Running head: IMMORAL IMPOSSIBILITY

Do children believe immoral events are magical?

Jonathan Phillips1 & Paul Bloom2

1 - Department of Psychology, Harvard University

2 - Department of Psychology, Yale University

**Abstract**

Previous research has observed that 3- to 5-year-old children judge that others ‘cannot’ do actions that are immoral. One interpretation of these findings is that children’s responses merely indicate their understanding that these actions should not be done. An alternative is that young children’s understanding of possibility differs from adults’ understanding, and they actually do not think such actions are possible. Across two studies, we provide evidence for the latter interpretation. We find that young children judge that moral violations are as ‘impossible’ as violations of the laws of physics, and judge that such moral violations would actually require ‘magic’ to occur.

For adults, the impossible and the immoral are clearly different categories. However, there is some intriguing evidence that young children may not clearly distinguish these two. Kalish (1998) and Chernyak and her colleagues (Chernyak, Kushnir & Wellman, 2010; Chernyak, Kushnir, Sullivan & Wang, 2013) have found that 3- to 5-year old children’s beliefs about which actions can be done are constrained by what is moral—they tend to say that agents can’t do something if it’s morally wrong (e.g., making a friend cry).

One possibility is that this pattern merely reflects the language being employed in these studies. All of this research tests children’s intuitions by asking them questions with the English modal auxiliaries ‘can’ and ‘could’. The problem here is that these linguistic constructions can not only be used to indicate that some action is not physically possible (e.g., ‘You can’t fly.’), but can also indicate that some action ‘should not’ be done (e.g., ‘You can’t hit your sister.’) (see Kratzer, 2012; Palmer, 2014). And so when children respond to a question by saying that one *can’t* do something immoral, they might be construing the question as being about morality, not possibility. This would suggest that they differ from adults only in their competence in understanding the question—adults are savvy enough to appreciate that they are being asked about possibility, not morality—not in their actual appreciation of what’s actually possible and what’s not.

Yet, another possibility is that, unlike adults, children actually don’t distinguish the immoral from the impossible. In fact, there is good reason to believe in general that young children’s understanding of possibility differs from adults’ understanding. Shtulman and Carey (2007) presented 4- to 8-year-old children and adults with examples of a number of different actions, some of which were physically impossible (e.g., eating lightning for dinner), some of which were highly improbable (e.g., finding an alligator under one’s bed), and some of which were completely ordinary (e.g., wearing a baseball cap). Participants were asked whether these actions could happen in real life. Intriguingly, younger children judged that both the physically impossible and the statistically improbable actions could not happen in real life. Subsequent research also found that young children judge that these statistically improbable events are ‘impossible’ and would actually require ‘magic’ to happen (Shtulman, 2009).

Here, we explore whether young children also represent immoral actions as impossible. Across two studies, we test whether young children judge that moral violations are as ‘impossible’ as violations of the laws of physics (Study 1), and whether such moral violations would actually require ‘magic’ to occur (Study 2).

**Study 1**

**Participants.** 40 4- to 5-year-old children (Mage(SDage) = 5.10 (0.45)) and 40 6- to 7-year-old children (Mage(SDage) = 6.91 (0.57)) were recruited to participate in the study either through a database of local families who live near New Haven, CT, or at local museums. 80 adult participants (Mage(SDage) = 30.19(9.75)) were recruited from Amazon’s Mechanical Turk ([www.mturk.com](http://www.mturk.com)) (Buhrmester, Kwang & Gosling, 2011).

**Methods.** The experiment was presented to children on an iPad which displayed the stories and accompanying pictures. Children indicated their answers by pressing one of two buttons on the screen corresponding to ‘possible’ and ‘impossible’. Children first completed six training trials. These trials were used to familiarize children with the task of judging an event to be possible or impossible. Three of the training trials involved events that were clearly possible (e.g., putting a small toy hammer in one’s pocket), and three of the training trials involved events that were clearly physically impossible (e.g., putting a large office chair in one’s pocket). Children were corrected if they incorrectly judged the possibility of these events on the training trials. After completing the training trials, children were read a series of eight stories involving young children, each with an accompanying picture of the characters in the story. After reading the story, children were presented with a new event that could follow the story and were asked whether that event was possible or impossible. The order in which the stories were presented was randomized, and the type of event children were asked about (physical violation, probability violation, moral violation, non-violation) was randomly selected on each trial.

Adult participants were presented with the same stories, pictures and test items as children. While adult participants completed the same six training trials, they were not given feedback on their answers. Additional demographic information was collected from the adult participants.

**Stimuli.** The stimuli consisted of eight stories that were presented along with an illustration of the scenario described in the story. For example, children were read the following story about a boy named Henry at a grocery store:

This is Henry. Henry is standing in line at a grocery store when he sees a candy bar that he really would like. He is very hungry after school and asks his mom if she will buy him the candy bar, but his mom says ‘No.’

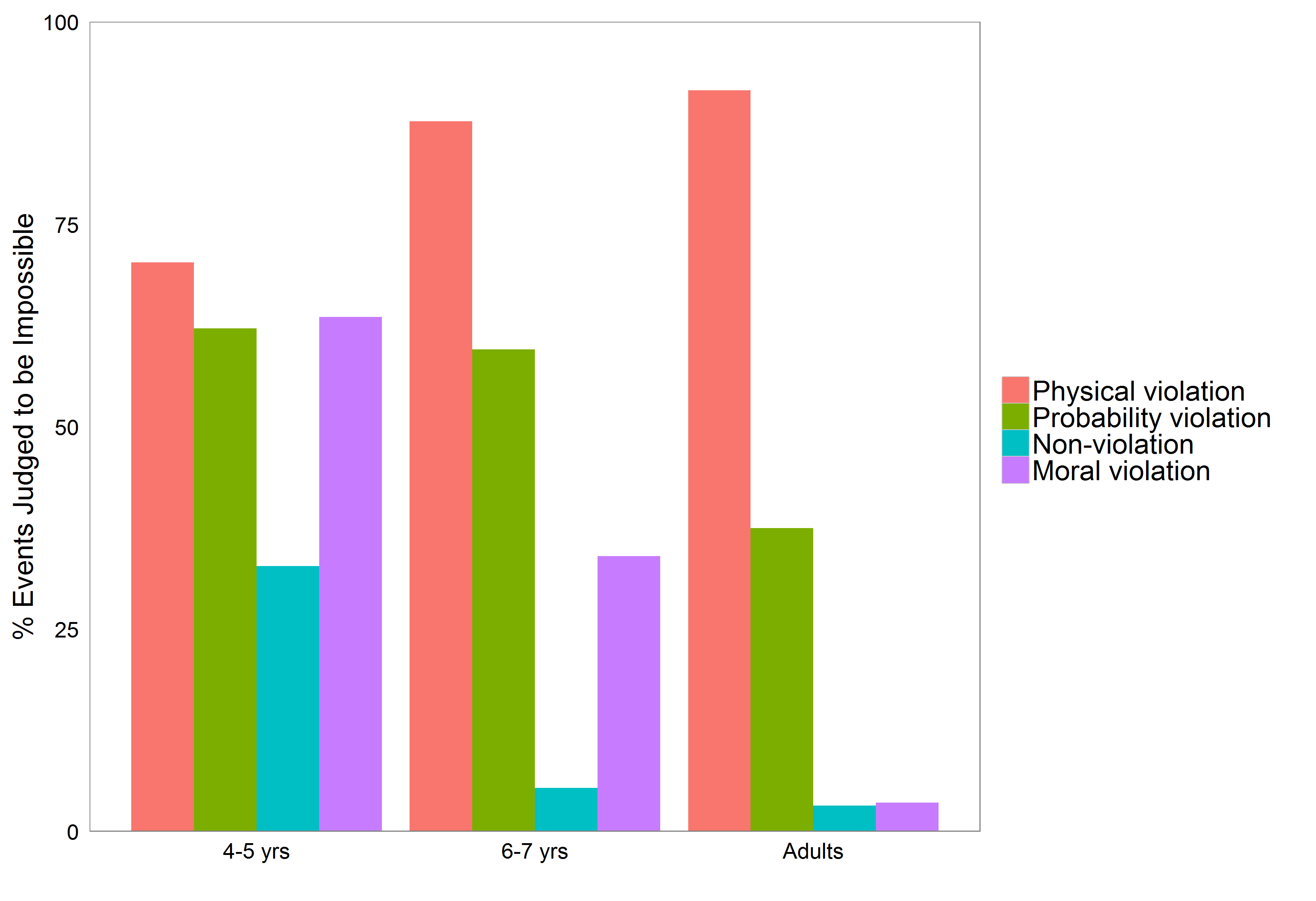
Two versions of each story were created, one with a male and one with a female protagonist. Children and adults were always read stories about protagonists who matched their sex. In each scenario, four test events were created, including one that was physically impossible (e.g., Henry throws his hat into the air and it turns into a candy bar), one that was highly improbable (e.g., Henry tells the man at the store that he should get all of the candy bars in the store for free, and the man gives them to him), one that involved an immoral action (e.g., Henry takes the candy bar without paying and doesn’t tell his mother), and one that was completely ordinary (e.g., Henry waits to get home to eat his favorite snack).

**Sample Size.** Sample sizes in previous research on children’s judgments of possibility have varied from 12 to 32 per age group (Kalish, 1998, Chernyak et al., 2013, respectively). As these studies are the first investigation into children’s impossibility and magic judgments of moral violations, we opted for a slightly larger sample, with ~20 children per age year and thus 40 children per age group (4- to 5-year-olds and 6- to 7-year-olds). In addition, we collected a matched sample of 80 adult participants for comparison (1 participant failed to complete the study).

**Exclusion Criteria.** No exclusion criteria were employed, and all participants were included in the analyses as long as they completed the study.

**Statistical approach.** The overall analyses were conducted with generalized linear mixed effects models with the structure: possibility.judgment ~ violation.type \* age.group + (violation.type | participant) + (1 | scenario) using the lme4 package in R (Bates, Maechler, & Bolker, 2012). This approach allowed for both participants and scenarios to be analyzed as random factors (Gelman & Hill, 2007; Jaeger, 2008). The significance of each effect is determined by a comparing a model that includes the relevant term in the model (as well as the other factors that are not currently being investigated) to a model that does not include that term (but does include all of the other factors that are not under investigation). The effect is determined to be significant if the fit of the model including the term differs significantly from the fit of the model that does not include that term (Barr, Levy, Scheepers, & Tily, 2013). For simplicity and clarity, follow-up pairwise comparisons have been included as a way of illustrating the patterns observed in the overall analysis. Data, analysis code, stimuli, and other supporting materials are available at: https://github.com/phillipsjs/deonticMagic

**Results**. The overall analysis revealed a main effect of Age, *Χ*2(2) = 14.29, *p* < .001, a main effect of Violation Type, *Χ*2(3) = 215.7, *p* < .001, and most importantly, an Age \* Violation Type interaction, *Χ*2(6) = 93.57, *p* < .001 (Fig. 1). Chi-squared pairwise-comparisons were used to characterize the critical interaction effect.

**

*Figure* 1. Percentage of events for each violation type that participants judged to be impossible as a function of age.

4- to 5-year-olds’ judgments of possibility did not differentiate between physical violations and moral violations *Χ*2(1) = 0.566, *p* = .452, *V* = .057. Additionally, they did not differentiate between probability violations and physical violations *Χ*2(1) = 1.376, *p* = .241, *V* = .094, or probability and moral violations, *Χ*2(1) = 0.099, *p* = .753, *V* = .025. They did, however, differentiate between each of these types of events and non-violations (*p*’s < .002).

In comparison, 6- to 7-year-olds’ possibility judgments did differentiate between the possibility of physical and moral violations, *Χ*2(1) = 37.248, *p* < .001, *V* = .500. Additionally, they also differentiated probability violations from physical violations *Χ*2(1) = 13.545, *p* < .001, *V* = .291, and probability violations from moral violations *Χ*2(1) = 7.846, *p* = .005, *V* = .214. However, they continued to judge each of these three violation-types to be significantly more impossible than non-violations (*p*’s < .001).

Adult participants did not differentiate between non-violations and moral violations, *Χ*2(1) < 0.001, *p* > .99, *V* < .001, and clearly did differentiate moral violations from physical violations *Χ*2(1) = 250.110, *p* < .001, *V* = .887. Somewhat surprisingly, they also continued to judge that probability violations were more impossible than non-violations *Χ*2(1) = 45.511, *p* < .001, *V* = .372.

**Discussion.** Overall, these results suggest that children not only think that immoral events *should* not happen, but also think that such events are *impossible*. One potential concern with this study, however, is that the youngest children may not have been familiar with the terms ‘possible’ and ‘impossible.’ While children were familiarized with these terms using only events that involved physical violations, it remains possible that the youngest children incorrectly employed the term ‘impossible’ to indicate that they did not think these events should happen, or did not want them to happen. To address this concern, participants in Study 2a were instead asked whether such events would require ‘magic’ to occur, as even young children have an intuitive understanding of magic (Johnson & Harris, 1994). Additionally, because adults judged a large percentage of the probability violations to be impossible, these items were replaced with less extreme probability violations to allow for a better comparison for how both morality and probability affect judgments of possibility over the course of development. Finally, in Study 2b, adult participants were asked to rank the relative likelihood of the events presented to children, providing a rough comparison of the likelihood of children experiencing similar moral or probability violations.

**Study 2a**

**Participants.** 20 older 3-year-olds (*M*age(*SD*age) = 3.66(0.27)), 40 4- to 5-year-olds (*M*age(*SD*age) = 4.95(0.57)), and 40 6- to 7-year-olds (*M*age(*SD*age) = 7.04(0.65)) were recruited to participate in the study through a database of local families who live near New Haven, CT, or at local museums. 80 adult participants (*M*age(*SD*age) = 30.57(9.53)) were recruited from Amazon’s Mechanical Turk ([www.mturk.com](http://www.mturk.com)).

**Methods.** The methods in Study 2a were identical to those used in Study 1 except that participants were instead asked to indicate whether a given violation would require magic to happen. Accordingly, the term ‘Impossible’ was replaced with the term ‘Magic’ and term ‘Possible’ was replaced with the phrase ‘No Magic’.

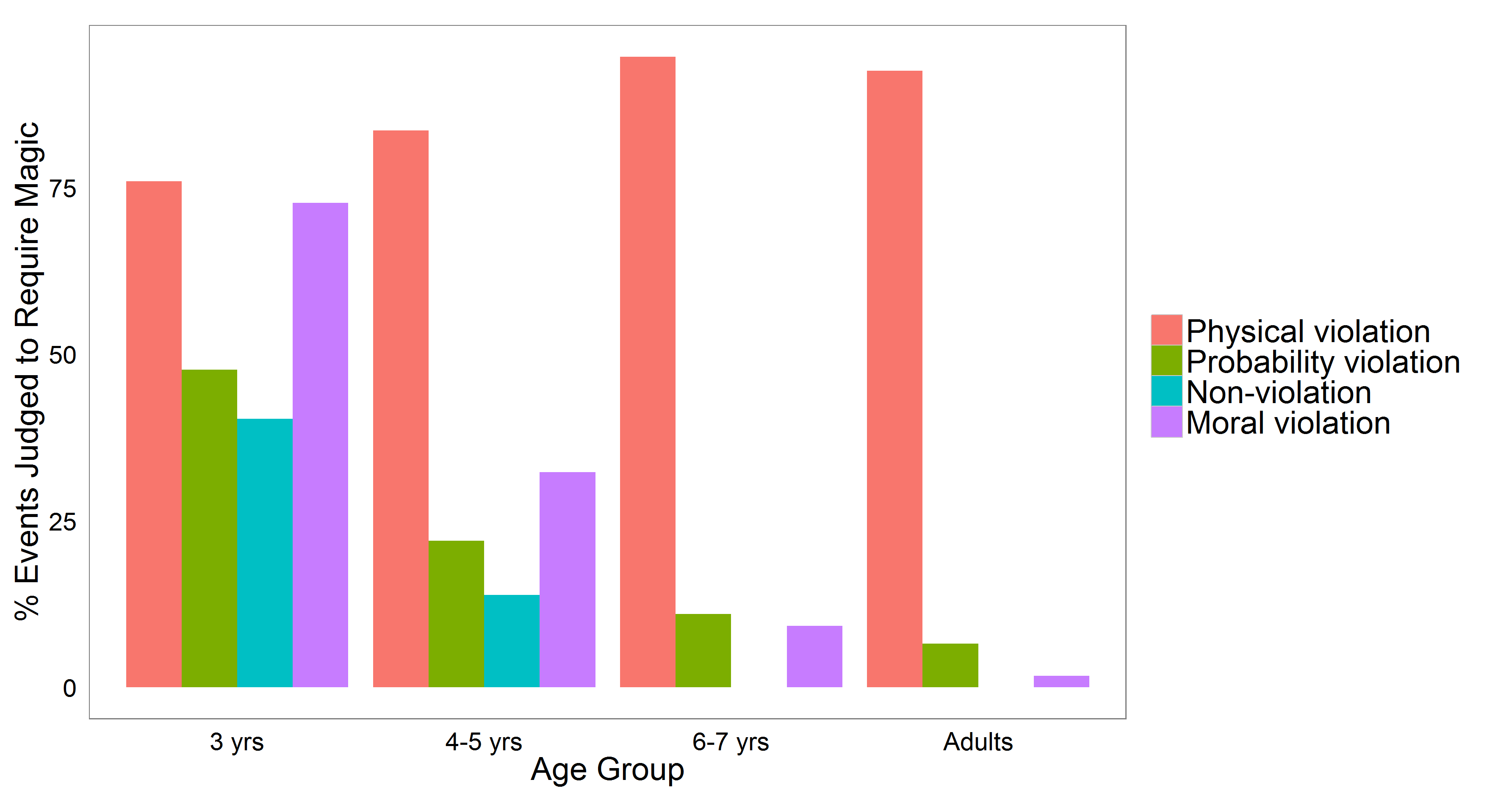
**Stimuli.** The stimuli used in Study 2 were identical to those in Study 1 except that the probability violation items used in Study 1 were replaced. The new probability violation items, while still improbable, were designed to be less extreme. For example, instead of the man in the store giving Henry all of the candy bars for free (as in Study 1), Henry could randomly be given the type of candy bar he wanted by another child upon leaving the store. The complete set of stimuli used in both studies are available in the supporting materials.

**Sample Size.** Sample sizes were determined in the same manner as Study 1 with 20 children per age year and an additional sample of 80 adults (2 participants failed to complete the study). Because this study did not involve the unfamiliar terms ‘Possible’ and ‘Impossible,’ but the simpler ‘Magic’ and ‘No Magic,’ we additionally collected a sample of 20 older 3-year-olds.

**Exclusion Criteria.** No exclusion criteria were employed, and all participants were included in the analyses as long as they completed the study.

**Statistical approach.** Study 2a used the same statistical approach as Study 1.

**Study 2a Results.** The overall analysis revealed a main effect of Age, *Χ*2(3) = 37.95, *p* < .001, a main effect of Violation Type, *Χ*2(3) = 315.85, *p* < .001, and most importantly, an Age\*Violation Type interaction, *Χ*2(9) = 93.47, *p* < .001 (Fig. 2). Chi-squared pairwise-comparisons were used to characterize the key interaction effect.

****

*Figure* 2. Percentage of events in each violation type that participants judged to require magic as a function of age.

3-year-olds’ judgments of magic did not distinguish physical violations from moral violations *Χ*2(1) = 0.206, *p* = .650, *V* = .050. They did, however, judge that physical violations required more magic than probability violations *Χ*2(1) = 3.886, *p* = .049, *V* = .220. Moreover, their magic judgments differentiated both physical and moral violations from non-violations (*p*’s < .001, *V*’s > .3), but did not distinguish between non-violations and probability violations, *Χ*2(1) = 1.459, *p* = .227, *V* = . 136 or moral violations and probability violations, *Χ*2(1) = 1.587, *p* = . 208, *V* = .144.

4- to 5-year-olds’ judgments of magic, by contrast, did differentiate between events involving physical and moral violations, *Χ*2(1) = 43.821, *p* < .001, *V* = .502, as well as events involving physical and probability violations *Χ*2(1) = 58.744, *p* < .001, *V* = .593. Similar to 3-year-olds, however, they continued to differentiate both physical and moral violations from non-violations (*p*’s < .01, *V*’s > .21), but not probability violations from non-violations, *Χ*2(1) = 2.127, *p* = .145, *V* = .121, or moral violations from probability violations, *Χ*2(1) = 1.512, *p* = . 219, *V* = .093.

6- to 7-year-olds’ magic judgments differentiated events involving physical violations from all other event types (*p*’s < .001, *V*’s > .86), but did not differentiate between any of the other event types (*p*’s > .05, *V*’s < .16).

Similarly, adult participants, differentiated physical violations from all other event types (*p*’s < .001, *V*’s > .84). Finally, while not predicted, adult participants’ magic judgments indicated that it would require more magic for a somewhat improbable event to happen than for an event that did not involve any violations, *Χ*2(1) = 7.922, *p* = .005, *V* = .161. They did not distinguish between any other event types.

**Study 2b**

**Participants.** 181adult participants (*M*age(*SD*age) = 31.86(10.97)) were recruited from Amazon’s Mechanical Turk ([www.mturk.com](http://www.mturk.com)).

**Methods.** Adult participants were presented with all of the items used in Study 2a. In this case, participants read scenarios and then ranked all four possible items for that scenario (physical violation, probability violation, moral violation, non-violation) in order of the relative likelihood of the events occurring. Additional demographic information was also collected from these adult participants.

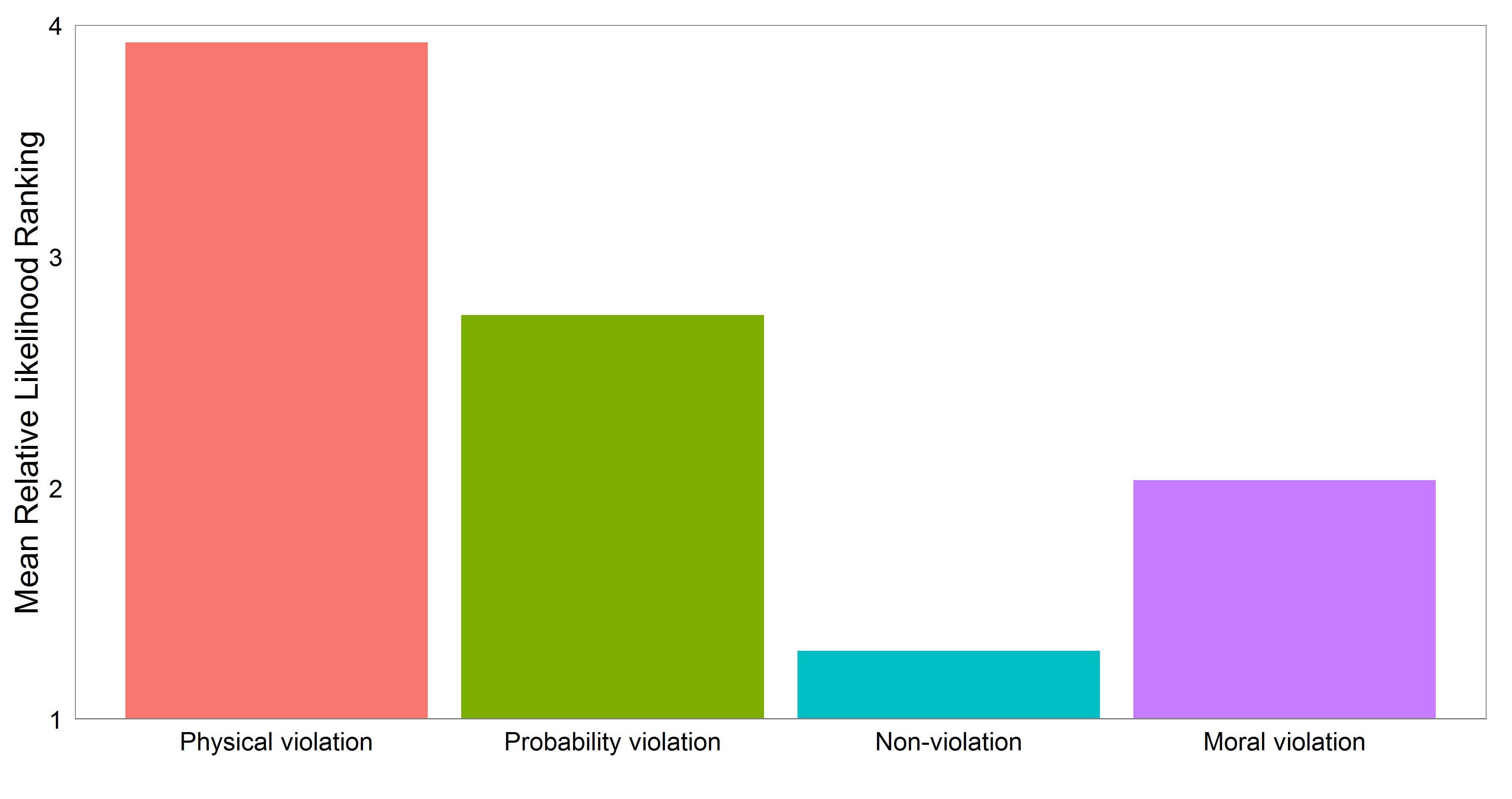
**Stimuli.** The stimuli used in Study 2b were identical to those used in Study 2a.

**Sample Size.** A sample of 181 adults were recruited to test the relative likelihood of the test items used in Study 2b. A comparatively large sample was recruited to ensure that each of the eight stories received no less than 20 relative likelihood judgments for the four test items used in each story.

**Exclusion Criteria.** No exclusion criteria were employed, and all participants were included in the analyses as long as they completed the study.

**Statistical approach.** Each of the 32 items judged by adults were assigned a mean score of the relative likelihood ranking (between 1 (most likely) and 4 (least likely)). Comparisons of relative likelihood were then performed at the level of the item. Data, analysis code, and other supporting materials are available at: https://github.com/phillipsjs/deonticMagic

**Study 2b Results and Discussion.** Item-wise comparisons of mean likelihood ranking revealed that adults judged the physically impossible events to be significantly more unlikely than the improbable events, *t*(7.164) = 19.01, *p* < .001, *d* = 9.50; improbable events were judged to be more unlikely than immoral events, *t*(14) = 6.57, *p* < .001, *d* = 3.28; and immoral events were judged to be more unlikely than events that did not involve any violations, *t*(14) = 6.97, *p* < .001, *d* = 3.49 (Fig. 2). To the extent that adults’ likelihood judgments are a rough estimate of what children themselves may experience, these results suggest that children would be more likely to have witnessed events similar to the moral violations than either the physical or probability violations.

****

*Figure* 3. Mean relative likelihood rankings (from 1 ‘most likely’ to 4 ‘least likely’ for each of the events by type.

**Study 2 Discussion.** Expanding on the previous study, these results suggest young children judge that immoral events require magic to happen. This finding is particularly striking in light of the fact that children have a relatively robust understanding of magic (Johnson & Harris, 1994), and are likely to have not only witnessed but actually preformed many of the immoral actions they judged to require magic (e.g., taking another child’s toy, lying to one’s parent, being mean to another child).

**General Discussion**

While previous research has demonstrated that children judge highly improbable events to be impossible, the present studies show that young children also represent immoral events as impossible (Study 1) and would require magic to happen (Study 2). Moreover, unlike earlier research (Chernyak, et al., 2013; Chernyak, et al., 2010; Kalish, 1998), the present results cannot be explained by an appeal to linguistic ambiguity as we did not employ modal auxiliaries such as ‘cannot’ or ‘could not.’

While the present studies provide strong evidence that young children think of immoral events as impossible, important questions remain about how we should understand children’s representation of these events. One interpretation of the present studies is that they demonstrate that children represent events that are *morally* wrong as involving a violation of some law of *physics* (and are thus impossible for this reason). However, it is not clear that such a strong conclusion is warranted. A more plausible alternative is that young children’s conception of impossibility is less sophisticated than that of adults’ – they simply do not distinguish between the different ways an event may be ruled out (because it involves physics being violated, because it is statistically improbable, or because it is morally wrong). On this alternative picture, young children may only represent events as being genuinely possible when they do not involve any kind of violation (physical, statistical, moral, etc.), and then collectively represent all of the events that do involve violations as impossible in a relatively undifferentiatedway. If correct, this picture suggests that what develops as children age is a more sophisticated sense of impossibility, one that eventually allows for adults to realize that immoral actions can happen, even if they should not.

**References**

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3), 255-278. doi:10.1016/j.jml.2012.11.001

Bates, D., Maechler, M. & Bolker, B. (2012). lme4: Linear mixed-effects models using S4 classes. R package version 0.999999-0[. http://CRAN.R-project.org/package=lme4](http://cran.r-project.org/package=lme4)

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk a new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science*, 6(1), 3-5.

Chernyak, N., Kushnir, T., Sullivan, K. M., & Wang, Q. (2013). A comparison of Nepalese and American children’s concepts of freedom of choice and social constraint. *Cognitive Science*, 37, 1343-1355.

Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press.

Johnson, C., & Harris, P. L. (1994). Magic: Special but not excluded. *British Journal of Developmental Psychology*, 12, 35–51.

Kalish, C. (1998). Reasons and causes: Children's understanding of conformity to social rules and physical laws. *Child Development*, 69(3), 706-720.

Kratzer, A. (2012). *Modals and conditionals: New and revised perspectives* (Vol. 36). Oxford University Press.

Levy, G. D., Taylor, M. G., & Gelman, S. A. (1995). Traditional and evaluative aspects of flexibility in gender roles, social conventions, moral rules, and physical laws. *Child Development*, 66(2), 515-531.

Palmer, F. R. (2014). *Modality and the English modals*. Routledge.

Shtulman, A. (2009). The development of possibility judgment within and across domains*. Cognitive Development*, 24, 293-309.

Shtulman, A., & Carey, S. (2007). Improbable or impossible? How children reason about the possibility of extraordinary events. *Child Development*, 78, 1015-1032.

Shtulman, A., & Tong, L. (2013). Cognitive parallels between moral judgment and modal judgment. *Psychonomic Bulletin & Review*, 20, 1327-1335.