Running head: INTENTION-BASED MORAL JUDGMENT

The Development of Intention-Based Morality: The Influence of Intention Salience and Recency, Negligence, and Outcome on Children's and Adults' Judgments

Gavin Nobes, Georgia Panagiotaki, and Paul Engelhardt
University of East Anglia

Author note

Gavin Nobes, School of Psychology, University of East Anglia, Norwich, UK; Georgia Panagiotaki, Department of Clinical Psychology, Norwich Medical School, University of East Anglia, Norwich, UK; Paul Engelhardt, School of Psychology, University of East Anglia, Norwich, UK.

Correspondence concerning this article should be addressed to Gavin Nobes, School of Psychology, University of East Anglia, Norwich Research Park, Norwich, NR4 7TJ, United Kingdom. Email: g.nobes@uea.ac.uk

Date of submission: 4th October 2016, revised 12th January 2017

Abstract

Two experiments were conducted to investigate the influences on 4-8-year olds' and adults' moral judgments. In both, participants were told stories from previous studies that had indicated that children's judgments are largely outcome-based. Building on recent research in which one change to these studies' methods resulted in substantially more intention-based judgment, in Experiment 1 (N = 75) the salience and recency of intention information were increased, and in Experiment 2 (N = 99) carefulness information (i.e., the absence of negligence) was also added. In both experiments even the youngest children's judgments were primarily intention-based, and in Experiment 2 punishment judgments were similar to adults' from 5-6 years. Comparisons of data across studies and experiments indicated that both changes increased the proportion of intention-based punishment judgments – but not acceptability judgments – across age-groups. These findings challenge and help to explain those of much previous research, according to which children's judgments are primarily outcome-based. However, younger participants continued to judge according to outcome more than older participants. This might indicate that young children are more influenced by outcomes than are adults, but other possible explanations are discussed.

Keywords: moral judgments; intention; outcome; negligence; salience

The Development of Intention-Based Morality: The Influence of Intention Salience and Recency, Negligence, and Outcome on Children's and Adults' Judgments

Intention-based moral judgment is a fundamental component of mature morality: adults typically judge actions according to whether they are well- or ill-intentioned. In contrast, researchers have repeatedly reported that young children base their moral judgments more on the outcomes of actions than on the intentions of the agents (e.g., Buchanan & Thompson, 1973; Cushman, Sheketoff, Wharton, & Carey, 2013; Elkind & Dabek, 1977; Farnill, 1974; Gummerum & Chu, 2014; Helwig, Zelazo, & Wilson, 2001; Imamoğlu, 1975; Killen, Mulvey, Richardson, Jampol, & Woodward, 2011; Piaget, 1932/1965; Walden, 1982; Yuill, 1984; Zelazo, Helwig, & Lau, 1996). However, there remains considerable debate about this issue, and several studies have reported that young children (e.g., Baird & Astington, 2004; Bearison & Isaacs, 1975; Chandler, Greenspan, & Barenboim, 1973; Nelson, 1980; Nobes, Panagiotaki, & Pawson, 2009; Nummedal & Bass, 1976; Vaish, Carpenter, & Tomasello, 2010), and even 8-month-olds (Hamlin, 2013), are influenced primarily by intentions.

Rather than seeking to add to this already large body of often-conflicting evidence, perhaps the main objective of researchers in this field should be to investigate the relative validity of previous studies' contrasting claims. Duncan, Engel, Claessens, and Dowsett (2014) argue that this should be done by testing the replicability and robustness of previous research. However, despite the scores of studies of children's moral judgment since Piaget's (1932 / 1965) seminal work, this approach has been taken remarkably rarely.

Two very similar studies that provide strong evidence for the claim that children judge primarily according to outcome are those of Helwig et al. (2001) and Zelazo et al. (1996). Children and adults were told stories about agents who accidentally made peers or pet animals happy or sad. The researchers reported that participants of all ages judged attempted harms (ill-intentioned actions with positive outcomes) to be good, and accidental harms (well-intentioned actions with negative outcomes) to be bad; moreover, children – though not adults – considered

the well-intentioned agents to deserve punishment, and the ill-intentioned agents to deserve none. That is, regardless of age, almost all participants made outcome-based acceptability judgements, and children (the oldest were nearly 8) made primarily outcome-based punishment judgments.

Nobes, Panagiotaki and Bartholomew (2016) replicated Helwig et al.'s and Zelazo et al.'s methods and corroborated their findings. However, when the wording of the acceptability question was changed from the original (e.g., "Is it okay for Kevin to give Rob a puppy?")[1] to being more agent-focused (e.g., "Is Kevin good, bad or just okay?"), the older children's and adults' responses were essentially reversed: they judged almost exclusively according to intentions. Moreover, despite Helwig et al.'s and Zelazo et al.'s other key question – whether the agent should be punished – remaining unchanged, punishment judgments also became substantially more intention-based. Even the youngest children's (4-5 years) acceptability and punishment judgments were now based approximately equally on intention and outcome.

Nobes et al. (2016) also asked participants who made outcome-based punishment judgments a "parental knowledge" question (e.g., "If her parents found out she tried to hit the dax, should they tell her off?"). The large majority now gave intention-based responses. This shows that even participants who give outcome-based judgments are usually aware of, and can base their judgments on, intentions. It also suggests that some – perhaps many – apparently outcome-based responses result from the belief that, since parents cannot know their children's intentions, they tend to assume that a bad outcome is culpable and punishable because it was probably deliberate.

These findings challenge both the robustness of Helwig et al.'s and Zelazo et al.'s findings and the validity of the claim that children's judgments are primarily outcome-based. However, they are consistent with a "weak" version of this account: while children's

[1] Helwig et al. and Zelazo et al. asked about the acceptability of acts in this way because, as well as investigating the relative influence of intention and outcome on moral judgments, they sought to address the separate issue of whether children judge according to acts (e.g., petting or hitting an animal) or the harm that results from them.

judgments are not *primarily* outcome-based, there is still an "outcome-to-intent shift" (Cushman et al., 2013) because, with increasing age, individuals judge less according to outcome, and more according to intention.

One explanation is that children's judgments are indeed less intention-focused, and more outcome-focused, than adults'. Cushman and colleagues (e.g., Cushman, 2008; Cushman et al., 2013) have proposed a dual-process model that comprises an early-developing, relatively automatic process that is sensitive to the causes of outcomes, and a later-developing process that is sensitive to mental states, especially intentions. Until about 5 years of age children have only the first, causal, process available and so are influenced almost entirely by outcomes; they then gradually become able also to take intentions into account.

An alternative explanation is that children are able to make intention-based moral judgments from an early age, but lack the cognitive resources (e.g., memory, executive functions, theory of mind) required to remember, understand and integrate intention information in their judgments, at least when told stories such as those of Helwig et al. and Zelazo et al. In particular, the salience of outcomes might be greater than that of intentions such that young children forget or fail to notice agents' intentions, or are unable to inhibit their emotional or intuitive responses to the outcomes (e.g., Buon, Seara-Cardoso, & Viding, 2016; Margoni & Surian, 2016). This problem of outcomes being more salient than intentions is common in this area of research, not least because outcomes (e.g., a victim's pleasure or pain; a desired gift or broken possession) are typically tangible and explicit, whereas intentions are less easily perceived and understood. Bearison and Isaacs (1975) and Nelson (1980) reported that children's judgments became more intention-based when the intentions were stated explicitly, and hence were more salient than when they were implicit, as in Piaget's (1932 / 1965) stories. In addition, since intentions precede outcomes, they are almost invariably presented both verbally and pictorially in this order, and so recency effects are likely. Feldman, Klosson, Parsons, Rholes and Ruble (1976) and Nummedal and Bass (1976) found that

children's judgments became more intention-based when they reversed the order of presentation from the usual intention then outcome, to outcome followed by intention. More recently, Gvozdic, Moutier, Dupoux and Buon (2016) reported that metacognitive training, and in particular the use of an executive alert (to "not focus too much" on the consequences), resulted in 5-8 year-old children making adult-like, primarily intention-based judgments. This finding supports a cognitive resources account because it indicates that children's outcome-based judgment arises not from an inability to understand and use intention information, but from failure to inhibit their focus on outcomes.

A third possible explanation of the intention-to-outcome shift is that children are able to make intention-based moral judgments from an early age, but are influenced in their judgments by presumed negligence of the agents (Nuñez, Laurent, & Gray, 2014). For example, Helwig et al. told a story about Ethan, who wanted to make his friend happy by giving him a puppy. However, the shopkeeper made a mistake so that Ethan accidentally gave his friend a gift box containing a tarantula. Despite his good intentions, mature, intention-based judges might have considered Ethan blameworthy because he was careless not to check the present before giving it to his friend; that is, they might have judged him naughty and deserving of punishment not because of the outcome per se, but because he was negligent. Since negligence co-varies with outcome (a bad outcome implies carelessness), these mature judgments would appear to be outcome-based. Nobes et al. (2009) found that, when carefulness was explicitly stated, even 3-4 year-olds judged primarily according to intention.

The current study. In this study we tested these separate – though not entirely mutually exclusive – accounts by building on Nobes et al.'s (2016) replications of Helwig et al.'s (2001) and Zelazo et al.'s (1996) studies. Our objectives were to evaluate the claim that young children base their moral judgments on the outcomes of actions rather than on the agents' intentions, and to investigate the reasons for their judgments.

Two experiments were conducted, in each of which a single change was made to

systematically test its effects on moral judgments. Both experiments were near-replications of Helwig et al.'s and Zelazo et al.'s studies, except that the acceptability question was rephrased (as in Nobes et al., 2016), intention salience and recency were increased, and in Experiment 2 carefulness (i.e., lack of negligence) information was added. Data from these experiments were then compared with those of Nobes et al. (2016) to determine the extent to which these factors separately and conjointly accounted for some children's continued outcome-based judgments, and the increase with age of intention-based judgment, that were reported there.

The approach of replicating studies that provided strong support for the view that children's moral judgments are primarily outcome-based, and systematically manipulating one factor at a time, enabled us to identify whether, and to what extent, each factor influenced judgments. It ensured that different findings – in particular, of intention-based judgment – could not be attributed to any other methodological differences, such as variations in the content or presentation of stories.

If it were found that children's moral judgments became largely or wholly intention-based when intention salience and recency were increased, or when carefulness (i.e., absence of negligence) information was added, this would provide strong support for the second (intention salience) or third (negligence) of the possible explanations outlined above. Since these factors have only rarely been investigated before – in the large majority of studies in this area outcomes have been more salient than intentions, and participants have been told nothing about carefulness or carelessness – these explanations would be directly applicable to most previous studies, too.

On the other hand, if there were little or no increase in children's use of intention information compared with that reported by Nobes et al. (2016), this would be consistent with accounts such as that of Cushman et al. (2013), that young children's judgments are strongly influenced by outcome.

Experiment 1

The first experiment investigated the impact on children's and adults' moral judgments of the relative salience and recency of intention and outcome information. The methods were almost identical to those of Helwig et al. and Zelazo et al., except that, as in Nobes et al. (2016), the acceptability question was changed to be more agent-focused than in the original studies. In addition, the salience and recency of intention information was increased.

It was predicted that participants of all ages would judge more according to intention than outcome, that is, accidental harms (positive intentions, negative outcomes) would be considered more acceptable and less punishable than attempted harms (negative intentions, positive outcomes). In particular, we expected that younger children's judgments would be primarily intention-based. We also expected that, when participants gave outcome-based responses, the majority would give intention-based responses to the parental knowledge question.

Method

Sample

The participants were 19 children (12 girls) aged 4-5 years (M = 62.5; range = 58-65 months), 24 (12 girls) aged 5-6 years (M = 74.0; range = 66-82 months); 20 (10 girls) aged 7-8 years (M = 97.1; range = 88-104 months), and 12 adults (5 women) (M = 42 years, range = 18–70 years[2]). The children attended five British state primary and junior schools. The adults were mainly university students and administrative staff. All participants were white except for five African Caribbeans and five South Asians. Children were excluded when parental consent was not given, on teachers' advice (e.g., because children were very shy, or their English was poor), or when children showed signs of boredom or distraction. Two of the youngest children withdrew early.

Nobes et al. (2016) indicated effect sizes of action valence (accidental or attempted

[2] In both experiments reported here there was no indication that adults' judgments changed with age, rs < .10 in magnitude, ps > .74.

harm) when the rephrased acceptability question was asked (as in the present study) of η_p^2 = .452 for acceptability (naughtiness) judgments, and η_p^2 = .185 for punishment judgments. An a priori power analysis using GPower (Faul, Erdfelder, Lang, & Buchner, 2007), with effect size specification as in Cohen (1988), indicated that a sample of 40 participants (i.e., 10 per age group) would be sufficient to detect the smaller effect (for punishment) with power (1 - β) set at 0.80 and α = .05.

Design

The independent variables were age group, action valence and source of story (Helwig et al. or Zelazo et al.) The dependent variables were acceptability (naughtiness) and punishment judgments, justifications, and "parental knowledge" judgments.

Measures

Each participant was told two illustrated stories from Helwig et al. (2001) and two from Zelazo et al. (1996) (Appendix). Each of these pairs comprised one accidental harm (positive intention, negative outcome) and one attempted harm (negative intention, positive outcome). In Helwig et al.'s accidental harm, Ethan had the good intention of giving his friend a puppy, but the shopkeeper put a big spider in the gift box by mistake, so Ethan made his friend scared and upset. In the attempted harm, ill-intentioned Chris wanted to give a big spider, but – because of the same less-than-fastidious shopkeeper – accidentally gave a puppy. Zelazo et al.'s accidental harm involved Sally, who wanted to make her pet (a "dax") happy by stroking it, but the dax jumped up so that Sally accidentally hit it and made it sad. In the attempted harm, Anne also had a dax which she wanted to hurt, but the dax wiggled away so Anne accidentally stroked it, making it happy.

During each story, participants were asked to predict the agents' behavior (e.g., "What is Ethan going to get Chris for his birthday?"), and, in the Helwig et al. stories, they were also asked to predict the recipient's emotions (e.g., "How do you think Chris felt when he got the big spider?") After each story, participants were asked to make two judgments, one about

acceptability (goodness or badness), and the other about the level of punishment that the agent deserved.

These methods were in all relevant respects identical to those used in the original studies (and shown by Nobes et al. (2016) to closely replicate their findings), except for two sets of changes: those made by Nobes et al., and those made for this experiment.

The changes made by Nobes et al. to the original studies were as follows: the acceptability questions were rephrased so that they focused on whether the agent was good or bad (e.g., "Is Sally good, bad or just okay? How good / bad?"); character information (e.g., "Sally is nice. She doesn't want to hurt anyone") was removed because it gave the answer to the rephrased intention question[3]; participants were asked to justify their punishment judgments; and, when they gave apparently outcome-based judgments, they were asked a "parental knowledge" question, e.g., "If her parents found out she tried to hit the dax, should they tell her off?" The justification and parental knowledge questions were asked after the judgments to ensure that they did not influence the judgments. In addition, some words were changed to improve comprehension by our British participants, for example, Anne "stroked" the dax rather than petted it; and Ethan accidentally gave a "big spider" rather than a tarantula.

The following additional changes were made for this experiment: First, three confirmation questions were added directly before the judgment questions. They concerned the outcome (e.g., "Did Sally hit the dax or stroke it?"), cause (e.g., "Why did Sally hit the dax? Did she want to stroke the dax, or did she stroke it by mistake?"), and intention ("What did Sally try to do? Did she try to stroke the dax, or did she try to hit it?"). Wrong responses were corrected by the interviewers. These questions were included to ensure that the key elements of the stories had been understood and remembered (this was not assessed in the original studies),

[3] Removal of character information would be expected to *reduce* intention-based judgment relative to the original studies because it implied the agents' intentions (e.g., Anne was nasty, so she wanted to hurt the dax).

and the second and third of these questions were also designed to increase the salience and recency of intention information relative to those of outcome information.

Second, another post-judgment question was asked in this experiment: participants who gave apparently outcome-based judgments (e.g., to punish a well-intentioned agent) were asked again about the agents' intentions (e.g., "What did Sally want to do the dax? Did she want to make it feel nice or did she want to hurt it?") This question was included to assess whether apparently outcome-based judgments occurred because participants had forgotten about the agents' intentions.

The third change made in this experiment was to exclude the comprehension and confirmation questions that were asked in the original studies. These referred to aspects of the stories that were unrelated to intentions, causes or outcomes and instead asked whether the dax in Zelazo et al. liked being petted or hit, or if the boys in Helwig et al. liked puppies or tarantulas. Since Nobes et al. (2016) reported that 98.5% of children's answers to these questions were correct, and Zelazo et al. approximately 96%, they were considered unnecessary and excluded to avoid the interviews becoming too long.

To summarize, the only relevant change from Nobes et al., (2016) was that the confirmation questions about cause and intention were added. As in Nobes et al., the acceptability question was rephrased so that it focused on the agent rather than the outcome, and character information was removed. All other changes were neutral because they concerned questions that were excluded because they were about irrelevant aspects of the stories, or that occurred after the judgments.

As in the original studies, acceptability judgments were scored from 1 (*really, really bad*), through 3 (*okay*), to 5 (*really, really good*), and punishment from 0 (*no trouble*), through 1 (*a little trouble*), to 2 (*a lot of trouble*). For example, responses that Sally (who was well-intentioned but who accidentally hurt her pet) is good and should not be punished would indicate intention-based judgment, whereas saying that Anne (who was ill-intentioned but who

accidentally made her pet happy) is good and should not be punished would indicate outcome-based judgment.

Justifications were coded according to whether they were based on intentions (e.g., "Because she wanted to smack it", "He didn't mean to, it was an accident"), outcomes (e.g., "Cos he gived a big horrid spider", "The animal was sad and cried"), negligence ("He should have checked inside the box", "She didn't hold on tight enough"), or other (e.g. "Cos she looks like my friend"; "Don't know"). Justifications were also coded according to whether they were factually correct. For example, if a participant said that Sally (who was well-intentioned) should be punished "Because she tried to hurt the dax", their justification would be coded as intention-based, but factually incorrect. A quarter of all justifications were coded by a second independent judge, and interrater reliability was 94.7% (Cohen's $\kappa = .91$).

Pictures were 20cm x 30cm sketches that illustrated the story characters, their intentions (e.g., the agent with, in a thought bubble, his smiling friend), likes and dislikes (e.g., the smiling friend with a puppy), causes (e.g., a shopkeeper places a big spider, instead of a puppy, in a gift box), and outcomes (e.g., the friend looking unhappy with a big spider). The Zelazo et al. pictures were kindly provided by the authors, but the Helwig et al. pictures are no longer available and so were redrawn according to the story texts, in the same style as the Zelazo et al. pictures.

Procedure

Participants were interviewed individually in quiet areas of their schools or university. They were first given an introduction and brief explanation, and then asked if they were happy to continue. The four stories were told in random order, except that either the pair from Helwig et al., or the pair from Zelazo et al., was told first.

Results

Comprehension. The percentages of correct responses to the prediction, confirmation and justification questions, and to the pre- and post-judgment confirmation questions, are

shown in Table 1. Of the 4-5- and 5-6 year-old children, 60.0% and 55.6% respectively gave one or more incorrect responses to the 12 (i.e., 3 for each of the 4 stories) confirmation questions that they were asked, and 48.0% and 22.2% gave 2 or more incorrect responses. All of the older children and adults answered all of these questions correctly.

Acceptability judgments. Preliminary analyses revealed no main or interaction effects of gender or of source of story (Zelazo et al. or Helwig et al.) and so these were excluded from further analyses. A 4 (Age group) x 2 (Action valence [accidental harm, attempted harm]) mixed ANOVA with repeated measures on action valence indicated a main effect of action valence on acceptability judgments, $F(1,69) = 125.67$, $p < .001$, $\eta_p^2 = .65$ (Figure 1). Participants rated well-intentioned actions with bad outcomes more acceptable than ill-intentioned actions with good outcomes, that is, their acceptability judgments were more intention- than outcome-based. This was qualified by an interaction between action valence and age group, $F(3,69) = 13.71$, $p<.001$, $\eta_p^2 =.37$. Pairwise comparisons indicated that adults made this distinction more clearly than all three child age groups (ps < .001), and 7-8 year olds more clearly than 5-6-year olds $p = .03$. Participants of all four age groups rated accidental harms more acceptable than attempted harms, Fs ≥ 8.0, ps ≤ .02, η_p^2s ≥ .28. The main effect of age-group did not approach significance.

Punishment judgments. The equivalent analyses were conducted on punishment ratings (Figure 2). Source of story and gender were again excluded following preliminary analyses. The mixed ANOVA showed a main effect of action valence, $F(1,69) = 56.82$, $p < .001$, $\eta_p^2 = .45$, indicating that well-intentioned actions were judged less punishable than ill-intentioned actions. There was also an interaction between action valence and age-group, $F(3,69) = 3.86$, $p = .013$, $\eta_p^2 = .14$. Pairwise comparisons indicated that adults distinguished action valences more clearly than the 4-5-year-olds and 5-6 year-olds, ps < .01, and marginally more clearly than 7-8-year-olds, $p < .07$. All three older age-groups considered accidental harms less punishable than attempted harms, Fs > 5.9, ps < .03, η_p^2s > .20, and 4-5-year-olds judged them

marginally less punishable, $F(1,18) = 3.77$, $p = .07$, $\eta_p^2 = .18$. The main effect of age-group did not approach significance.

Justifications. Table 2 shows the percentages of participants' justifications of their punishment judgments that were based on intention, negligence and outcome. From 5-6 years justifications were significantly more intention- than outcome-based (binomial *p*s < .01). Adults' justifications were based on intention rather than outcome more than were those of any of the child age-groups', $\chi^2(1)$s > 6.9, *p*s < .01.

Parental knowledge. When outcome-based punishment judgments were made (i.e., an accidental harm-doer should be punished, or an attempted harm-doer should not), participants were asked whether the agent's parents should punish if they knew the agent's intentions**.** Table 3 shows the frequencies of outcome-based punishment judgments and of intention-based responses to the parental knowledge question (i.e., they should not punish an accidental harm-doer, or they should punish an attempted harm-doer).

Discussion

Consistent with the first hypothesis, participants' acceptability and punishment judgments were primarily intention-based: even the youngest children considered accidental harms significantly more acceptable, and marginally more punishable, than the attempted harms. However, there was a substantial effect of age: older participants' judgments were considerably more intention-based than younger participants'.

The second hypothesis was also supported: when participants made outcome-based punishment judgments, the majority at all ages gave intention-based responses to the parental knowledge question.

Responses to the confirmation questions indicated that approximately 10% of each of the key aspects of the stories – intention, cause and outcome – were initially misunderstood by 4-6-year-olds. This suggests that, since these questions were not asked in the original or previous studies – nor, of course, wrong answers corrected – approximately 20% of young children's

judgments in those studies were based on incorrect information. This would have led children who actually made intention-based judgments, but misunderstood agents' intentions, to make apparently outcome-based judgments. Equally, children who made outcome-based judgments, but misunderstood the outcomes, would have made apparently intention-based responses.

Responses to the post-judgment intention questions indicated that approximately a quarter of the youngest children's apparently outcome-based responses, and 10% of the 5-6-year olds', are likely to have been based on misinterpreted or forgotten intention. That is, when these children said that a well-intentioned agent was naughty and punishable, their judgments could actually have been based on the misunderstood or misremembered intentions of the agents.

These findings indicate that the combination of the rephrased acceptability question and increased salience and recency of intention information accounts for almost all of the outcome-based judgment by older participants that was reported by Helwig et al. and Zelazo et al. It also explains most of young children's outcome-based judgment, but not all: there was still a strong effect of age, in that older children's and adults' judgments were more intention-based than young children's. The possibility that this might be due to younger participants assuming that accidental harms were caused by negligence was investigated in Experiment 2.

Experiment 2

This experiment built on Experiment 1 to investigate whether the outcome-based judgments reported there and in Nobes et al. (2016) resulted from assumptions – particularly by young children – that accidental harms were caused by negligence. If so, despite their being well-intentioned, these actions would be considered blameworthy and punishable. Since negligence co-varies with outcome, the result would be that mature, negligence-based moral judgments would appear to be immature, outcome-based judgments.

This possibility was tested by replicating Experiment 1 and adding information about carefulness, that is, the absence of negligence. This information was inserted into each story by

explicitly stating that agents were careful, and by explaining how they were careful. In addition, pictures were added in which the agents were shown being careful. If the negligence account is correct, inclusion of this verbal and pictorial information should decrease the frequency of apparently outcome-based (but actually negligence-based) judgments of accidental harms.

The same carefulness information was added to the attempted harms. Its inclusion might lead to harsher (i.e., more intention-based) judgments since being careful about causing harm could emphasize the agent's malice; on the other hand, participants might consider carefulness to be praiseworthy, regardless of ill-intention, in which case judgments would be less harsh, i.e., apparently more outcome-based.

The first prediction was that, as in Nobes et al. (2016) and Experiment 1, acceptability and punishment judgments would be primarily intention-based. Second, we predicted that, as in Experiment 1, even 4-5 year olds' judgments would be based more on intention than outcome. If young children's tendency to judge more than adults according to outcome resulted largely or wholly from their judgments being negligence-based (i.e., they assumed that bad outcomes resulted from negligence, and judged accordingly), then the addition of carefulness information should result in the differences between younger and older participants' judgments and justifications largely or wholly disappearing.

The third prediction was that the inclusion of carefulness information in this experiment would draw participants' attention to the issue of negligence, and therefore increase the frequency of references to negligence when justifying their judgments.

And fourth, intention-based responses would be elicited from the parental knowledge question, despite its being asked only when punishment judgments were outcome-based.

Sample. The participants were 26 children aged 4-5 years (12 girls, *M* = 61.40; range = 56-65 months), 35 aged 5-6 years (19 girls, *M* = 75.31, range = 66-82 months), 20 aged 7-8 years (10 girls, M = 99.8, range = 94-106 months) and 18 adults (10 women, M = 28, range =

18-54 years). The children attended six British state primary and junior schools in rural and urban areas. Thirteen 4-6 year-olds withdrew early.

An a priori power analysis indicated that a sample of 20 (i.e., 5 per age-group) would be sufficient to detect the effect of action valence with effect size set at $\eta_p^2 = .45$ (as in Experiment 1 for punishment judgments) and power at .95.

Design, measures and procedure. All methods were identical to those of Experiment 1, - including both the agent-focused acceptability question used there and in Nobes et al. (2016), and increased salience and recency of intention information – except that carefulness (i.e., lack of negligence) information was added (Appendix, last column). This was done by explicitly stating that the agent was careful, and explaining how they were careful. For example, Sally, who wanted to make her pet happy by stroking it, "held it very carefully to make sure she didn't hurt it". In addition, another confirmation question was added (e.g., "Did Sally hold the dax carefully to make sure she didn't hurt it?") before the outcome, cause and intention confirmation questions. This was done to assess understanding and awareness of the negligence information; to ensure that children understood that the agent was careful – incorrect responses were corrected; and to increase the likelihood of participants remembering this information. Finally, another picture was added to each story between those showing the intentions and outcomes in which, for example, Sally held the pet carefully on a table with two hands.

A second independent judge coded 25% of the justifications, and interrater reliability was 92.9% (Cohen's $\kappa = .87$).

Results

Comprehension. The percentages of correct predictions, confirmations and judgment justifications are shown in Table 4. Despite being told explicitly that the agents were careful, large proportions of children, and even 10% of adults, considered them to be careless. This was

especially the case regarding Helwig et al.'s characters, who were considered by 53.1% (4-5 years), 49.1% (5-6 years) and 37.5% (7-8 years) of children to have been careless to give the wrong present.

Of the 4-5, 5-6, and 7-8 year-old children, 46.7%, 12.9% and 35.0%, respectively, gave one or more incorrect responses to the 12 (i.e., 3 for each of the 4 stories) outcome, cause and intention confirmation questions that they were asked. 30.0% of 4-5-year-olds, 3.2% of 5-6-year-olds, and 10.0% of the 7-8-year-olds gave 2 or more incorrect responses. All of the adults answered all of these questions correctly.

Acceptability judgments. Gender and source of story were excluded following preliminary analysis. A 4 (Age group) x 2 (Action valence) mixed ANOVA with repeated measures on action valence indicated that the well-intentioned accidental harms were judged more acceptable than the ill-intentioned attempted harms, $F(1,82) = 221.86$, $p < .001$, $\eta_p^2 = .70$ (Figure 3). This was qualified by an interaction between action valence and age group, $F(3,82) = 17.96$, $p < .001$, $\eta_p^2 = .36$. Pairwise comparisons indicated that adults distinguished between accidental and attempted harms more clearly than all three child age-groups (ps $< .001$), and the youngest children did so less clearly than the 5-6-year-olds, $p < .02$, and 7-8-year-olds, $p < .05$. Participants of all age groups considered accidental harms significantly more acceptable than attempted harms, Fs ≥ 7.82, ps $\leq .012$, η_p^2 s $\geq .29$. The main effect of age-group did not approach significance.

Punishment judgments. The equivalent analysis of punishment ratings (Figure 4) also showed a main effect of action valence, $F(1,77) = 91.24$, $p < .001$, $\eta_p^2 = .54$, and a marginally significant interaction between action valence and age-group, $F(3,77) = 2.55$, $p = .06$, $\eta_p^2 = .09$. Pairwise comparisons indicated that the 4-5 year-olds distinguished between accidental and attempted harms less clearly than the 5-6 year-olds, $p = .04$, and adults, $p = .01$. Accidental harms were considered less punishable than attempted harms by all four age-groups, $F \geq 7.95$, ps $\leq .011$, η_p^2s $\geq .31$. The main effect of age-group did not approach significance.

Justifications. The 4-5-year-olds' (binomial $p = .052$) and all three other age-groups' (ps $< .01$) justifications were based more on intention than on outcome (Table 5). This distinction was smaller for the 4-5-year-olds than for the other age groups, $\chi^2(1)$s ≥ 7.7, ps $< .01$.

Parental knowledge. Participants gave a total of 85 outcome-based punishment judgments, of which 65 (76.5%) were changed to intention-based in response to the parental knowledge question (Table 6). This proportion was similar for all age-groups (4-5 years: 68.4%; adults: 88.8%).

Discussion

Experiment 2 replicated Experiment 1 except that participants were told explicitly that agents were careful, shown additional pictures of them being careful, and then asked an extra question about whether the agents were careful. When this question was answered incorrectly, participants were reminded that the agents were careful.

The first prediction – that acceptability and punishment judgments would be primarily intention-based – was supported. The second prediction was also supported: the 4-5 year-olds considered accidental harms more acceptable and less punishable than attempted harms. Moreover, from 5-6 years, children's punishment judgments were based as much on intention as were adults'. However, there remained substantial differences between the age-groups' acceptability judgments, with adults' being more intention-based than all three children's age-groups'. Together, these findings replicate those of Experiment 1, and also suggest that telling children that agents are careful leads to increases in intention-based punishment judgment.

The third prediction, concerning justifications, was not supported. The proportion of justifications that referred to negligence remained very low in this experiment, despite carefulness being stated explicitly. It is possible that, since participants were corrected when they gave the wrong answer to the carefulness confirmation question, nearly all considered the agents to be careful when they made their judgments, and so did not consider negligence to be relevant to their judgments. An alternative explanation is that, even when negligence influences

judgments, this influence is rarely made explicit in the justifications of these judgments.

Responses to the parental knowledge question corroborated the findings of Nobes et al. (2016) and Experiment 1: at all ages a large majority of apparently outcome-based judgments were changed to intention-based responses when participants were informed that parents knew of the agents' intentions. The fourth prediction was therefore supported.

Despite having just been told that the agents were careful, about a third of the 4-6-year-olds' responses to the care confirmation question were that the agents were careless. Since this question has not been asked in previous studies it is not possible to know whether their participants also considered agents to be negligent, but it seems likely that they did, especially as, with very few exceptions (e.g., Nobes et al., 2009; Schleifer, Shultz, & Lefebvre-Pinard, 1983; Shultz, Wright & Schleifer, 1986), participants were not told that agents were careful. That is, in the absence of carefulness information, it seems likely that participants were even more likely to assume that the agents were negligent. However, it is also possible that the explicit statement and depiction of carefulness merely drew the agents' attention to the agents' negligence but, for some reason (e.g., young children's inability to inhibit intuitive responses), this did not lead them to consider them to be careful.

The findings that children were considerably more likely to consider the Helwig et al. agents (who gave gift boxes containing the wrong animals) to be careless than the Zelazo et al. agents (who accidentally stroked or hit their pets), and that the source of story (Helwig et al. and Zelazo et al.) was not associated with judgments together indicate that perceived extra carelessness does not influence moral judgments. This would seem to be inconsistent with the negligence account, but respondents who said the agents were careless were corrected before they made their judgments, and it is possible that many were persuaded that agents were careful, after all.

Considered individually, the findings of these experiments cannot reveal the separate and combined influence of the two changes – increasing the salience and recency of intention

information, and adding carefulness information – on moral judgments. To do so, it is necessary to compare judgments when neither, one, and both of these changes were made. These comparisons were made in the next stage of this study.

Comparison of data between studies and experiments

To investigate the influence of increasing the salience and recency of intention information, the findings of Experiment 1 were compared with those of Nobes et al. (2016). These experiments were identical except that different participants were tested, and in Experiment 1 intention information was more salient and recent.

The influence of carefulness (i.e., the absence of negligence) on moral judgments was investigated by comparing the findings of Experiment 1 with those of Experiment 2. Apart from other participants being tested, these differed only in that in Experiment 2 participants were told explicitly that, and how, the agents were careful, that is, they were not negligent.

By comparing the findings of Experiment 2 with those of Nobes et al. (2016) we also investigated the influence of the combination of more salient and recent intention and inclusion of carefulness information.

Measures of the relative extent to which each participant in each of the three experiments judged according to intention or outcome – the difference scores – were obtained by calculating, separately for acceptability and punishment, the difference between their mean ratings of accidental and attempted harms. ANOVAs were run on the data as above except that experiment was included as an additional factor. Higher difference scores indicated more intention-based judgment.

Results

Acceptability judgments. Figure 5 shows the mean acceptability difference scores (accidental harm scores minus attempted harm scores) in the three experiments. A 3 (Age-group) x 2 (Action valence [accidental harm, attempted harm]) x 3 (Experiment [Nobes et al., 2016, Experiment 1, Experiment 2]) mixed ANOVA with repeated measures on action valence

indicated a main effect of action valence on acceptability judgments, $F(1,298) = 343.08$, $p<.001$, $\eta_p^2 = .54$, and a main effect of age-group, $F(3,298) = 2.77$, $p = .04$: the youngest children's judgements were slightly harsher than those of the other three age-groups', $ps < .04$. There was also an interaction between action valence and age-group, $F(3,298) = 31.15$, $p<.001$, $\eta_p^2 = .24$. Participants of all four age-groups rated the well-intentioned accidental harms more acceptable than ill-intentioned attempted harms, $Fs \geq 9.41$, $ps \leq .003$, $\eta_p^2s > .10$. Pairwise comparisons showed that adults made this distinction more clearly than all age-groups of children, $ps < .001$; the 7-8-year-olds more clearly than the 5-6-year-olds, $p = .03$, and the 4-5-year-olds, $p < .001$; and the 5-6-year-olds marginally more clearly than the 4-5-year-olds, $p = .06$. There were no main or interaction effects of experiment. This was also the case when accidental and attempted harms were analysed separately.

A post hoc power analysis was conducted to test whether the non-significant effect of experiment could be attributed to a lack of statistical power. This indicated that, with power $(1 - \beta)$ set at 0.80, $\alpha = .05$, 2-tailed, and observed $\eta_p^2 = .003$, would have required more than 10 times the sample size ($N = 3205$) across all three experiments to reach statistical significance. It is therefore very unlikely that this null result resulted from limited sample size.

Punishment judgments. The mean punishment difference scores (attempted harm scores minus accidental harm scores) for the three experiments are shown in Figure 6. The equivalent ANOVA as for acceptability ratings was run on punishment ratings and showed a main effect of action valence, $F(1,298) = 153.31$, $p < .001$, $\eta_p^2 = .34$. There was also an interaction between action valence and age-group, $F(3,298) = 10.48$, $p < .001$, $\eta_p^2 = .10$. Pairwise comparisons indicated that adults distinguished more clearly than all three age groups of children, $ps \leq .005$, and the 4-5-year-olds less clearly than all three older age-groups $ps \leq .02$). All three older age-groups – but not the youngest group – judged accidental harms less punishable than attempted harms, $Fs \geq 21.48$, $ps < .001$, $\eta_p^2s > .20$. There was also an interaction between action valence and experiment, $F(2, 298) = 9.84$, $p < .001$, $\eta_p^2 = .06$. Participants in Experiment 2 (mean

difference score = .78) distinguished between accidental harms and attempted harms marginally more clearly than did participants in Experiment 1, $M = .58$, $p = .107$, and significantly more than those in Nobes et al. (2016), $M = .33$, $p < .001$, and this distinction was made more clearly in Experiment 1 than in Nobes et al. (2016), $p = .022$. The main effects of age-group and experiment, the interaction between them, and their 3-way interaction with action valence, did not approach significance.

Separate analyses showed that punishment judgments of both accidental harms, $F(3, 308) = 3.14$, $p = .045$, $\eta_p^2 = .02$, and attempted harms, $F(3, 306) = 5.34$, $p = .005$, $\eta_p^2 = .034$, differed according to experiment. Pairwise comparisons indicated that, compared with Nobes et al. (2016), participants in Experiment 2 judged accidental harms to be less punishable, $p = .013$, and attempted harms to be more punishable, $p = .002$. In addition, attempted harms were considered marginally more punishable in Experiment 2 than in Experiment 1, $p = .06$.

Discussion

Comparison of data between studies indicated that increasing the salience and recency of intention information resulted in punishment judgments, but not acceptability judgments, being more intention-based. The absence of an interaction between experiment and age-group indicates that this effect was general to all age-groups, although, while in Nobes et al. (2016) the youngest children's acceptability and punishment judgments were approximately equally intention- and outcome-based, in Experiment 1 these children's judgments were based more on intention than on outcome.

Comparisons between the experiments reported here indicated that, although adding carefulness information did not influence acceptability judgments, punishment judgments were marginally more intention-based in Experiment 2 than in Experiment 1.

The combination of increased intention salience and added carefulness information resulted in punishment judgments, but not acceptability judgments, being more intention-based in Experiment 3 than in Nobes et al. (2016). Comparisons of mean difference scores suggests

that both changes contributed approximately equally to the increase in intention-based judgment.

General discussion

Two experiments were conducted to examine possible reasons for children's and adults' apparently outcome-based moral judgments. Participants were told four stories from two studies (Helwig et al., 2001; Zelazo et al., 1996), both of which strongly supported the claim that children's moral judgments are primarily outcome-based. Nobes et al. (2016) replicated these studies, and found that changing the wording of the acceptability question resulted in substantially more intention-based judgment at all ages. However, some judgments were still based on outcomes: in particular, young children persisted in judging approximately as much according to outcome as to intention. The present study investigated possible reasons for this persisting outcome-based judgment with the aim of helping to explain the findings not only of the original studies, but also those of the many other studies in this area that have used similar methods and reported primarily outcome-based judgment by children.

In the first experiment, the Nobes et al. study was replicated except that the salience and recency of intention information were increased by asking – and if necessary correcting – participants about agents' intentions directly before they judged the agents. When participants' awareness and understanding of the agents' intentions was raised in this way, they judged accidental harms (positive intentions, negative outcomes) to be more acceptable and less punishable than attempted harms (negative intention, positive outcome): that is, they judged primarily according to intention. Even the youngest children's (4-5 years) acceptability judgments were based significantly, and punishment judgments marginally, more on intentions than outcomes. However, many of the younger children continued to judge according to outcomes, and there remained a marked increase with age in the extent to which both types of judgment were intention-based.

The second experiment investigated whether some or all of this still-persisting outcome-

based judgment could be attributed to participants assuming that well-intentioned agents who accidentally caused harm were blameworthy because they were negligent, rather than because of the outcome per se. Experiment 1 was replicated except that participants were also told that, and how, agents were careful, that is, they were not negligent. The results of Experiment 1 were replicated, except that the 4-5-year-olds' punishment judgments (as well as their acceptability judgments) were based significantly more on intentions than on outcomes. Moreover, from 5-6 years children's punishment judgments were as intention-based as were adults'. However, many of the 4-5 year-olds' judgments, and even some of the older children's acceptability judgments, remained based on outcome rather than intention.

Data from the two experiments were then pooled with those from Nobes et al. (2016) to investigate the independent and combined influences of, first, increasing the salience and recency of intention information (by comparing the results of Nobes et al. with those of Experiment 1); second, adding carefulness information (by comparing data from Experiment 1 and Experiment 2); and, third, both these changes (by comparing data from Nobes et al. with Experiment 2).

Regarding acceptability judgments, there were no differences at any age between the three experiments. Although this indicates that neither factor (intention salience and negligence) has a discernible impact on acceptability judgments, in Experiments 1 and 2 (but not Nobes et al., 2016) even 4-5 year-olds considered accidental harms to be more acceptable than attempted harms, that is, they based these judgments more on intention than on outcome. This suggests that the effect of increasing the salience of intention information was large enough to raise young children's judgments "above the bar" of intention-based acceptability judgment (i.e., they judged accidental harms significantly more acceptable than attempted harms), but was too small to reach significance when studies were compared.

In contrast, punishment judgments were more intention-based when intentions were more salient and recent, and still more intention-based when negligence information was added. The

combination of these changes resulted in 4-5 year-olds' punishment judgments being based significantly more on intentions than on outcomes, and 5-8 year-old's punishment judgments being as intention-based as were adults'.

The picture that emerges from these experiments is that acceptability judgments are hardly influenced, if at all, by either of the two factors investigated here. In contrast, both changes resulted in more intention-based punishment judgments, and the combination of both resulted in substantially more intention-based punishment judgments. Surprisingly, the changes were equally influential across age-groups, suggesting that the increase with age in intention-based punishment judgment is not attributable to youngsters being more influenced by the salience of outcome information, or by assumptions of negligence, than older participants. Rather, participants at all ages were equally likely to be influenced by these factors.

These factors therefore explain many outcome-based punishment judgments. However, they do not refute the "weak" form of the outcome-to-intent shift discussed above, according to which, although children's judgments are not primarily influenced by outcomes, they are more inclined to judge according to outcome than are adults. To this limited extent, then, the findings of this study are consistent with accounts of moral development such as Piaget's (1932 / 1965) and Cushman et al.'s (2013), according to which children's moral judgments are fundamentally different from adults'.

However, the data reported here suggest other possible reasons for some, and possibly all, of the outcome-based judgment by children that persisted even in Experiment 2. First, some young children misunderstood or forgot the agents' intentions so that their apparently outcome-based judgments might actually have been based on misinterpreted intentions. In particular, despite having been reminded of, and if necessary corrected on, the agents' intentions directly before the judgments, in Experiment 1 about a quarter of 4-5 year-olds and 10% of 5-6 year-olds who gave outcome-based punishment judgments said directly afterwards that the well-intentioned agents were ill-intentioned, or vice versa. Moreover, in Experiment 2

approximately a third of children considered the agents negligent, despite having just been told that they were careful. It is likely that many of these children based their judgments on misinterpretations of agents' intentions, or assumptions that the agents were negligent, and therefore punishable. The implication is that their intention- or negligence-based responses would have given the impression that their judgments were outcome-based. These factors are even more likely to have influenced judgments in previous studies (including Helwig et al., Zelazo et al., and Nobes et al., 2016) since the relevant confirmation questions were not asked and, of course, erroneous responses were not corrected.

A second possible reason for the remaining outcome-based responses was revealed by the parental knowledge question. When participants made outcome-based judgments, the majority in all age-groups gave intention-based responses to the parental knowledge question. This suggests that many outcome-based judgments occur because participants assume that punishers did not know about the agents' intentions. Indeed, in reality parents and other authorities often do not know what was intended, and so can only infer intentions – rightly or wrongly – from the outcomes of actions. If this interpretation is correct, it indicates that even the findings of this experiment substantially underestimate the true incidence of intention-based reasoning. However, this finding must be treated with caution because the reason why parental knowledge questions elicit intention-based responses might be that they remind the participants of agents' intentions, or emphasize their importance. But whatever its explanation, this finding shows that even those participants who make outcome-based judgments are sensitive to intentions, and are able to judge according to them.

The unexpected finding that acceptability judgments were not influenced by either factor, while punishment judgments were influenced by both, suggests that these two forms of judgment are affected by different factors and are driven by different processes. Cushman et al. (2013) have proposed that punishment judgments are influenced by both the early-developing causal process and the later-developing mental-state process, while acceptability judgments are

influenced only by the latter. They would therefore predict that acceptability judgments become intention-based earlier in development than punishment judgments, but our findings indicate the opposite.

However, a possible reason for negligence information influencing punishment – but not acceptability – judgments might be that punishment judgments are associated with perceptions of causality, as Cushman and colleagues suggest. When participants are told that agents were careful, this might attenuate their assumption that accidental harm-doers caused the accident, which would reduce their tendency to consider them punishable; conversely, it might reinforce their view that attempted harm-doers are responsible, in which case these agents would be deemed more punishable. The result would be as observed, namely that punishment judgments, but not acceptability judgments, become more intention-based when negligence information is added.

An alternative reason why increased intention salience and added negligence information might influence punishment but not acceptability judgments might lie in the different rating scales. Acceptability was measured on a bidirectional 5-point scale from *very, very bad* to *very, very good*, with *okay* at the centre. This might have invited neutral *okay* responses that required relatively little consideration of intention, negligence and outcome. In contrast, the punishment scale was unidirectional, from *no trouble* to *lots of trouble*. Since there was no neutral response, participants might have had to make more considered judgments, in which case the increased salience of intention and the addition of negligence information might have had more impact. This possibility could be tested by measuring response latencies (Imamoğlu, 1975), or by removing *okay* in the acceptability scale so that participants were forced to make either positive or negative judgments.

A limitation of this study is that there were sampling differences within and between experiments. In particular, children from different schools were interviewed, and it is likely that this accounted for some of the variance in judgments reported here. For example, many of

the young children in Experiment 2 attended small infant schools in relatively well-off villages, while the 7-8 year-olds were at larger junior schools, one of which was in a less affluent urban area. These differences might account for the relatively poor comprehension of the 7-8-year-olds, which could have impacted their moral judgments. In addition, testing conditions differed between schools, not least in terms of background noise, which is likely to have affected some children's concentration. Buon, Jacob, Loissel and Dupoux (2013) report that even adults' moral judgments become outcome-based under cognitive load, and the noise and distractions of everyday school life might have a similar effect on children.

Another limitation is that, while participants were asked about intentions after their judgments, they were not asked about outcomes. It is possible that these were sometimes forgotten or misunderstood, too. As Feldman et al. (1976) point out, "Without memory checks, it is not possible to assess what information the subject was in fact using when the judgment was made. Thus, preference for intent- or consequence-based judgments without recall data does not necessarily reflect the subjects' awareness of the information." (p. 559). Future researchers are encouraged to include confirmation questions about intention and outcome both before and after judgments, or to ask respondents to retell the story (Nelson, 1980).

Although the advantages of replications have been stressed, this study also illustrates some disadvantages. In particular, since it was important to replicate Helwig et al.'s and Zelazo et al.'s studies as closely as possible, it was not possible to simplify their stories to aid young children's comprehension. For example, the agents in the Helwig et al. stories gave the wrong animals as presents because they had false beliefs about the contents of the gift boxes. In effect, then, these stories included a "deceptive box" theory of mind task. Since large proportions of 4-year-olds fail such tasks (Perner, Leekham, & Wimmer, 1987; Wellman & Liu, 2004) it is likely that many young children in this study were confused about the agents' intentions, in which case their apparently outcome-based judgments would actually have been based on misunderstood intentions.

A further limitation was revealed by the high proportions of younger children who, despite the increased salience and recency of intention information, continued to misunderstand or forget the agents' intentions. For these participants, then, this manipulation appears to have lacked effectiveness. Given the efficacy of similar changes in previous research (e.g., Bearison & Isaacs, 1975; Feldman et al., 1976; Nelson et al., 1980; Nummedal & Bass, 1976), this is surprising, and suggests that failure to understand or recall occurred even more frequently in the large majority of previous research because outcomes are usually more salient and recent than intentions. However, Gvozdic et al. (2016) propose that much of children's outcome-based judgment occurs because of the related issue of failure to inhibit automatic responses to outcomes. This would explain the effectiveness of their metacognitive training approach relative to ours, and suggests that instructing children to "not focus too much" on consequences would have resulted in higher levels of intention-based judgment by children.

Much the same point applies to the addition of negligence information in Experiment 2, since many young children and even some adults continued to consider the agents to be careless. A similar approach to that used here (i.e., telling participants that, and how, agents were careful) was shown to be effective by Nobes et al. (2009), but the present findings indicate that some participants' persisting assumptions of negligence might have continued to result in apparent outcome-based judgments both in that study and here.

Our finding that 4-5 year-olds often misunderstood the stories led to our decision not to include even younger children. Had we used different stories that did not cause these problems, we would have included 3-year-olds. If they were found to judge primarily according to outcome, this would have provided support for an outcome-to-intent shift, albeit considerably earlier than any researchers (to our knowledge) have proposed. Replications of other studies that used less challenging stories are required to investigate this possibility.

Another issue for future research is to determine why the parental knowledge question resulted in so many participants effectively changing from outcome-based judgments to

intention-based responses. This might be done by telling participants *before* they judged that potential punishers understood agents' intentions. If this resulted in more intention-based judgment, then this would indicate the extent to which apparently outcome-based judgments in this and other studies were based on the assumption that punishers could not have known the agents' true intentions.

While our findings indicate that even young children's judgments are primarily intention-based, it is also clear that there are some wide disparities between the judgments of children of the same age: a small number persisted in making outcome-based judgments even in Experiment 2. The reasons for these intriguing individual differences have received very little attention from researchers, and remain poorly understood. It is possible that such children are severely disadvantaged in their social interactions because they fail to understand others' intentions, or are unaware of the significance of intentions when evaluating their own and others' actions. Indeed, an important precursor of aggressive behavior is hostile attribution bias, that is, the tendency to misattribute hostility to others' benign intentions (Arsenio, Adams, & Gold, 2009; Crick & Dodge, 1994). Future research should investigate the potential for measures of intention-based judgment – such as those used here and in similar studies – to identify individuals who are at risk for these misattributions and the antisocial behavior that results. These measures might also be adapted to develop interventions aimed at preventing or mitigating these misattributions, and hence reducing aggressive behavior.

In summary, the findings of these experiments indicate that, despite using very similar methods to Helwig et al. (2001) and Zelazo et al. (1996) – both of which reported primarily outcome-based moral judgment at all ages – participants' acceptability and punishment judgments were primarily intention-based. They therefore corroborate those of Nobes et al. (2016) and extend them in several ways: in particular, they indicate that even 4-5 year-olds' judgments are primarily intention-based, and that by 5-6 years children's punishment judgments are as intention-based as are adults'. The evidence presented here also indicates that,

in addition to Nobes et al.'s finding that high proportions of outcome-based judgments resulted from the wording of the acceptability question, some outcome-based punishment (but not acceptability) judgments in the original studies resulted from intention information being insufficiently salient, and from participants sometimes assuming that negative outcomes were caused by negligence. These findings shed light on the reasons for apparently outcome-based judgment not only in Helwig et al. and Zelazo et al., but also in the large majority of other studies in this area because they, too, gave no information about negligence, and included outcomes that were more salient than intentions.

These findings are inconsistent with the strong form of the outcome-to-intention shift espoused originally by Piaget (1932 / 1965) and subsequently corroborated by most researchers in this area, according to which young children's moral judgments are primarily outcome-based. However, even in Experiment 2 a large minority of the youngest children's judgments remained outcome-based, as did some older children's acceptability judgments. Neither intention salience nor negligence information either independently or in combination can account for this persisting outcome-based judgment. Moreover, these factors do not explain the differences between age-groups, since young children's judgements were influenced no more (nor less) than other age-groups by either. The findings are therefore consistent with the weak form of the outcome-to-intent shift, according to which, while not being primarily outcome-based, children's judgments are more outcome-based than are adults'.

But the findings presented here suggest other possible reasons for outcome-based judgments. In particular, they reveal a high level of misunderstanding by children of the key elements of the stories, perhaps because they forgot, or failed to integrate information about intentions, negligence and outcomes. Their apparently outcome-based judgments could therefore actually have been based on incorrect beliefs about the agents' intentions or level of negligence. In addition, when asked the parental knowledge question, most participants at all ages who made outcome-based judgments gave intention-based responses, which shows that

almost all children are at least sensitive to intention information, and are capable of basing their moral judgments on it. These, and perhaps other possible reasons for their higher rates of outcome-based judgments must be tested before we can be sure about the extent to which children's judgments are actually based on intention and outcome.

References

Arsenio, W. F., Adams, E., & Gold, J. (2009). Social information processing, moral reasoning, and emotion attributions: relations with adolescents' reactive and proactive aggression. *Child Development, 80*(6), 1739-1755. doi:10.1111/j.1467-8624.2009.01365.x

Baird, J. A., & Astington, J. W. (2004). The role of mental state understanding in the development of moral cognition and moral action. *New Directions for Child and Adolescent Development, 103,* 37–49. doi:10.1002/cd.96

Bearison, D., & Isaacs, L. (1975). Production deficiency in children's moral judgments. *Developmental Psychology, 11,* 732-737. doi:10.1037/0012-1649.11.6.732

Buchanan, J. P., & Thompson, S. K. (1973). A quantitative methodology to examine the development of moral judgment. *Child Development, 44,* 186–189. doi.org/10.2307/1127700

Buon, M., Jacob, P., Loissel, E., & Dupoux, E. (2013). A non-mentalistic cause-based heuristic in human social evaluations. *Cognition, 126*(2), 149-155. doi:10.1016/j.cognition.2012.09.006

Buon, M., Seara-Cardoso, A., & Viding, E. (2016). Why (and how) should we study the interplay between emotional arousal, Theory of Mind, and inhibitory control to understand moral cognition? *Psychonomic Bulletin & Review*, 1-21. doi:10.3758/s13423-016-1042-5

Chandler, M. J., Greenspan, S., & Barenboim, C. (1973). Judgments of intentionality in response to video-taped and verbally presented moral dilemmas: The medium is the message. *Child Development, 44*, 315–320. doi:10.2307/1128053

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Crick, N. R., & Dodge, K. A. (1994). A review and reformulation of social information-processing mechanisms in children's social adjustment. *Psychological Bulletin, 115*(1),

74. doi:10.1037/0033-2909.115.1.74

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

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

Duncan, G. J., Engel, M., Claessens, A., & Dowsett, C. J. (2014). Replication and robustness in developmental research. *Developmental Psychology*, *50*(11), 2417. doi:10.1037/a0037996

Elkind, D., & Dabek, R. F. (1977). Personal injury and property damage in moral judgments of children. *Child Development, 48*, 518-522. doi:10.2307/1128648

Farnill, D. (1974). The effects of social-judgment set on children's use of intent information. *Journal of Personality, 42*, 276–289. doi:10.1111/j.1467-6494.1974.tb00674.x

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*, 175-191. doi:10.3758/BF03193146

Feldman, N. S., Klosson, E. C., Parsons, J. E., Rholes, W. S., & Ruble, D. N. (1976). Order of information presentation and children's moral judgments. *Child Development, 47*, 1203-1215. doi:10.2307/1128821

Gummerum, M., & Chu, M. T. (2014). Outcomes and intentions in children's, adolescents', and adults' second-and third-party punishment behavior. *Cognition*, *133*(1), 97-103. doi:10.1016/j.cognition.2014.06.001

Gvozdic, K., Moutier, S., Dupoux, E., & Buon, M. (2016). Priming children's use of intentions in moral judgement with metacognitive training. *Frontiers in Psychology, 7*, 190. http://doi:10.3389/fpsyg.2016.00190

Hamlin, J. K. (2013). Failed attempts to help and harm: Intention versus outcome in preverbal

infants' social evaluations. *Cognition, 128,* 451-474. doi:10.1016/j.cognition.2013.04.004

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

Imamoğlu, E. O. (1975). Children's awareness and usage of intention cues. *Child Development, 46,* 39–45. doi:10.2307/1128831

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

Margoni , F., & Surian, L., (2016). Explaining the U-shaped development of intent-based moral judgments. *Frontiers in Psychology, 7,* 219, 1-6. doi:10.3389/fpsyg.2016.00219

Nelson, S. A. (1980). Factors influencing young children's use of motives and outcomes as moral criteria. *Child Development, 51,* 823-829. doi:10.2307/1129470

Nobes, G., Panagiotaki, G., & Bartholomew, K. J. (2016). The influence of intention, outcome and question-wording on children's and adults' moral judgments. *Cognition*, *157*, 190-204. http://dx.doi.org/10.1016/j.cognition.2016.08.019

Nobes, G., Panagiotaki, G., & Pawson, C. (2009). The influence of negligence, intention and outcome on children's moral judgments. *Journal of Experimental Child Psychology, 104*, 382–397. doi:10.1016/j.jecp.2009.08.001

Nummedal, S.G., & Bass, S.C. (1976). Effects of the salience of intention and consequence on children's moral judgments. *Developmental Psychology, 12,* 475-476. doi:10.1037/0012-1649.12.5.475

Nuñez, N., Laurent, S., & Gray, J. M. (2014). Is negligence a first cousin to intentionality? Lay conceptions of negligence and its relationship to intentionality. *Applied Cognitive Psychology*, *28*(1), 55-65. doi:10.1002/acp.2957

Perner, J., Leekam, S. R., & Wimmer, H. (1987). Three-year-olds' difficulty with false belief:

The case for a conceptual deficit. *British Journal of Developmental Psychology*, *5*(2), 125-137. doi:10.1111/j.2044-835X.1987.tb01048.x

Piaget, J. (1932/1965). *The moral judgment of the child*. Trans. M. Gabain. New York: Free Press.

Schleifer, M., Shultz, T. R., & Lefebvre-Pinard, M. (1983). Children's judgements of causality, responsibility and punishment in cases of harm due to omission. *British Journal of Developmental Psychology*, *1*(1), 87-97. doi:10.1111/j.2044-835X.1983.tb00546.x

Shultz, T. R., Wright, K., & Schleifer, M. (1986). Assignment of moral responsibility and punishment. *Child Development*, 177-184. URL: http://www.jstor.org/stable/1130649

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

Walden, T. A. (1982). Mediation and production deficiencies in children's judgments of morality. *Journal of Experimental Child Psychology*, *33*, 165–181. doi:10.1016/0022-0965(82)90012-1

Wellman, H. M., & Liu, D. (2004). Scaling of Theory-of-Mind tasks. *Child Development*, *75*(2), 523-541. doi:10.1111/j.1467-8624.2004.00691.x

Yuill, N. (1984). Young children's coordination of motive and outcome in judgments of satisfaction and morality. *British Journal of Developmental Psychology, 2*, 73–81. doi:10.1111/j.2044-835X.1984.tb00536.x

Zelazo, P. D., Helwig, C. C., & Lau, A. (1996). Intention, act, and outcome in behavioral prediction and moral judgment. *Child Development*, *67*, 2478–2492. doi:10.2307/1131635

Table 1

Percentages of correct predictions, confirmations and justifications (Experiment 1). Example questions in parentheses.

	4-5 years	5-6 years	7-8 years	Adults
Predictions				
Behavior ("What is Kevin going to get Rob?")	88.9	91.7	96.3	100
Emotion ("How did Rob feel when he got the puppy?")	94.9	97.8	100	100
Confirmations				
Outcome ("Did Kevin give Rob a puppy or a spider?")	82.0	86.1	100	100
Cause ("Why did Kevin give Rob a puppy?")	91.0	94.4	100	100
Intention ("What did Kevin try to get Rob?")	88.0	100	100	100
Intention post-judgment ("What did Kevin want to do to Rob?")*	75.6	89.7	95.7	100
Justification**	86.0	94.4	98.8	100

* Post-judgment intention questions were asked only when judgments were outcome-based.

** A correct justification is factually correct, regardless of whether it is based on intention, outcome, etc.

Table 2

Percentages of punishment justifications based on intention, negligence and outcome, by age-group (Experiment 1)

Justification basis	4-5 years	5-6 years	7-8 years	Adults	Total
Intention	40.7	57.7	64.9	86.8	62.5
Negligence	0	0	0	4.2	1.0
Outcome	29.6	25.9	24.3	6.8	21.7
Other / DK	29.7	16.4	10.9	2.3	14.8

Table 3

Frequencies of outcome-based punishment judgments, and frequencies of intention-based responses to the subsequent parental knowledge question, by age-group in Experiment 1. (The parental knowledge question was asked only when the punishment judgments were outcome-based.)

	4-5 years	5-6 years	7-8 years	Adults	Total
Outcome-based punishment judgments	38	26	21	5	90
Intention-based parental knowledge response	26	17	17	4	64

Table 4

Percentages of correct predictions, confirmations and justifications (Experiment 2)

	4-5 years	5-6 years	7-8 years	Adults
Predictions				
Behavior	84.9	98.1	94.9	100
Emotion	96.0	98.3	97.5	100
Confirmations				
Care	60.8	67.0	71.8	89.9
Outcome	84.6	96.4	91.3	100
Cause	92.3	98.0	98.8	100
Intention	96.0	98.0	97.5	100
Justification*	87.1	90.0	98.7	100

* A correct justification is factually correct, regardless of whether it is based on intention, outcome, etc.

Table 5

Percentages of punishment justifications based on intention, negligence and outcome, by age-group and action valence (Experiment 2)

Justification basis	4-5 years	5-6 years	7-8 years	Adults	Total
Intention	32.7	54.1	78.4	83.7	62.2
Negligence	0.8	3.3	0.0	0.0	1.0
Outcome	20.8	12.1	15.7	10.9	14.9
Other	45.8	30.7	6.0	5.4	22.0

Table 6

Frequencies of outcome-based punishment judgments, and of intention-based responses to the subsequent parental knowledge question, by age-group and action valence in Experiment 2. (The parental knowledge question was asked only when the punishment judgments were outcome-based.)

	4-5 years	5-6 years	7-8 years	Adults	Total
Outcome-based punishment judgments	38	20	18	9	85
Intention-based ‘parental knowledge’ response	26	16	15	8	65

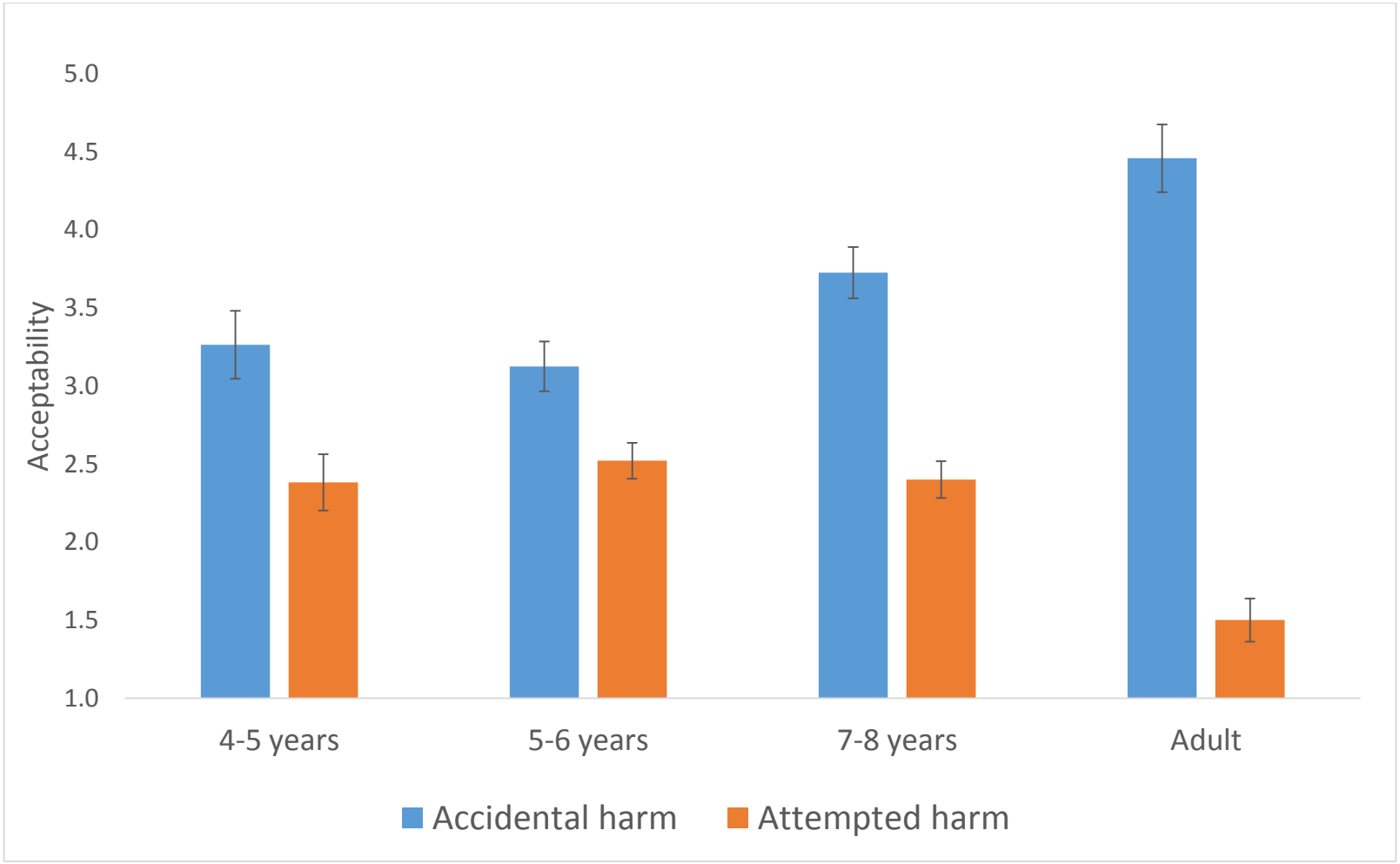

Figure 1. Mean (+SE) acceptability ratings (1 = really, really bad; 5 = really, really good) of accidental and attempted harms by age group when intention salience and recency are increased (Experiment 1).

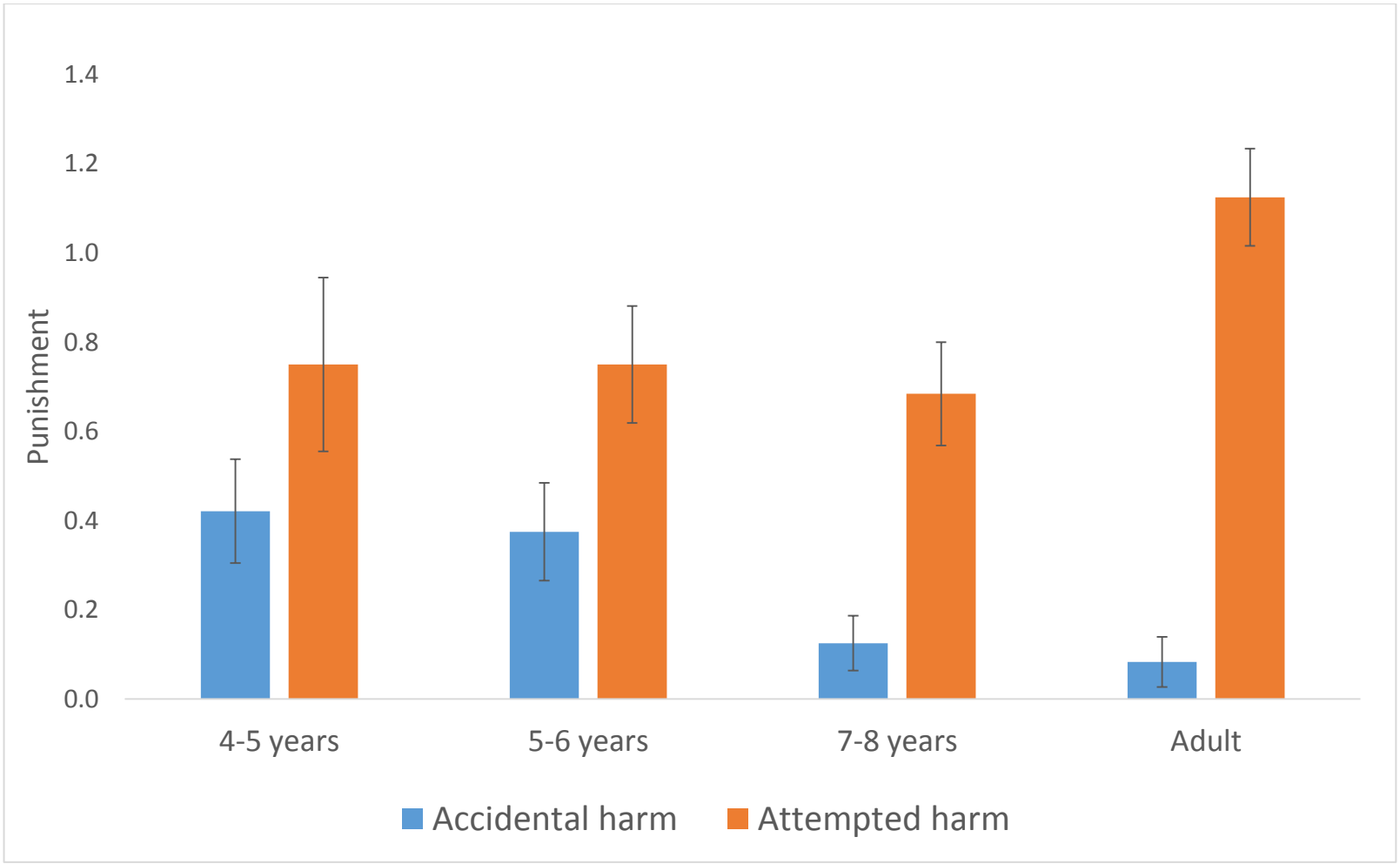

Figure 2. Mean (+SE) punishment ratings (0 = none; 2 = a lot) of accidental and attempted harms by age group when intention salience and recency are increased (Experiment 1).

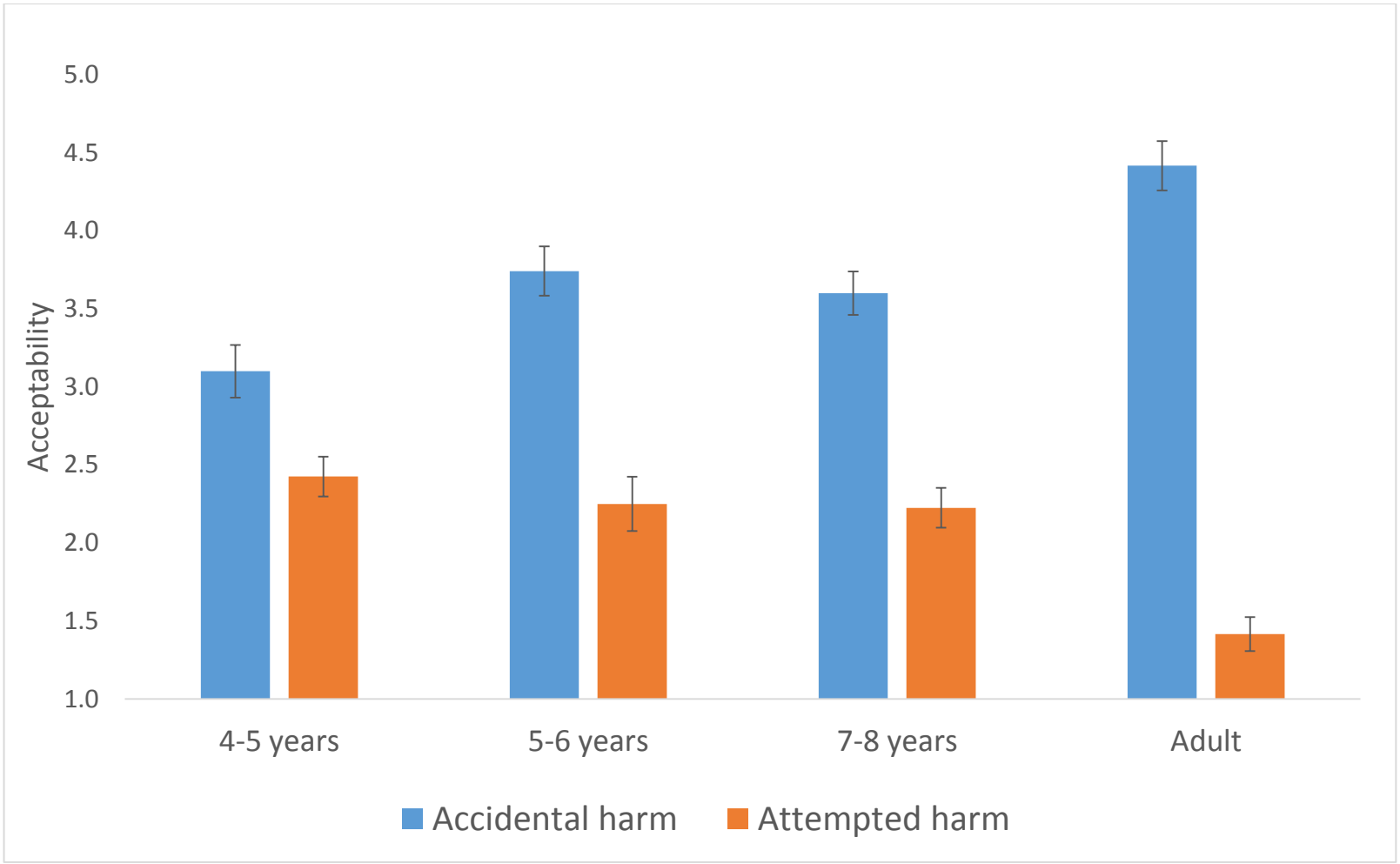

Figure 3. Mean (+SE) acceptability (1 = really, really bad; 5 = really, really good) ratings of accidental and attempted harms by age-group when intention salience is increased and carefulness information added (Experiment 2).

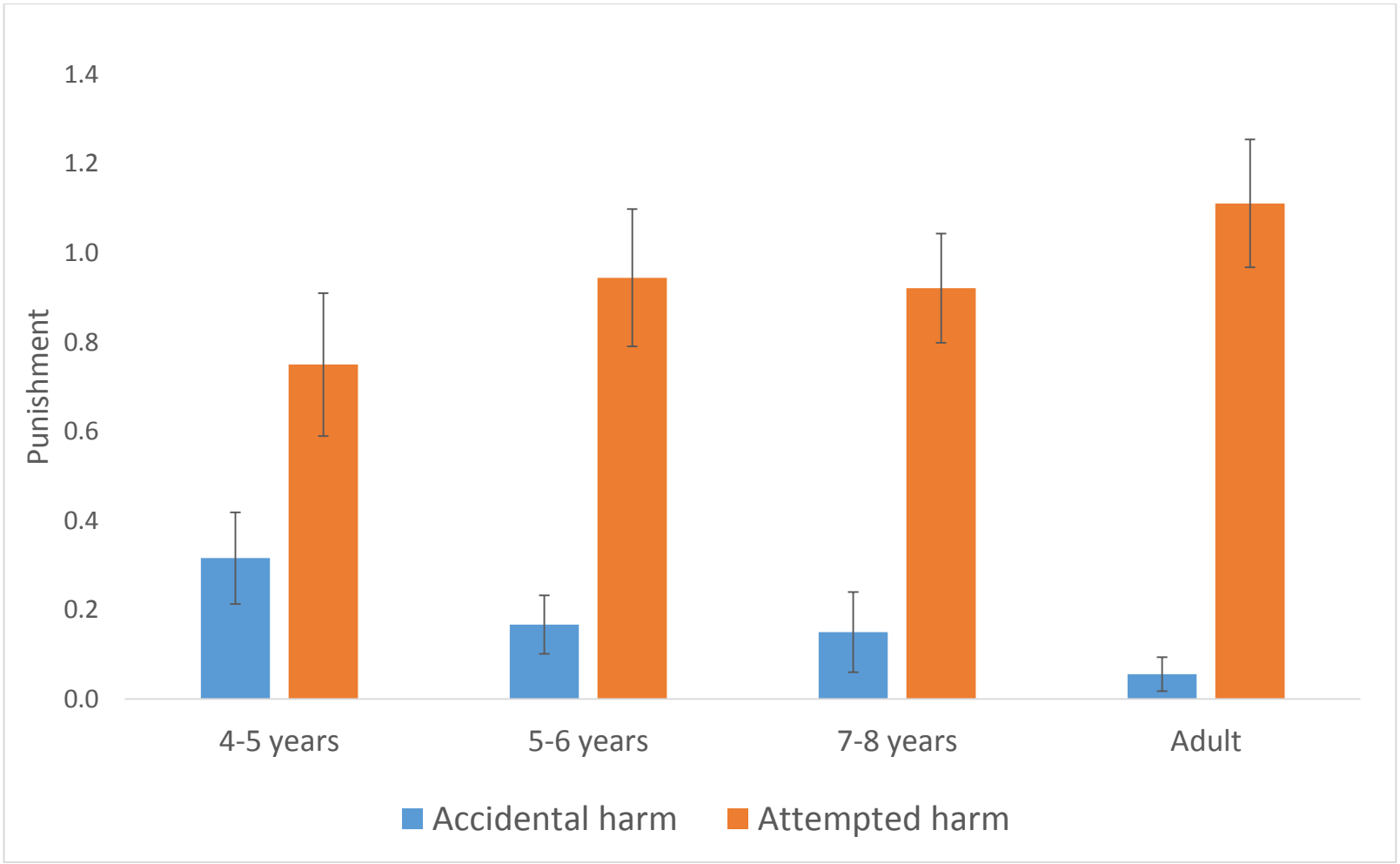

Figure 4. Mean (+SE) punishment (0 = none; 2 = a lot) ratings of accidental and attempted harms by age-group when intention salience is increased and carefulness information added (Experiment 2).

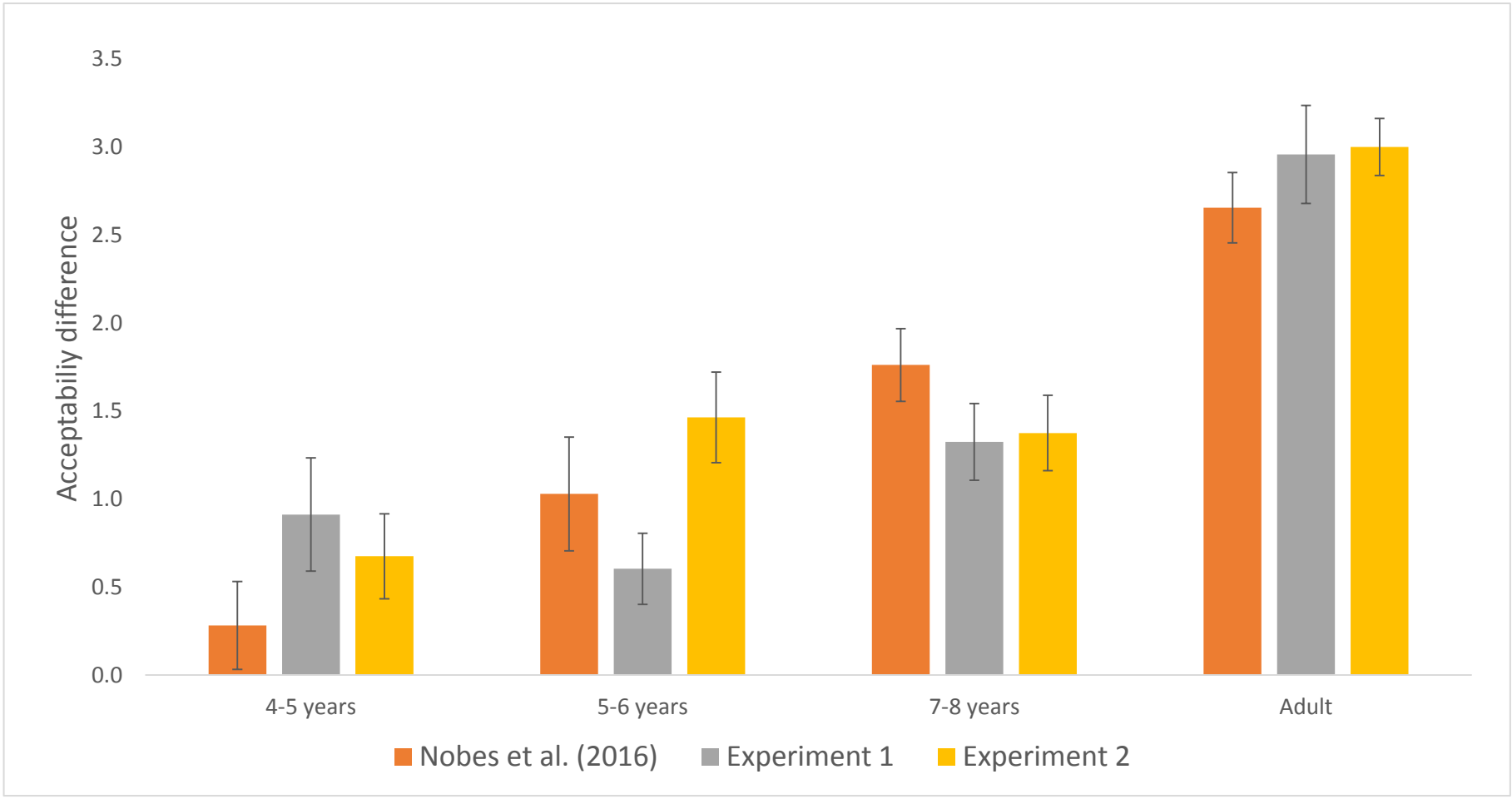

Figure 5. Mean (+SE) differences in acceptability judgments of accidental harms and attempted harms by age group and experiment. Positive scores indicate that accidental harms are considered more acceptable than attempted harms; higher scores indicate more intention-based judgment.

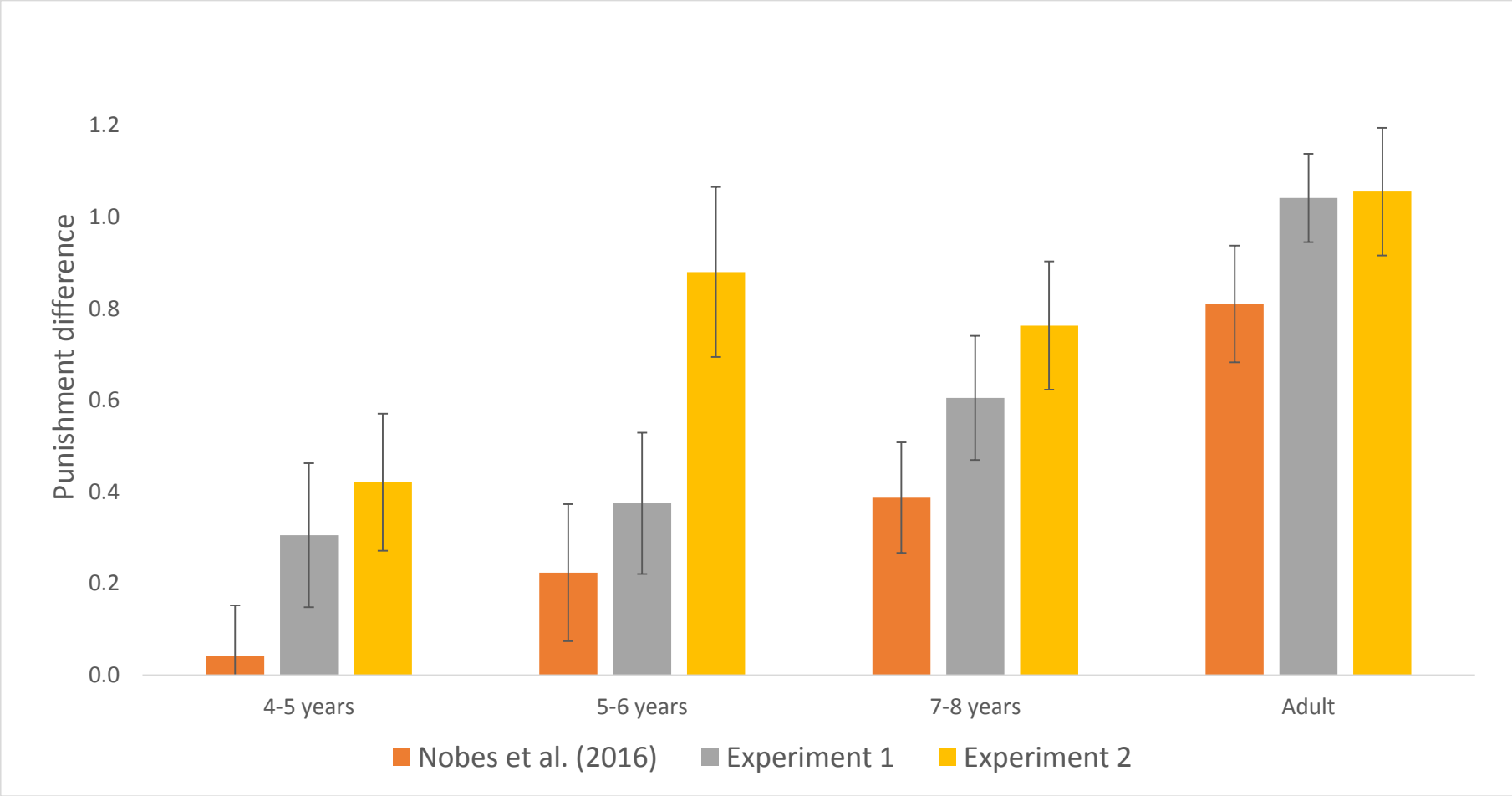

Figure 6. Mean (+SE) differences in punishment judgments of accidental harms and attempted harms by age group and experiment. Positive scores indicate that accidental harms are considered less punishable than attempted harms; higher scores indicate more intention-based judgment.

Appendix: Example interview schedules

1. Accidental harm (positive intention; negative outcome), adapted from Helwig et al. (2001)

Issue / question	Nobes et al. (2016)	Experiment 1	Experiment 2
Preference: Puppies	Here's Ethan. Ethan has a friend named Chris. Chris really likes puppies. He likes to read about them and play with them. When Chris sees puppies, he feels happy because he likes them.		
Comprehension 1: Puppies	How does Chris feel when he sees puppies?		
Preference: Spiders	Chris doesn't like spiders though. When Chris sees big spiders, he is afraid. Big spiders scare Chris. When Chris sees big spiders he is afraid and he cries.		
Comprehension 2: Spiders	How does Chris feel when he sees big spiders?		
Intention	When Chris invited Ethan to his birthday party, Ethan wanted to bring a present that would make Chris happy.		
Confirmation 1: Spiders	Now, how does Chris feel when he sees big spiders?		
Confirmation 2: Puppies	How does he feel when he sees puppies?		
Knowledge	Now, Ethan knows that Chris likes puppies. He knows that Chris is scared and cries when he sees big spiders and is happy and smiles when he sees puppies.		
Behavioral prediction	What is Ethan going to get Chris for his birthday? Is he going to get Chris a puppy or a spider?		
Intention	Well, let me tell you what happened. Ethan wanted to make Chris happy and he knew Chris liked puppies, so Ethan decided to get Chris a puppy for his birthday.		
Care			Ethan went to the pet shop and asked for a puppy. Ethan was very careful to make sure that he got Chris a puppy. He couldn't look in the box because it was very well wrapped up, and so he

			asked the man in the shop if he was *sure* there was a puppy in the box. The man said "Don't worry, there's a puppy in the box".
Cause	But someone at the pet shop made a mistake and put a big spider in the box instead.		But actually the man in the pet shop made a mistake and put a big spider in the box instead.
Outcome - act	So Ethan gave Chris a big spider for his birthday.		
Emotional state prediction	How do you think Chris felt when he got the big spider?		
Outcome - emotion	When Chris got the big spider he was upset. Chris was scared by the spider.		
Confirmation 3: Care			Was Ethan careful to make sure he gave Chris a puppy and not a spider?
Confirmation 4: Outcome		Did Ethan give Chris a puppy or a spider?	
Confirmation 5: Cause (deliberate / accidental)		Why did Ethan give Chris a spider? Did he want to give him the spider, or did he give it to him by mistake?	
Confirmation 6: Intention		What did Ethan *try* to get Chris? Did he try to get him a puppy or a spider?	
Acceptability	Is Ethan good, bad or just OK? How good/bad? Is he really, really good/bad or just a little good/bad or just okay?		
Punishment	Should Ethan get in trouble? A little trouble or a lot of trouble?		
Justification	*Why* should/n't he get in trouble?		
Confirmation 7: Intention		[If should get in trouble:] What did Ethan want to do to Chris? Did he want to make Chris feel happy or scared?	
Parental knowledge	[If should get in trouble:] If his parents found out he tried to give Chris a puppy, *should* they tell him off? Why?		

2. Attempted harm (negative intention; positive outcome), adapted from Zelazo et al. (1996)

Issue / question	Nobes et al. (2016)	Experiment 1	Experiment 2
Introduction	Here's Anne. Anne's parents went on a trip to Brazil, far, far away. You know what they found there? They found a special kind of animal called a dax and they brought it back to Anne.		
Preference: Stroking	Now, a dax is pretty normal, it has skin just like you and me. When you stroke a dax, it feels good and it smiles		
Comprehension 1: Stroking	What does a dax do when you stroke it?		
Preference: Hitting	It doesn't like to be hit, though. That really, really hurts a dax, when you hit it. When you hit it, it hurts and it cries.		
Comprehension 2: Hitting	What does a dax do when you hit it?		
Intention	When Anne's parents gave her the dax she wanted to hurt it.		
Confirmation 1: Stroking	Now, what does a dax do when you stroke it?		
Confirmation 2: Hitting	And what does it do when you hit it?		
Knowledge	Now, Anne knows that a dax is normal. She knows that it cries when you hit it and that it smiles when you stroke it.		
Behavioral prediction	What is Anne going to do?		
Knowledge	That's right. Anne wanted to make the dax sad and she knew it didn't like to be hit, so		
Care			she held it very carefully to make sure it couldn't get away, and
Intention	she tried to hit it.		
Cause	But, you know what? When she tried to hit it, the dax wiggled away		
Outcome - act	so she ended up stroking it by mistake		
Outcome - emotion	and the dax smiled.		

Confirmation 3: Care		Did Anne hold the dax carefully to make sure it couldn't get away?
Confirmation 4: Outcome		Did Anne hit the dax or stroke it?
Confirmation 5: Cause (deliberate / accidental)		Why did Anne stroke the dax? Did she want to stroke the dax, or did she stroke it by mistake?
Confirmation 6: Intention		What did Anne try to do? Did she try to stroke the dax, or did she try to hit it?
Acceptability	Is Anne good, bad or just okay? How bad/good? Is she really, really bad/good or just a little bad/good/ or just okay?	
Punishment	Should Anne get in trouble? A little trouble or a lot of trouble?	
Justification	*Why* should/n't she get in trouble?	
Confirmation 7: Intention		[If shouldn't get in trouble:] What did Anne want to do to the dax? Did she want to make it feel nice or did she want to hurt it?
Parental knowledge	[If shouldn't get in trouble:] If her parents found out she tried to hit the dax, *should* they tell her off? Why?	