., Shaw, A., & Kinzler, K. D. (2016). Intuitive jurisprudence: Early reasoning about the functions of punishment. *Journal of Empirical Legal Studies*, *13*, 693–717. https://doi.org/10.1111/jels.12130

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

Bussey, K. (1999). Children's categorization and evaluation of different types of lies and truths. *Child Development*, *70*, 1338–1347. https://doi.org/10.1111/1467-8624.00098

Choe, K. S., Keil, F. C., & Bloom, P. (2005). Children's understanding of the Ulysses conflict. *Developmental Science*, *8*, 387–392. https://doi.org/10.1111/j.1467-7687.2005.00426.x

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

Darley, J. M., & Shultz, T. R. (1990). Moral rules: Their content and acquisition. *Annual Review of Psychology*, *41*, 525–556. https://doi.org/10.1146/annurev.ps.41.020190.002521

DeScioli, P., & Kurzban, R. (2009). Mysteries of morality. *Cognition*, *112*, 281–299. https://doi.org/10.1016/j.cognition.2009.05.008

DeScioli, P., & Kurzban, R. (2013). A solution to the mysteries of morality. *Psychological Bulletin*, *139*, 477–496. https://doi.org/10.1037/a0029065

Dunham, Y., Baron, A. S., & Banaji, M. R. (2008). The development of implicit intergroup cognition. *Trends in Cognitive Sciences*, *12*, 248–253. https://doi.org/10.1016/j.tics.2008.04.006

Engelmann, J. M., Herrmann, E., & Tomasello, M. (2012). Five-year olds, but not chimpanzees, attempt to manage their reputations. *PLoS ONE*, *7*, e48433. https://doi.org/10.1371/journal.pone.0048433

Engelmann, J. M., Over, H., Herrmann, E., & Tomasello, M. (2013). Young children care more about their reputation with ingroup members and potential reciprocators. *Developmental Science*, *16*, 952–958. https://doi.org/10.1111/desc.12086

Feiman, R., Carey, S., & Cushman, F. (2015). Infants' representations of others' goals: Representing approach over avoidance. *Cognition*, *136*, 204–214. https://doi.org/10.1016/j.cognition.2014.10.007

Fessler, D. M. T., & Haley, K. J. (2003). The strategy of affect: Emotions in human cooperation. In P. Hammerstein (Ed.), *Genetic and cultural evolution of cooperation* (pp. 7–36). Cambridge, MA: MIT Press.

Fu, G., Heyman, G. D., Chen, G., Liu, P., & Lee, K. (2015). Children trust people who lie to benefit others. *Journal of Experimental Child Psychology*, *129*, 127–139. https://doi.org/10.1016/j.jecp.2014.09.006

Geraci, A., & Surian, L. (2011). The developmental roots of fairness: Infants' reactions to equal and unequal distributions of resources. *Developmental Science*, *14*, 1012–1020. https://doi.org/10.1111/j.1467-7687.2011.01048.x

Grocke, P., Rossano, F., & Tomasello, M. (2015). Procedural justice in children: Preschoolers accept unequal resource distributions if the procedure provides equal opportunities. *Journal of Experimental Child Psychology*, *140*, 197–210. https://doi.org/10.1016/j.jecp.2015.07.008

Guala, F. (2012). Reciprocity: Weak or strong? What punishment experiments do (and do not) demonstrate. *Behavioral and Brain Sciences*, *35*, 1–15. https://doi.org/10.1017/S0140525X11000069

Haidt, J. (2012). *The righteous mind: Why good people are divided by politics and religion*. New York, NY: Pantheon Books.

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

Hamlin, J. K., Wynn, K., Bloom, P., & Mahajan, N. (2011). How infants and toddlers react to antisocial others. *Proceedings of the National Academy of Sciences of the United States of America*, *108*, 19931–19936. https://doi.org/10.1073/pnas.1110306108

Helwig, C. C., & Jasiobedzka, U. (2001). The relation between law and morality: Children's reasoning about socially beneficial and unjust laws. *Child Development*, *72*, 1382–1393. https://doi.org/10.1111/1467-8624.00354

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, *33*, 61–83. https://doi.org/10.1017/S0140525X0999152X

Heyes, C. M., & Frith, C. D. (2014). The cultural evolution of mind reading. *Science*, *344*, 124–309. https://doi.org/10.1126/science.1243091

Heyman, G., Barner, D., Heumann, J., & Schenck, L. (2014). Children's sensitivity to ulterior motives when evaluating prosocial behavior. *Cognitive Science*, *38*, 683–700. https://doi.org/10.1111/cogs.12089

Heyman, G. D., & Legare, C. H. (2005). Children's evaluation of sources of information about traits. *Developmental Psychology*, *41*, 636–647. https://doi.org/10.1037/0012-1649.41.4.636

Hussar, K. M., & Harris, P. L. (2010). Children who choose not to eat meat: A study of early moral decision-making. *Social Development*, *19*, 627–641. https://doi.org/10.1111/j.1467-9507.2009.00547.x

Isella, M., Kanngiesser, P., & Tomasello, M. (2018). Children's selective trust in promises. *Child Development*. https://doi.org/10.1111/cdev.13105

Jordan, J. J., Sommers, R., Bloom, P., & Rand, D. G. (2017). Why do we hate hypocrites? Evidence for a theory of false signaling. *Psychological Science*, *28*, 356–368. https://doi.org/10.1177/0956797616685771

Kushnir, T., Xu, F., & Wellman, H. M. (2010). Young children use statistical sampling to infer the preferences of other people. *Psychological Science*, *21*, 1134–1140. https://doi.org/10.1177/0956797610376652

Lane, J. D., Wellman, H. M., & Gelman, S. A. (2013). Informants' traits weigh heavily in young children's trust in testimony and in their epistemic inferences. *Child Development*, *84*, 1253–1268. https://doi.org/10.1111/cdev.12029

Leimgruber, K. L., Shaw, A., Santos, L. R., & Olson, K. R. (2012). Young children are more generous when others are aware of their actions. *PLoS ONE*, *7*, e48292. https://doi.org/10.1371/journal.pone.0048292

Liu, D., Gelman, S. A., & Wellman, H. M. (2007). Components of young children's trait understanding: Behavior-to-trait inferences and trait-to-behavior predictions. *Child Development*, *78*, 1543–1558. https://doi.org/10.1111/j.1467-8624.2007.01082.x

Meltzoff, A. N. (1995). Understanding the intentions of others: Re-enactment of intended acts by 18-month-old children. *Developmental Psychology*, *31*, 838. https://doi.org/10.1037/0012-1649.31.5.838

Miller, S. A. (2009). Children's understanding of second-order mental states. *Psychological Bulletin*, *135*, 749. https://doi.org/10.1037/a0016854

Mills, C. M., & Keil, F. C. (2005). The development of cynicism. *Psychological Science*, *16*, 385–390. https://doi.org/10.1111/j.0956-7976.2005.01545.x

Perner, J., & Wimmer, H. (1985). "John *thinks* that Mary thinks that. . ." attribution of second-order beliefs by 5-to 10-year-old children. *Journal of Experimental Child Psychology*, *39*, 437–471. https://doi.org/10.1016/0022-0965(85)90051-7

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

Reyes-Jaquez, B., & Echols, C. H. (2015). Playing by the rules: Self-interest information influences children's trust and trustworthiness in the absence of feedback. *Cognition*, *134*, 140–154. https://doi.org/10.1016/j.cognition.2014.10.002

Rhodes, M., & Chalik, L. (2013). Social categories as markers of intrinsic interpersonal obligations. *Psychological Science*, *24*, 999–1006. https://doi.org/10.1177/0956797612466267

Rizzo, M. T., Cooley, S., Elenbaas, L., & Killen, M. (2018). Young children's inclusion decisions in moral and social–conventional group norm contexts. *Journal of Experimental Child Psychology*, *165*, 19–36. https://doi.org/10.1016/j.jecp.2017.05.006

Rochat, P., Dias, M. D., Liping, G., Broesch, T., Passos-Ferreira, C., Winning, A., & Berg, B. (2009). Fairness in distributive justice by 3-and 5-year-olds across seven cultures. *Journal of Cross-Cultural Psychology*, *40*, 416–442. https://doi.org/10.1177/0022022109332844

Schmidt, M. F., Rakoczy, H., & Tomasello, M. (2012). Young children enforce social norms selectively depending on the violator's group affiliation. *Cognition*, *124*, 325–333. https://doi.org/10.1016/j.cognition.2012.06.004

Schmidt, M. F., Svetlova, M., Johe, J., & Tomasello, M. (2016). Children's developing understanding of legitimate reasons for allocating resources unequally. *Cognitive Development*, *37*, 42–52. https://doi.org/10.1016/j.cogdev.2015.11.001

Shaw, A., Li, V., & Olson, K. R. (2013). Reputation is everything. In M. Banaji & S. Gelman (Eds.), *Navigating the social world* (pp. 220–224). New York, NY: Oxford University Press.

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*, 363–375. https://doi.org/10.1037/a0031247

Shaw, A., & Olson, K. (2014). Fairness as partiality aversion: The development of procedural justice. *Journal of Experimental Child Psychology*, *119*, 40–53. https://doi.org/10.1016/j.jecp.2013.10.007

Shaw, A., & Olson, K. (2015). Whose idea is it anyway? The importance of reputation in acknowledgement. *Developmental Science*, *18*, 502–509. https://doi.org/10.1111/desc.12234

Shtulman, A., & Phillips, J. (2018). Differentiating "could" from "should:" Developmental changes in modal cognition. *Journal of Experimental Child Psychology*, *165*, 161–182. https://doi.org/10.1016/j.jecp.2017.05.012

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* (pp. 1–83). Chicago, IL: University of Chicago Press.

Silver, I., & Shaw, A. (2018). Pint-sized public relations: The development of reputation management. *Trends in Cognitive Sciences*, *22*, 277–279. https://doi.org/10.1016/j.tics.2018.01.006

Smetana, J. G., & Braeges, J. L. (1990). The development of toddlers' moral and conventional judgments. *Merrill-Palmer Quarterly*, *36*, 329–346. Retrieved from https://www.jstor.org/stable/23087284

Starmans, C., & Bloom, P. (2016). When the spirit is willing, but the flesh is weak: Developmental differences in judgments about inner moral conflict. *Psychological Science*, *27*, 1498–1506. https://doi.org/10.1177/0956797616665813

Tasimi, A., Dominguez, A., & Wynn, K. (2015). Do-gooder derogation in children: the social costs of generosity. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.01036

Tasimi, A., & Wynn, K. (2016). Costly rejection of wrongdoers by infants and children. *Cognition*, *151*, 76–79. https://doi.org/10.1016/j.cognition.2016.03.004

Turiel, E. (1998). The development of morality. In W. Damon & N. Eisenberg (Eds.), *Handbook of child psychology* (pp. 95–130). New York, NY: Academic Press.

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

Vaish, A., Herrmann, E., Markmann, C., & Tomasello, M. (2016). Preschoolers value those who sanction non-cooperators. *Cognition*, *153*, 43–51.

Wainryb, C., & Recchia, H. (2013). Moral lives across cultures. In M. Killen & J. G. Smetana (Eds.), *Handbook of moral development* (pp. 259–278). New York, NY: Psychology Press.

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, *13*, 103–128. https://doi.org/10.1016/0010-0277(83)90004-5

Yuill, N., & Perner, J. (1988). Intentionality and knowledge in children's judgments of actor's responsibility and recipient's emotional reaction. *Developmental Psychology*, *24*, 358. https://doi.org/10.1037/0012-1649.24.3.358

Zahavi, A. (1975). Mate selection—a selection for a handicap. *Journal of Theoretical Biology*, *53*, 205–214. https://doi.org/10.1016/0022-5193(75)90111-3

Supporting Information

Additional supporting information may be found in the online version of this article at the publisher's website:

Appendix S1. Replication of Study 1 without a prediction measure preceeding the punishment measure.