



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



Contents lists available at [ScienceDirect](https://www.sciencedirect.com)

Journal of Experimental Child Psychology

journal homepage: www.elsevier.com/locate/jecp



Children's judgments of interventions against norm violations: COVID-19 as a naturalistic case study



Young-eun Lee^{a,*}, Julia Marshall^b, Paul Deutchman^b, Katherine McAuliffe^b, Felix Warneken^a

^a Department of Psychology, University of Michigan, Ann Arbor, MI 48109, USA

^b Department of Psychology and Neuroscience, Boston College, Chestnut Hill, MA 02467, USA

ARTICLE INFO

Article history:

Received 9 December 2021

Revised 11 April 2022

Available online 21 April 2022

Keywords:

Social norm

Punishment

Intervention

Development

ABSTRACT

The coronavirus pandemic has had a significant influence on social interactions, introducing novel social norms such as mask-wearing and social distancing to protect people's health. Because these norms and associated practices are completely novel, it is unknown how children assess what kinds of interventions are appropriate under what circumstances and what principles they draw on in their decisions. We investigated children's reasoning about interventions against individuals who failed to adhere to COVID-19 norms. In this pre-registered study ($N = 128$), 4- to 7-year-olds heard stories about a norm violator, that is, a person who refuses to wear a mask in class (COVID condition) or wear indoor shoes in class when his or her shoes are muddy (Muddy Shoes condition). Children evaluated four different interventions—giving a mask/indoor shoes (Giving), preventing the person from entering (Exclusion), throwing a paper ball at the person (Throwing), and not intervening (Doing Nothing)—in terms of their rightness, niceness, and effectiveness. We found that across measures children evaluated Giving most positively, whereas they viewed Throwing most negatively. Doing Nothing and Exclusion received mixed evaluations across measures, revealing nuanced judgments of these interventions in children. In most measures, there was no difference between the COVID and Muddy Shoes conditions, suggesting that children's evaluations are not specific to the novel COVID-19 context. Together, our results show that

* Corresponding author.

E-mail address: yl4898@columbia.edu (Y.-e. Lee).

children dynamically evaluate each intervention, taking multiple factors into account. The current study has implications for the development of interventions against norm violations.

© 2022 Elsevier Inc. All rights reserved.

Introduction

The COVID-19 pandemic has had a major impact on our daily lives. One major change has been the emergence of novel social norms surrounding public health measures such as mask-wearing and social distancing (Fischer et al., 2021; Katz, Sanger-Katz, & Quealy, 2020; Pew Research Center, 2020). These new norms have led to public discussion over the right way to apply these norms and how to interact with people who do not follow them. Although common in everyday life and political debate, there is no systematic study about the way in which children conceptualize these novel COVID-19-related norms and how they evaluate different forms of norm enforcement. The question of norm enforcement is of particular importance because empirical research identifies the study of norm enforcement as central for the scientific understanding of the role of norms in social life (Fehr & Fischbacher, 2004; Gelfand, Harrington, & Jackson, 2017; Smetana, 2013). Specifically, the study of norm enforcement helps to assess how children conceptualize moral violations. For instance, psychological research has shown that children readily identify moral violations (e.g., hitting another person) as more deserving of punishment than conventional violations (e.g., not sitting in the designated place) (Smetana, 1981), suggesting that norm enforcement is one of the important criteria determining whether a violation is moral or socioconventional. Therefore, we use measures of norm enforcement as a means to gain a better understanding of how children make sense of these novel norms surrounding COVID-19.

We focused our investigation on children because the recently introduced COVID-19 norms provide developmental researchers with an unprecedented historical opportunity to study how completely novel norms and norm enforcement practices take hold in children's lives. Prior work on children's concepts of norms and norm enforcement has studied traditional norm violations such as immoral acts of hindering another person, property violations, and refusal to share in a fair manner (e.g., Hamlin, Wynn, Bloom, & Mahajan, 2011; McAuliffe, Jordan, & Warneken, 2015; Vaish, Herrmann, Markmann, & Tomasello, 2016). Therefore, most work has focused on "classic" problems of human social life with long-standing traditions in what norms apply, for example, distributive justice principles or moral and legal rules about violence. Thus, compared with the COVID-19 norms, children may be more familiar with classic sociomoral problems and the corresponding norms are deeply embedded in the socialization of children, with a rich set of principles to draw on and a wide range of examples, moral parables, stories, songs, and educational practices that are specifically geared toward the moral socialization of children. Although it is possible that children went through extensive socialization processes regarding COVID-19 norms during the past few years (e.g., by talking to their parents or peers), no such long-standing historical traditions exist for COVID-19, leaving open how children make sense of these novel norms.

One possibility is that children would have difficulties in understanding the novel COVID-19 norms based on their existing sociomoral reasoning skills, potentially due to an opaque causal relationship between a harmful action and its consequence inherent in the COVID-19 norm violation, which is different in nature from traditional moral problems. For example, at least at a superficial level, whether to put on a mask may seem like a harmless personal choice (like what clothes to wear). However, compared with other traditional moral transgressions, the actual consequences of not adhering to COVID-19 norms are potentially far more devastating (e.g., death). Furthermore, unlike traditional transgressions, there is a delay between the norm violation and its harm (e.g., a victim could show coronavirus-related symptoms several days after being exposed to people not wearing a mask), potentially making the norm violation even more opaque. Therefore, it is possible that the delayed moral

consequences in combination with seemingly harmless norm-violating actions might keep children from conceptualizing COVID-19 norms as a social norm whose violation requires an intervention.

Another possibility is that with proper guidance from parents, teachers, and/or media, children would understand these novel norms without difficulty, potentially because they apply their existing sociomoral reasoning skills to a novel norm violation context. Past research supports this possibility. For example, a recent study (Leotti, Pochinki, Reis, Bonawitz, & LoBue, 2021) found that 4- to 7-year-old U.S. children tested after the COVID-19 pandemic demonstrated greater understanding of a contagious illness such as the common cold than children tested before the pandemic, suggesting that the pandemic familiarized children with the concept of the transmission of a contagious illness and its underlying mechanism. Moreover, 5- to 12-year-olds in the United States understand the causal relation between not wearing a mask and making another person sick (Labotka & Gelman, 2022). Specifically, when asked the reason why wearing a mask would protect people from getting coronavirus, a majority of children provided explanations in terms of germ movement, showing that children understand that a facial mask is important for blocking coronavirus. Hence, based on children's understanding of a contagious illness, it is possible that children would conceptualize wearing a mask as a new social norm with ease. COVID-19 norm violations provide us with a unique opportunity to address how children reason about norm enforcement of novel norms with less transparent moral consequences in real life. Therefore, one major goal of this study was to provide the first systematic study of how children conceptualize and evaluate the norm enforcement of novel COVID-19 health measures.

A second major goal of our study was to conduct a more comprehensive investigation of children's reasoning about a whole range of different types of norm enforcement. Examining different interventions offers a more inclusive approach to studying what adults and children consider to be the most appropriate and effective forms of third-party intervention. As described in detail by previous studies with children and adults (Dhaliwal, Patil, & Cushman, 2021; Lee & Warneken, 2020; Raihani & Bshary, 2015), the vast majority of studies on norm enforcement have focused heavily on third-party punishment while paying little attention to other types of intervention. More recent studies have looked at alternative interventions other than punishment, primarily focusing on comparing compensation of victims with punishment of transgressors. Overall, these studies found that adults and children prefer to compensate a victim rather than punish a transgressor (Chavez & Bicchieri, 2013; Jordan, Hoffman, Bloom, & Rand, 2016; Riedl, Jensen, Call, & Tomasello, 2015; Yang, Wu, & Dunham, 2021; but see FeldmanHall, Sokol-Hessner, Van Bavel, & Phelps, 2014; McAuliffe & Dunham, 2021; and Stallen et al., 2018, for different results). Similarly, both adults and children evaluate those who compensated a victim more positively than those who punished a transgressor (Dhaliwal et al., 2021; Lee & Warneken, 2020; Liu, Yang, & Wu, 2021; Raihani & Bshary, 2015). Specifically, from around 5 years of age, children prefer a compensator over a punisher (Lee & Warneken, 2020). Together, these recent findings highlight the importance of alternative interventions, which might be more prevalent than punishment in real life.

Motivated by insights from this previous work, the current study included a broad range of possible reactions to a COVID-19 norm violation that could be observed in real life: (a) not letting a norm violator enter a space, (b) throwing a paper ball at a norm violator, (c) giving a new mask to a norm violator, and (d) doing nothing. To elaborate, we chose a sanction (i.e., not letting a norm violator enter a space) because this type of intervention has been one of the most common interventions in real life (e.g., preventing people not wearing a mask from entering a public space). We also added an unjustified aggressive reaction (i.e., throwing a paper ball at a norm violator) to assess how children reason about an aggressive behavior that is irrelevant to improving the situation in the context of COVID-19. Furthermore, we chose giving a new mask to a norm violator as a helpful intervention. To be specific, unlike previous research that identified compensation of a victim as a feasible alternative intervention in a fairness or ownership violation context (e.g., Lee & Warneken, 2020; Liu et al., 2021), we identified that giving a new mask to a person not wearing a mask is a more relevant helpful reaction in the COVID-19 context than compensating victims (i.e., giving a new mask to a potential victim who is already wearing one). Finally, we included doing nothing (no intervention) to examine whether children think that an intervention is even needed in the context.

A third major goal of this study was to provide a comprehensive assessment of children's evaluations of interventions beyond mere evaluations of niceness or rightness. Existing work rarely measured children's perceived effectiveness of interventions to change people's behavior, focusing instead on a unidimensional measure that assessed children's liking or endorsement of interventions (e.g., Lee & Warneken, 2020; Liu et al., 2021). Importantly, how effective an intervention is in reforming a transgressor's behavior and in removing harm of a transgression (e.g., whether it effectively undoes the transgression) might play a key role in evaluating an intervention. However, children's understanding of these aspects of an intervention cannot be captured by measuring only its rightness or niceness judgments.

We are aware of a few developmental studies assessing children's beliefs about the effectiveness of punishment. These studies show that 6- to 8-year-olds assume that an individual who acted in a mean way would become nicer after receiving punishment such as incarceration or time out (Dunlea & Heiphetz, 2021) and that punishment would prevent a future transgression (Bregant, Shaw, & Kinzler, 2016). In addition, children who are 6 years of age and older consider spanking to be the most effective way to change a person's misbehavior in the short term, although they rated it as least fair compared with other disciplinary methods such as time out, withdrawal of privileges, and reasoning (Vittrup & Holden, 2010). Moreover, children aged 5 to 7 years indicate that transgressors are less likely to reoffend after having been punished in a communicative way (Marshall, Yudkin, & Crockett, 2021), demonstrating that children understand punishment's potential to deter future misbehavior.

In sum, these studies suggest that children aged 5 years and older understand how punishment can shape a transgressor's future behavior. Importantly, it is around the same age that children begin to adjust their own behavior when there is punishment. For example, 6- and 7-year-olds show an increase in fair sharing when a third-party punishes unfair sharing (Martin, Martin, & McAuliffe, 2021). Together, by going beyond rightness or niceness measurements, these studies provided important insight into children's understanding of the effectiveness of punishment in reforming transgressors. However, most of these prior studies included only punishment and no other type of intervention. Moreover, they assessed whether the intervention changes the perpetrator's behavior or moral traits, but not whether the intervention is an effective solution to removing potential harm for other members of the group. Therefore, measuring children's understanding of effectiveness in removing harm for the group as well as effectiveness in reforming transgressors will allow us to gain a more nuanced understanding of children's evaluations while assessing potential dissociation in children's evaluations (e.g., judging an intervention as nice but ineffective).

The current study

We examined how 4- to 7-year-old children evaluate different reactions to those who refuse to follow COVID-19 norms. We identified 4 to 7 years as an age range that reflects important milestones in children's emerging norm development. Specifically, at this age range children readily distinguish moral norm violations from conventional ones (Smetana, 1981; Smetana, Jambon, Conry-Murray, & Sturge-Apple, 2012), evaluate punishment as adequate when norms are being violated (Lee & Warneken, 2020; Vaish et al., 2016), and start to engage in punishment themselves (Lee & Warneken, 2022; Marshall et al., 2021; McAuliffe et al., 2015; Yudkin, Van Bavel, & Rhodes, 2020). Moreover, this is around the same time that children begin to show an understanding of how punishment might change transgressors' behavior (Bregant et al., 2016; Dunlea & Heiphetz, 2021; Marshall et al., 2021; Martin et al., 2021). We did not test children aged 8 years and older because prior work suggested that children's evaluations of punishment remain relatively similar beyond 7 years of age (Bregant et al., 2016; Lee & Warneken, 2020).

We examined only the COVID-19 norm of mask-wearing because it is well known and provides a clear visual cue to children of whether a given protagonist followed a norm or not. Children heard scenarios detailing four different reactions to a person not wearing a mask in class. These four reactions included giving a new mask to the person (hereafter Giving), not letting the person in the classroom (hereafter Exclusion), throwing a paper ball at the person (hereafter Throwing), and not intervening in

the situation (hereafter Doing Nothing). By including four reactions to the norm violation, we sought to understand children's reasoning about different interventions that could be prevalent in real life.

Another important aim was to provide a multidimensional understanding of interventions in four different aspects. Specifically, we asked children to rate not only (a) the rightness of each intervention but also (b) its niceness, (c) its effectiveness for the group, and (d) its effectiveness for making the norm violator follow the norm.

Furthermore, because the current study involves a novel norm violation in an unusual setting (i.e., pandemic), we examined the degree to which children's responses to COVID-19 norm violations are generalizable to another context. To assess this, we compared COVID-19 norm violations with non-health-related norm violations. We created a scenario in which a character refused to wear indoor shoes in the classroom while wearing muddy shoes that could make the floor dirty. In the Muddy Shoes condition, we matched the interventions to those in the COVID-19 scenario: Giving (new indoor shoes), Exclusion, Throwing, and Doing Nothing.

We chose the muddy shoes scenario because at the superficial level the COVID-19 and muddy shoes scenarios share some common aspects. Specifically, in both scenarios, the norm violator refuses to wear either a mask or indoor shoes in class. By not wearing a mask or indoor shoes, there is a harm done to other people. That is, the coronavirus could spread between people and make people sick in the COVID-19 scenario, whereas the mud could spread to the floor and make other people's shoes dirty and make the floor slippery in the Muddy Shoes scenario. We believe that neither scenario involves completely moral or completely socioconventional violations. Rather, the two scenarios share some similarities with moral violations (e.g., the violation could potentially harm others' health or property) and share some aspects with socioconventional violations (e.g., a rule that may change depending on contexts) (Smetana et al., 2012). By including the Muddy Shoes condition, we could assess whether children's assessments of interventions are specific to the COVID-19 case or might apply to other contexts as well, selecting one intuitive contrast. Our goal was *not* to assess exhaustively whether children apply domain-general reasoning, which would require many more comparisons and the inclusion of norm violations far exceeding the scope of a single study.

We hypothesized that in their rightness judgments (i.e., how right or wrong an intervention is), children would evaluate Giving positively, whereas they would regard Doing Nothing and Throwing as negative, regardless of condition. This prediction was based on earlier findings that in fairness contexts children prefer a helpful intervention over a punitive one (Lee & Warneken, 2020). By contrast, we predicted that children would evaluate Exclusion more positively in the COVID condition than in the Muddy Shoes condition based on the assumption that children would be more accepting of a punitive intervention when the norm violation directly threatens group members' health than when it does not. We also expected that children would be more likely to endorse Exclusion with increasing age in the COVID condition because older children might better understand the opaque causal relationship between norm violation and its harm.

For niceness judgments (i.e., how nice or harsh an intervenor is to a norm violator), we expected children to make niceness judgments based on perceived aggression toward the norm violator. Hence, we hypothesized that regardless of condition, children would perceive non-aggressive interventions (i.e., Giving and Doing Nothing) as nice and would judge aggressive interventions (i.e., Exclusion and Throwing) as harsh. In addition to rightness judgments, we included a measure of niceness to capture a potential dissociation between the two measures. For example, what is considered as a nice behavior is not necessarily the right thing to do in a norm violation setting. Often a harsh action should be taken to stop the violation.

In judgments of effectiveness for groups, we expected children to infer whether an intervention effectively removes a potential threat (e.g., spread of coronavirus, getting a muddy floor) from other people in class. Therefore, the hypothesis was that across conditions children would rate Giving and Exclusion as effective solutions for the class because these interventions remove potential harm by either giving a new mask (or indoor shoes) or not allowing the person in the room, whereas children would rate Doing Nothing and Throwing as ineffective because these reactions could not get rid of potential harm from the group. Moreover, we predicted that this pattern of results would become pronounced with age because older children might better understand the opaque link between the norm violation and its harm.

Lastly, in judgments of effectiveness for changing norm violators, we expected children to make the judgments based on whether an intervention would be aversive to the norm violator. Our hypothesis was that across conditions children would consider aversive interventions (i.e., Exclusion and Throwing) to be effective in changing the transgressor's behavior, whereas they would consider non-aversive interventions (i.e., Giving and Doing Nothing) to be ineffective. The prediction was based on earlier findings that children aged 5–10 years consider aversive interventions such as time out, monetary punishment, spanking, incarceration to be effective in reforming a transgressor (Bregant et al., 2016; Dunlea & Heiphetz, 2021; Marshall et al., 2021; Martin et al., 2021; Vittrup & Holden, 2010).

Method

Participants

Our final sample consisted of 128 4- to 7-year-old children ($M = 71.27$ months, range = 48–95; $n = 32$ 4-year-olds, $n = 32$ 5-year-olds, $n = 33$ 6-year-olds, $n = 31$ 7-year-olds; 56 male and 72 female).¹ These children were tested from March 2021 to May 2021 when the Centers for Disease Control and Prevention required people to wear a mask indoors before coronavirus vaccinations became widely available. All children were living in the United States at the time of study. Based on zip codes reported by parents, a majority of the participants (63%) were from Michigan (see online supplementary material [SOM] for the distribution of participants across the United States). Most children were recruited from the University of Michigan Health Research database, in which parents sign up voluntarily to participate in research. Parents of the child participants received a \$5 Amazon e-gift card for their participation. Children were tested via an online meeting platform (Zoom) because in-person research was not possible. Demographic information such as race, education, and income were not obtained.

An additional 7 children were excluded because of parental interference ($n = 3$) or inattention ($n = 2$) or because they did not complete the test session ($n = 2$). Power analyses established that our sample size ($n = 64$ per condition) was large enough to detect a medium-sized condition effect ($d = .50$) with sufficient power (.80) at alpha = .05 (see our pre-registration at <https://aspredicted.org/7xf5h.pdf>).

Experimental design

Children were randomly assigned to the COVID condition or the Muddy Shoes condition (between participants; $n = 64$ per condition). In each condition, children heard a short story about a character who refused to wear a mask (COVID condition) or indoor shoes (Muddy Shoes condition). In both conditions, children heard four types of intervention against those who refused to wear a mask or indoor shoes: Giving, Exclusion, Doing Nothing, and Throwing. There were four measures for each intervention type: (a) rightness judgment of each intervention, (b) niceness judgment of each intervention, (c) its effectiveness for groups (i.e., whether each intervention is an effective solution for other people in class), and (d) its effectiveness for norm violators (i.e., whether the norm violator will wear a mask or indoor shoes in the future). Lastly, after seeing all four interventions, children were asked to rank each intervention from the best to the worst.

¹ Upon the completion of the current study, we asked children a few questions about their daily practice of mask-wearing to collect data for a separate study (Gollwitzer et al., 2022). Of 128 children, 61 were willing to answer these questions. When asked how often they wear a mask outside of the house, a vast majority of children reported that they always wear a mask outside of the house (51%) or wear it most of the time (25%). Those who reported sometimes (13%) or never (8%) wearing a mask were relatively rare. In addition, 36 children who were attending school completely or partially in-person at the time reported that at school they needed to wear a mask all the time (53%) or most of the time (33%). A minority of children reported that they needed to wear a mask some of the time (11%) or none of the time (3%) at school. Furthermore, children reported that their parents said they should always (51%) or most of the time (20%) wear a mask outside of the house. Thus, a vast majority of children practiced mask-wearing in their everyday life and also received guidance from their parents about mask-wearing in a public space.

Procedure

After parents provided consent online, researchers scheduled an online testing session in which children participated with a digital device (e.g., computer, tablet). At the beginning of the session, the experimenter double-checked that children could see our experimental stimuli from their device (see SOM).

During a *practice phase*, children learned how to indicate their judgments on three 4-point Likert scales that were going to be used during the test phase: rightness, niceness, and perceived likelihood scales. The practice trials confirmed that children could use each scale properly (see SOM).

During the *introduction phase*, children in the COVID condition heard a story about coronavirus with images of child characters at a summer camp. Participants were asked four questions to assess their understanding of coronavirus-related norms: (a) whether they think getting coronavirus is good or bad (98% reported that it is bad), (b) whether they think wearing a mask can keep people from getting coronavirus (97% reported that it can keep people from getting coronavirus), (c) whether they think the children at the camp should wear a facial mask in class (97% reported that they should wear a mask), and (d) whether it is okay or not okay to not wear a mask in class (97% reported that it is not okay). Those in the Muddy Shoes condition were asked identical questions about norms related to wearing muddy shoes in class: (a) whether children think getting muddy shoes is good or bad (94% reported that it is bad), (b) whether they think wearing indoor shoes can keep the floor from getting dirty (80% reported that it can keep the floor from getting dirty), (c) whether they think the children at the camp should wear indoor shoes in class (86% reported that they should wear indoor shoes), and (d) whether it is okay or not okay to not wear indoor shoes in class (94% reported that it is not okay). During this phase, the experimenter did not provide an explicit definition of coronavirus, masks, or indoor shoes because we were interested in assessing children's default understanding of the situations. This procedure confirmed that children overall endorsed wearing a mask or indoor shoes as a part of norms that people need to follow (see SOM for analysis examining the difference between conditions in these comprehension questions). In both conditions, the introduction phase ended with a scene in which a camp counselor told the story characters that everyone should wear a mask or indoor shoes in class. This scene was included to further strengthen the norm of wearing a mask or indoor shoes in the context.

During the following *test phase*, children were presented with four intervention scenarios in random order. All four scenarios started with a character (hereafter *norm violator*) who had a mask (COVID condition) or indoor shoes (Muddy Shoes condition) but intentionally refused to bring them. Then, the norm violator showed up in class without wearing a mask or indoor shoes.

A character who was introduced as a class leader responded to the norm violator in one of four ways (see Fig. 1). In a Giving scenario, the class leader gave a new mask (COVID condition) or a new pair of indoor shoes (Muddy Shoes condition) to the norm violator. In an Exclusion scenario, the class leader kept the norm violator from entering the classroom by closing the door. In a Throwing scenario, the class leader threw a paper ball at the norm violator's head. In a Doing Nothing scenario, the class leader did nothing to the norm violator.

Across four intervention scenarios, we introduced class leaders wearing a badge as an intervenor. Although a teacher might be most likely to intervene against norm violations in a classroom setting, we were concerned that introducing an authority figure as an intervenor would justify any actions that the person shows, resulting in children's positive evaluations irrespective of intervention types. Hence, we introduced a class leader as an intervenor because class leaders are peers in class and thus do not have the authority as strong as a teacher. At the same time, class leaders are often considered to have responsibilities to lead the class. This idea is consistent with a finding that 3- and 4-year-olds are more likely to punish their group members when an experimenter gave children a sheriff's badge and told them they were in charge compared with when there was no such manipulation (Yudkin et al., 2020).

After hearing each intervention scenario, children answered four test questions in a random order using a 4-point Likert scale: (a) rightness judgment of the class leader's behavior (from *really wrong* to *really right*), (b) niceness judgment of the class leader's behavior (from *really harsh* to *really nice*), (c) effectiveness of the intervention for the group (from *definitely no* to *definitely yes*), and (d) effectiveness

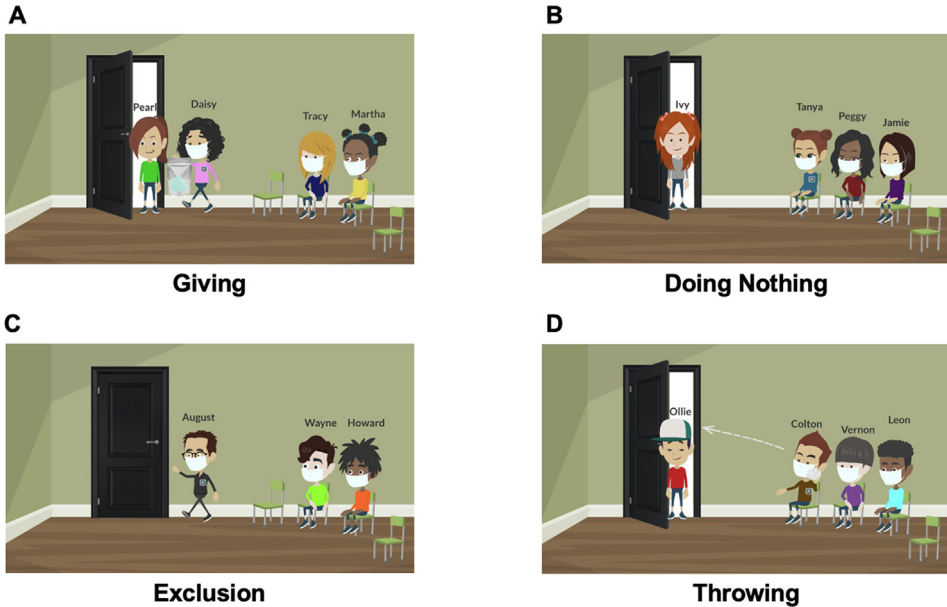


Fig. 1. Example of four intervention scenarios in the COVID condition. (A) The class leader gives a new mask to the person not wearing a mask. (B) The class leader does nothing to the person not wearing a mask. (C) The class leader closes the door and does not allow the person not wearing a mask to enter the room. (D) The class leader throws a paper ball at the person not wearing a mask. Children saw two intervention scenarios with a group of male characters and two other intervention scenarios with a group of female characters. We counterbalanced the intervention scenarios of these groups. Intervention scenarios in the Muddy Shoes condition were identical to those in the COVID condition except that (a) everyone but the norm violator was wearing indoor shoes and (b) a pair of new indoor shoes was given to the norm violator in a Giving scenario. © 2019 GoAnimate, Inc. Images are copyrighted by and used by permission of VYOND™. VYOND is a trademark of GoAnimate, Inc., registered in Australia, Brazil, the European Union, Norway, the Philippines, Singapore, Switzerland and the United Kingdom.

of the intervention on the norm violator (from *definitely no* to *definitely yes*). The test questions in the Muddy Shoes condition were identical except that the questions involved a norm violator who refused to wear indoor shoes in class (see Fig. 2).

At the end of the test session, children were reminded of the four intervention scenarios and were asked to rank class leaders who did the best job to the worst job when they saw someone not wearing a mask or indoor shoes (with 1 being the best and 4 being the worst).

In each intervention scenario, children were asked comprehension questions to confirm that they paid attention to the stories. Across four scenarios, children correctly identified the norm violator (99%) and the class leader who intervened (95%). In a minority of trials in which children did not identify them correctly, the experimenter provided a correct answer before asking test questions.

The order of intervention scenarios and all test questions was randomized by Qualtrics. All children saw two intervention scenarios with a group of male characters and two intervention scenarios with a group of female characters. Importantly, we counterbalanced the role of these characters.

Data coding and analyses

Children’s responses were automatically recorded by Qualtrics. All statistical analyses were conducted with R statistical software (R Version 4.1.1; R Core Team, 2021).

Before data collection, we pre-registered our research questions, measures, and analyses, but not specific hypotheses (see <https://aspredicted.org/7xf5h.pdf>). All data and protocols are available through the Open Science Framework (<https://osf.io/p5m7j>).

We analyzed four dependent measures (rightness, niceness, effectiveness for a group, and effectiveness on norm violator, respectively) as linear mixed models (LMMs) using the R package “lme4”

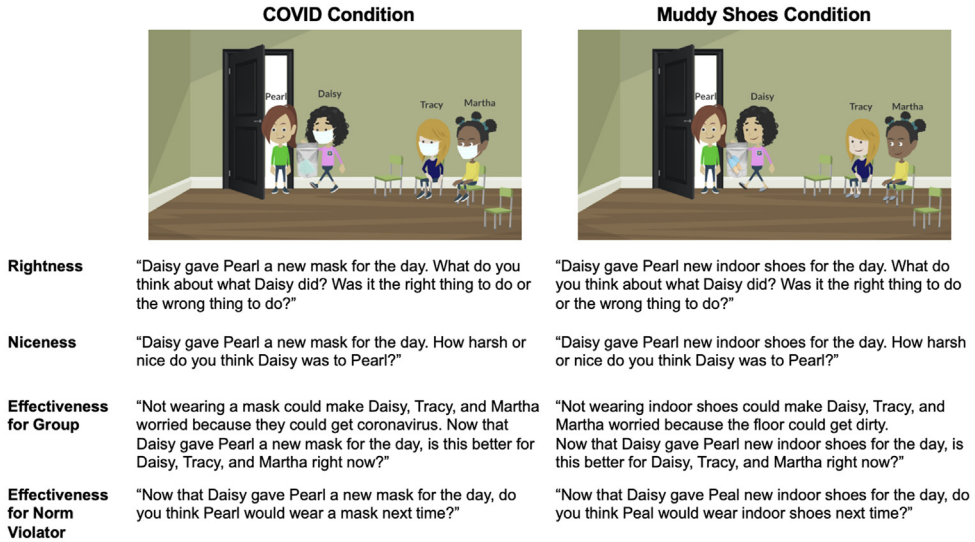


Fig. 2. Example of test questions in a Giving scenario across conditions. In this example, Daisy was a class leader who gave a new mask (COVID condition) or new indoor shoes (Muddy Shoes condition) to Pearl, who was not wearing a mask or indoor shoes. There were two other bystanders, Tracy and Martha. © 2019 GoAnimate, Inc. Images are copyrighted by and used by permission of VYOND™. VYOND is a trademark of GoAnimate, Inc., registered in Australia, Brazil, the European Union, Norway, the Philippines, Singapore, Switzerland and the United Kingdom.

(Bates, Maechler, Bolker, & Walker, 2015). All our full models included children’s age in months as a continuous variable, intervention type (Giving, Exclusion, Doing Nothing, or Throwing), condition (COVID or Muddy Shoes), and all interactions among the variables as fixed effects plus participant identity (ID) as a random effect to account for repeated testing. We first compared the full model with a null model that included only participant ID as a random intercept. If the full model provided a significantly better fit to the data than the null model, we then conducted hypothesis-driven tests to examine the role of individual predictors by sequentially dropping them from the full model and assessed changes in model fit using likelihood ratio tests (LRTs). Pairwise comparisons were performed using the R package “emmeans” (Lenth, 2021). We also conducted one-sample *t* tests (two-tailed) to examine whether children’s ratings differed significantly from a neutral valence of 0.²

To analyze the ranking of interventions, we ran a generalized estimating equation (GEE) for ordinal responses with the R package “multgee” (Touloumis, 2015).³ Our full models included children’s age in months, intervention type (Giving, Exclusion, Doing Nothing, or Throwing), condition (COVID or Muddy Shoes), and the interactions among the variables. We first compared the full model with a null model that included only an intercept. If the full model provided a significantly better fit to the data than the null model, we then conducted hypothesis-driven tests to examine the role of individual predictors by sequentially dropping them from the full model and assessed changes in model fit by performing a Wald-type goodness-of-fit test between two nested models.

In most of our measures, we found nonsignificant effects involving condition (COVID vs. Muddy Shoes). Thus, for the sake of brevity, we report the statistics involving condition only where it was sig-

² Because our pre-registered analyses included age as a critical variable, we did not include one-sample *t* tests in our pre-registration (given that they are meaningful only when collapsing across age for exploratory purposes). However, to provide more initial information on the valence of children’s evaluations as above or below the neutral value of zero, we included these analyses along with our descriptive statistics.

³ We had pre-registered to run an ordinal logistic regression on children’s ranking of interventions. However, due to an issue with model identifiability, we were unable to run this analysis. Thus, following the recommendation from two statistical consultants, we ran GEE models as an alternative. We then unpacked a significant interaction effect between age and intervention type by running an ordinal logistic regression for each intervention type with an R package “MASS” (Venables & Ripley, 2002).

nificant (see SOM for nonsignificant effects involving condition). In addition, all our full models were significantly different from the null models. For conciseness, we do not report each individual test in the main text but rather refer to the SOM for details.

Results

We first summarize analyses for each measure separately before we compare across measures to gain a fuller picture of children’s responses. For each of the measures, we first ran one-sample *t* tests that compared children’s ratings with a neutral valence of 0 to assess whether children’s rating significantly differed from a neutral rating. We next report analyses that examine potential effects of age, intervention type, and condition on children’s ratings and ordinal rankings, respectively.

Rightness judgments and ranking

As shown in Fig. 3A, children evaluated Giving overall positively, $t(127) = 22.88, p <.001$, whereas they evaluated Doing Nothing, Exclusion, and Throwing negatively, $t(127) = -2.48, p <.05, t(127) = -4.21, p <.001$, and $t(127) = -17.83, p <.001$, respectively.

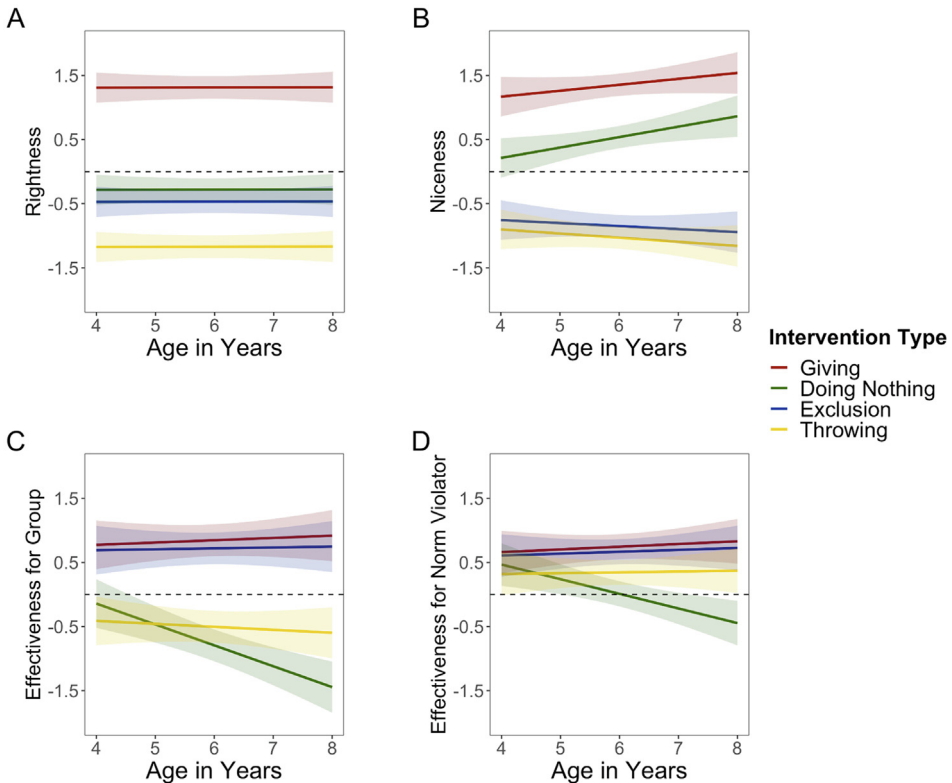


Fig. 3. Children’s estimated ratings across four measures. (A) Estimated ratings of rightness judgments based on the model including age. On the y axis, +1.5 indicates really right and -1.5 indicates really wrong. (B) Estimated ratings of niceness judgments based on the final model. On the y axis, +1.5 indicates really nice and -1.5 indicates really harsh. (C) Estimated ratings of effectiveness for groups based on the final model (collapsed across conditions). On the y axis, +1.5 indicates definitely yes and -1.5 indicates definitely no. (D) Estimated ratings of effectiveness for changing norm violators based on the final model. On the y axis, +1.5 indicates definitely yes and -1.5 indicates definitely no. In all graphs, the dashed line indicates a neutral valence of 0, and the confidence bands represent 95% confidence intervals.

When including age, intervention type, and condition as predictors, we found a significant main effect of intervention type, $LRT, \chi^2(3) = 229.02, p < .001$. Children rated Giving ($M = 1.31, SD = 0.65$) most positively, followed by Doing Nothing ($M = -0.28, SD = 1.29$) and Exclusion ($M = -0.47, SD = 1.26$). They rated Throwing most negatively ($M = -1.17, SD = 0.74$). Pairwise comparisons between intervention types were significant (all $ps < .001$) except for ratings of Exclusion versus Doing Nothing, $t(387) = 1.49, p > .44$.

The findings from children's ranking of four intervention types were largely similar to those from rightness ratings (with 1 being the best and 4 being the worst). That is, children ranked Giving highest ($M = 1.23, SD = 0.46$), followed by Doing Nothing ($M = 2.45, SD = 1.00$), Exclusion ($M = 2.98, SD = 0.88$), and Throwing ($M = 3.34, SD = 0.68$). In addition, we found a marginally significant interaction of intervention type and age, $\chi^2(3) = 7.76, p = .05$, such that with increasing age children tended to be more likely to rank Giving higher, $\chi^2(1) = 6.01, b = 0.04, SE = 0.02, p < .05$, and rank Doing Nothing lower, $\chi^2(1) = 4.52, b = -0.03, SE = 0.01, p < .05$.

In sum, these results suggest that children considered Giving as the most appropriate intervention against the norm violations and considered Throwing as the least appropriate reaction. Interestingly, although not letting someone enter a room without a mask is an effective and widely used intervention against the spread of COVID-19 in real life, children viewed it as similarly inappropriate as Doing Nothing.

Niceness judgments

As shown in Fig. 3B, whereas children evaluated Giving and Doing Nothing as nice, $t(127) = 29.01, p < .001$ and $t(127) = 4.91, p < .001$, respectively, they perceived Exclusion and Throwing as harsh; $t(127) = -10.30, p < .001$ and $t(127) = -14.53, p < .001$, respectively. Overall, children perceived Giving to be the nicest ($M = 1.34, SD = 0.52$), followed by Doing Nothing ($M = 0.52, SD = 1.21$), Exclusion ($M = -0.85, SD = 0.94$), and Throwing ($M = -1.03, SD = 0.80$).

Our regression analyses revealed an interaction of intervention type and age, $LRT, \chi^2(3) = 8.17, p < .05$. To unpack the interaction effect, we ran a separate linear model for each intervention and found that with age children were more likely to perceive Giving as nice ($b = 0.01, SE = 0.003$), $F(1, 125) = 5.69, p < .05$. There were no significant age effects for other interventions (all $ps > .07$). These results indicate that children viewed Giving and Doing Nothing as nice acts toward a norm violator, whereas they viewed Exclusion and Throwing as harsh acts.

Effectiveness for groups

We examined whether children thought the interventions were an effective solution for the group (e.g., "Not wearing a mask could make Daisy, Tracy, and Martha worried because they could get coronavirus. Now that Daisy gave Pearl a new mask for the day, is this better for Daisy, Tracy, and Martha right now?"; see Fig. 2). As displayed in Fig. 3C, results showed that whereas children evaluated Giving ($M = 0.97, SD = 0.92$) and Exclusion ($M = 0.55, SD = 1.06$) as effective for the group, $t(127) = 11.89, p < .001$ and $t(127) = 5.92, p < .001$, respectively, they perceived Doing Nothing ($M = -0.62, SD = 1.12$) and Throwing ($M = -0.55, SD = 1.06$) as ineffective, $t(127) = -6.24, p < .001$ and $t(127) = -5.82, p < .001$, respectively.

Regression analyses showed a significant interaction between intervention type and age, $LRT, \chi^2(3) = 16.51, p < .001$, and a significant interaction between intervention type and condition, $LRT, \chi^2(3) = 9.77, p < .05$. Subsequent separate linear models for each intervention showed that with age children became less likely to perceive Doing Nothing as an effective solution for the group ($b = -0.03, SE = 0.01$), $F(1, 125) = 16.93, p < .001$. Moreover, children showed a marginally significant tendency to perceive Exclusion as a more effective solution for the group in the COVID condition than in the Muddy Shoes condition ($b = -0.33, SE = 0.19$), $F(1, 125) = 3.11, p = .08$, whereas they showed a marginally significant tendency to perceive Doing Nothing as a less effective intervention in the COVID condition than in the Muddy Shoes condition ($b = 0.31, SE = 0.19$), $F(1, 125) = 2.81, p = .10$.

These findings show that children judged Giving and Exclusion to be effective for the group, whereas they judged Throwing to be ineffective. Moreover, we found an age effect in which children

were less likely to consider Doing Nothing as an effective solution for the group with age, suggesting that they gain a better understanding of the group-level consequences of onlooking over development.

Effectiveness for norm violators

We examined children's ratings to the question asking whether each intervention is an effective solution for changing the norm violator's behavior (e.g., "Now that Daisy gave Pearl a new mask for the day, do you think Pearl would wear a mask next time?"; see Fig. 2). As shown in Fig. 3D, children rated Giving ($M = 0.75$, $SD = 0.74$), Exclusion ($M = 0.67$, $SD = 0.88$), and Throwing ($M = 0.35$, $SD = 1.03$) as significantly more effective in making behavioral changes in norm violators than neutral, $t(127) = 11.43$, $t(127) = 8.64$, and $t(127) = 3.87$, respectively, all $ps < .001$.

Regression analyses indicated a significant interaction between intervention type and age, LRT , $\chi^2(3) = 13.51$, $p < .01$. Subsequent linear models by each intervention type revealed that with age children were less likely to perceive Doing Nothing as effective in making the norm violator wear a mask or indoor shoes next time ($b = -0.02$, $SE = 0.01$), $F(1, 125) = 7.73$, $p < .01$. There were no significant age effects in the other interventions, all $ps > .46$.

In sum, these findings suggest that as they grow older, children increasingly understand how onlooking would negatively influence the violator's future behavior. Interestingly, children rated not only aversive interventions (i.e., Throwing and Exclusion) but also Giving to be effective in making norm violators wear a mask next time, showing that children view the carrot (Giving) as well as the stick (Exclusion and Throwing) as effective in making violators follow the norms.

Discussion

The current study provides the first systematic assessment of children's reasoning about the appropriate way to intervene against norm violations in the context of COVID-19 health measures. Our results show that children as young as 4 to 7 years already have a sophisticated understanding of the sociomoral dilemmas that can result from enforcing the novel COVID-related health measures. Despite the fact that norms surrounding mask-wearing were introduced in the United States only in recent history, young children appear to apply their existing sociomoral reasoning abilities competently to this novel context.

Our findings reveal that children have a nuanced ability to reflect on trade-offs between different types of norm enforcement. Specifically, there were two interventions that consistently received either positive or negative evaluations. For example, Giving was viewed positively across all measures; children rated that Giving is the right behavior, a nice behavior, and an effective solution for the group and for reforming the norm violator. In contrast, Throwing was viewed negatively in three of the four measures; throwing was viewed as a wrong, harsh, and ineffective behavior for the group, although it was viewed as effective in making the violator follow the norm in the future. By contrast, there were two other interventions that received mixed evaluations from children. For example, even though Exclusion was perceived to be wrong and harsh, children considered it to be an effective solution for the group and for changing the norm violator's behavior. In addition, although Doing Nothing was viewed as a nice act, it was viewed as wrong and ineffective for the group and for changing the behavior of the norm violator. Taken together, our results suggest that children evaluate interventions in a sophisticated manner; they do not evaluate interventions uniformly positive or negative in all respects but rather are capable of nuanced reasoning about the intervention (e.g., a certain measure might be effective and nevertheless wrong and harsh).

Another noteworthy finding was that both Exclusion and Doing Nothing were rated as wrong. The results were surprising given the fact that Exclusion is one of the common ways of enforcing mask-wearing in real life. We speculate that children rated Exclusion negatively potentially because of perceived aggression or harshness. This possibility is consistent with our findings that children considered Exclusion as a harsh act in their niceness judgments. Moreover, our results highlight that children prefer helpful interventions over punitive ones when alternative interventions other than punishment are available. These results contrast with earlier findings that infants and children prefer

norm enforcers over non-enforcers when the only mode of intervention is punishment or verbal reproach of a transgressor (Hamlin et al., 2011; Vaish et al., 2016). Instead, the current results are largely in line with more recent findings on children's judgment in the context of fairness or ownership violations (Lee & Warneken, 2020; Liu et al., 2021) showing that children's preference for a helpful intervention over a punitive one extends to the context of health-related harm (although the prior studies involved a preference for a helpful intervention directed toward a victim, not toward a transgressor). Hence, the current study highlights the need to study children's judgments of interventions other than punishment in various moral contexts.

One of the unique features of the current study was the measures of children's perceived effectiveness of the interventions in terms of both preventing harm for other group members in the moment and changing the norm violator's behavior in the future. With these measures, we found that with age children were less likely to judge Doing Nothing as an effective solution for the group and for changing the norm violator's behavior. These results suggest that in the age range of 4 to 7 years, children understand potential negative consequences of not intervening in a norm violation. Our findings provide converging evidence to the existing literature that identified 4–7 years as an important period for the development of norm enforcement. To be specific, 4- and 5-year-olds start to enforce moral norms (Marshall et al., 2021; Yudkin et al., 2020) and evaluate those who enforce norms positively (Lee & Warneken, 2020; Vaish et al., 2016). At 5–8 years of age, children gain a better understanding of how punishment would influence a norm violator's future behavior (Bregant et al., 2016; Dunlea & Heiphetz, 2021; Marshall et al., 2021). Our findings add to the literature by showing that during this period children develop an understanding of how different interventions would affect other group members as well as a norm violator.

Another major finding was that even though the COVID-19 norms are novel, children quickly adapt to these novel norms. That is, children do not necessarily develop sociomoral reasoning skills specific to the novel COVID-19 norms. Rather, they seem to draw their sociomoral reasoning skills and apply them in a new context. This is apparent from the findings that children responded quite similarly to the COVID and the Muddy Shoes conditions across our measures. The Muddy Shoes condition was similar to the COVID condition in that the norm violation has negative consequences for the classmates, although not wearing indoor shoes might be perceived as more immediately harmful but less severe than the act of not wearing a mask. Despite potential differences in the severity or perceived harm of norm violations between the two conditions, in the vast majority of our measures children's responses did not differ significantly between the two conditions. There were only two instances with a condition difference (although both were marginally significant), namely with children judging Exclusion as a more effective solution for the group for the COVID condition than for the Muddy Shoes condition, whereas they tended to judge Doing Nothing as a less effective intervention in the COVID condition than in the Muddy Shoes condition. Overall, our results show that children's evaluations are not specific to the novel context of mask-wearing. In general, our results are more consistent with the notion that children's evaluations stem from a general reasoning process applied across contexts.

One potential limitation of our study is that the findings might not be generalizable to other populations. There are both between-nation and within-nation variations in the degree to which people follow the norm of mask-wearing (Fischer et al., 2021; Katz et al., 2020). Thus, parents' attitudes toward mask-wearing and their practice of it could be transferred to their children. In the current study, we addressed this issue by asking children in our U.S. sample whether they think people should wear a mask in class and whether it is okay or not okay to not wear a mask in class. Our study confirmed that, at least in our sample of children, a vast majority (97%) of them living in the United States reported that people should wear a mask and it is not okay to not wear a mask. Future research could investigate whether the current findings can be generalized to other samples (e.g., children from a U.S. state or another country with a low mask-wearing rate).

Another limitation is that some aspects of the current research might not be familiar to all children in the same way. For example, although our comprehension checks and measures confirmed that children did not have difficulty with understanding the scenarios, children might differ in their experience with wearing indoor shoes. Similarly, although children seem to experience some form of leadership role at schools such as student leaders, school captains, and head boy/girl (Billsberry, Escobar Vega, & Molineux, 2019; The Teacher Toolkit, n.d.), it is still possible that children in the current study might

have different experiences with a class leader. In addition, even though our practice trials confirmed that children did not have difficulty with evaluating the niceness or harshness of an intervention (see SOM), it is still plausible that children did not conceptualize harshness as an opposite end of niceness.

In addition, although the comparison between the COVID and Muddy Shoes conditions allowed us to examine whether children's evaluations are specific to the context of COVID-19 or not, the conditions were not entirely parallel. That is, the COVID scenario differed from the Muddy Shoes scenario in a few aspects. For example, children might perceive the COVID scenario as a more serious violation than the Muddy Shoes scenario. In addition, compared with the Muddy Shoes condition, emotional information of characters was less available in the COVID condition by covering up characters' mouths with a facial mask (see Fig. 2). However, despite the few differences between the two conditions, it is noteworthy that we did not find a significant difference between the conditions in children's evaluations in a majority of measures.

One remaining question is to explain the cognitive processes underlying children's reasoning about interventions. Because we did not ask children to justify their evaluations, our data could not explain the reason underlying children's judgments. Why did children evaluate Giving positively? Why did they rate Exclusion as negative as Doing Nothing in rightness judgments? Although the current study was not designed to address these questions directly, the result from an exploratory analysis shows that the more children perceived an intervention to be effective and nice, the more they viewed it as the right thing to do (see SOM). This finding suggests that children consider effectiveness for groups, effectiveness for norm violators, and niceness of an intervention simultaneously when deciding the degree to which the intervention is the right thing to do. Therefore, future studies could systematically vary these factors to further investigate the cognitive process underlying children's judgments and/or to investigate the reason underlying their judgments (e.g., by asking for their reasoning directly).

Another remaining question is whether children consider the violations in the current study as a moral or socioconventional violation. The current research did not directly assess whether children conceptualize these violations as a moral or conventional violation. It is well documented that children perceive moral transgressions to be more generally wrong, independent of rules and authority, more serious, and less alterable compared with conventional violations (Smetana et al., 2012). Hence, building on the prior work, future research could investigate how children's conceptualization of COVID and Muddy Shoes scenarios differs from classic moral transgressions such as stealing and hitting another person.

Future research should also examine how children evaluate other types of intervention. The four interventions explored in the current study are not the only ways to respond to a transgression. In real life, reproaching a transgressor verbally or reporting a transgressor to an authority figure could be an alternative way to intervene in the situation. It would be interesting to investigate how these interventions are evaluated compared with a direct punishment. Relatedly, future research could investigate how children intervene against a COVID-19 norm violation when they are given a chance to intervene and how children's actual behavior is related to their evaluations.

Lastly, it would be important to study how children were introduced to the novel COVID-19 norms. We conducted the current study about a year after the mask-wearing practice was introduced. We found that children readily accept this novel practice as a part of social norms that people should follow. However, the current study cannot address how children have become socialized to the novel norms. We speculate that children might have received numerous instructions and guidance from multiple sources in their everyday life such as their parents, teachers, peers, and media (Leotti et al., 2021). Future research should examine the process in which children are introduced and socialized to novel norms.

In conclusion, our study contributes to the existing literature by addressing how children judge four different types of intervention against a novel but naturalistic social norm. Our study showed that 4- to 7-year-olds reason about interventions dynamically, taking multiple factors into account simultaneously.

Acknowledgments

This research was funded in part by a Rackham Research Grant awarded to Y.L. We thank Sofia Lopez, Rachel Miller, Amy Nowack, and Allison Sparpana for their help with data collection. We also

thank the Consulting for Statistics, Computing and Analytics Research (CSCAR) at the University of Michigan for statistical advice.

References

- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>.
- Billsberry, J., Escobar Vega, C., & Molineux, J. (2019). Think of the children: Leader development at the edge of tomorrow. *Journal of Management and Organization*, 25, 378–381.
- 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.
- Chavez, A. K., & Bicchieri, C. (2013). Third-party sanctioning and compensation behavior: Findings from the ultimatum game. *Journal of Economic Psychology*, 39, 268–277.
- Dhaliwal, N. A., Patil, I., & Cushman, F. A. (2021). Reputational and cooperative benefits of third-party compensation. *Organizational Behavior and Human Decision Processes*, 164, 27–51.
- Dunlea, J. P., & Heiphetz, L. (2021). Children's and adults' views of punishment as a path to redemption. *Child Development*, 92, e398–e415.
- Fehr, E., & Fischbacher, U. (2004). Third party punishment and social norms. *Evolution and Human Behavior*, 25, 63–87.
- FeldmanHall, O., Sokol-Hessner, P., Van Bavel, J. J., & Phelps, E. A. (2014). Fairness violations elicit greater punishment on behalf of another than for oneself. *Nature Communications*, 5, 5306.
- Fischer, C. B., Adrien, N., Silguero, J. J., Hopper, J. J., Chowdhury, A. I., & Werler, M. M. (2021). Mask adherence and rate of COVID-19 across the United States. *PLoS One*, 16 e249891.
- Gelfand, M. J., Harrington, J. R., & Jackson, J. C. (2017). The strength of social norms across human groups. *Perspectives on Psychological Science*, 12, 800–809.
- Gollwitzer, A., Marshall, J., Lee, Y., Deutchman, P., Warneken, F., & McAuliffe, K. (2022, March 20). Parent and community political partisanship predict children's preventative COVID-19 responding. <https://doi.org/10.31234/osf.io/djhkp>.
- 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.
- Jordan, J., Hoffman, M., Bloom, P., & Rand, D. (2016). Third-party punishment as a costly signal of trustworthiness. *Nature*, 530, 473–476.
- Katz, J., Sanger-Katz, M., & Quealy, K. (2020, July 17). A detailed map of who is wearing masks in the U.S. *The New York Times*. <https://www.nytimes.com/interactive/2020/07/17/upshot/coronavirus-face-mask-map.html>.
- Labotka, D., & Gelman, S. A. (2022). *Children's and parents' beliefs concerning viral disease transmission: A comparison of COVID-19 and the common cold*. Unpublished manuscript, University of Michigan.
- Lee, Y., & Warneken, F. (2020). Children's evaluations of third-party responses to unfairness: Children prefer helping over punishment. *Cognition*, 205 104374.
- Lee, Y.-e., & Warneken, F. (2022). Does third-party punishment in children aim at equality? *Developmental Psychology*. In *Advance online publication*. <https://doi.org/10.1037/dev0001331>.
- Lenth, R. V. (2021). *emmeans: Estimated marginal means, aka least-squares means*. R package, Version 1.7.0. <https://CRAN.R-project.org/package=emmeans>.
- Leotti, L., Pochinki, N., Reis, D., Bonawitz, E., & LoBue, V. (2021). Learning about germs in a global pandemic: Children's knowledge and avoidance of contagious illness before and after COVID-19. *Cognitive Development*, 59 101090.
- Liu, X., Yang, X., & Wu, Z. (2021). To punish or to restore: How children evaluate victims' responses to immorality. *Frontiers in Psychology*, 12. <https://doi.org/10.3389/fpsyg.2021.696160>.
- Marshall, J., Yudkin, D. A., & Crockett, M. J. (2021). Children punish third parties to satisfy both consequentialist and retributive motives. *Nature Human Behaviour*, 5, 361–368.
- Martin, J. W., Martin, S., & McAuliffe, K. (2021). Third-party punishment promotes fairness in children. *Developmental Psychology*, 57, 927–939.
- McAuliffe, K., Jordan, J. J., & Warneken, F. (2015). Costly third-party punishment in young children. *Cognition*, 134, 1–10.
- McAuliffe, K., & Dunham, Y. (2021). Children favor punishment over restoration. *Developmental Science*, 24 e13093.
- Pew Research Center. (2020). *More Americans say they are regularly wearing masks in stores and other businesses*. <https://www.pewresearch.org/fact-tank/2020/08/27/more-americans-say-they-are-regularly-wearing-masks-in-stores-and-other-businesses>.
- R Core Team (2021). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org>.
- Raihani, N. J., & Bshary, R. (2015). Third-party punishers are rewarded, but third-party helpers even more so. *Evolution*, 69, 993–1003.
- Riedl, K., Jensen, K., Call, J., & Tomasello, M. (2015). Restorative justice in children. *Current Biology*, 25, 1731–1735.
- Smetana, J. G. (1981). Preschool children's conceptions of moral and social rules. *Child Development*, 52, 1333–1336.
- Smetana, J. G. (2013). Young children's moral and conventional understanding. In M. Benaji & S. 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., Conry-Murray, C., & Sturge-Apple, M. L. (2012). Reciprocal associations between young children's developing moral judgments and theory of mind. *Developmental Psychology*, 48, 1144–1155.
- Stallen, M., Rossi, F., Heijne, A., Smidts, A., De Dreu, C. K. W., & Sanfey, A. G. (2018). Neurobiological mechanisms of responding to injustice. *Journal of Neuroscience*, 38, 2944–2954.
- The Teacher Toolkit. (n.d.). *Student leaders*. Retrieved April 7, 2022, from <https://www.theteachertoolkit.com/index.php/tool/student-leaders>.
- Touloumis, A. (2015). R package multgee: A generalized estimating equations solver for multinomial responses. *Journal of Statistical Software*, 64(8), 1–14 <https://www.jstatsoft.org/v64/i08>.

- Vaish, A., Herrmann, E., Markmann, C., & Tomasello, M. (2016). Preschoolers value those who sanction non-cooperators. *Cognition*, *153*, 43–51.
- Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with S* (4th ed.). New York: Springer.
- Vittrup, B., & Holden, G. W. (2010). Children's assessments of corporal punishment and other disciplinary practices: The role of age, race, SES, and exposure to spanking. *Journal of Applied Developmental Psychology*, *31*, 211–220.
- Yang, X., Wu, Z., & Dunham, Y. (2021). Children's restorative justice in an intergroup context. *Social Development*, *30*, 663–683.
- Yudkin, D. A., Van Bavel, J. J., & Rhodes, M. (2020). Young children police group members at personal cost. *Journal of Experimental Psychology: General*, *149*, 182–191.