British Journal of Developmental Psychology (2015)

The British Psychological Society

www.wileyonlinelibrary.com

Young children use pedagogical cues to modulate the strength of normative inferences

Lucas P. Butler[1]*, Marco F. H. Schmidt[2], Jessica Bürgel[2] and Michael Tomasello[2]

[1]University of Maryland, College Park, Maryland, USA
[2]Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany

Young children understand pedagogical demonstrations as conveying generic, kind-relevant information. But, in some contexts, they also see almost any confident, intentional action on a novel artefact as normative and thus generic, regardless of whether this action was pedagogically demonstrated for them. Thus, although pedagogy may not be necessary for inferences to the generic, it may nevertheless be sufficient to produce inductive inferences on which the child relies more strongly. This study addresses this tension by bridging the literature on normative reasoning with that on social learning and inductive inference. Three-year-old children learned about a novel artefact from either a pedagogical or non-pedagogical demonstration, and then, a series of new actors acted on that artefact in novel ways. Although children protested normatively in both conditions (e.g., 'No, not like that'), they persisted longer in enforcing the learned norms in the face of repeated non-conformity by the new actors. This finding suggests that not all generic, normative inferences are created equal, but rather they depend – at least for their strength – on the nature of the acquisition process.

Social norms are central to the functioning of human society, governing everything from our moral behaviour and judgement (Killen & Rutland, 2011; Nucci & Turiel, 1978; Piaget, 1932; Turiel, 1983, 2002) to how to navigate everyday personal interactions (Nucci & Turiel, 1978; Searle, 1995), and providing the necessary foundation on which cooperation is based (Fehr & Fischbacher, 2004; Tomasello, 2009). In recent years, developmental research has focused on how young children understand and apply social norms, finding that children show fairly sophisticated understanding of social norms from a very young age (Rakoczy & Schmidt, 2013; Schmidt & Tomasello, 2012). From around age 3, and in some cases earlier, children understand the normative structure of simple novel actions and games (Rakoczy, Warneken, & Tomasello, 2008), selectively modulate their enforcement of norms depending on the nature of those norms and the identity of their violators (Schmidt, Rakoczy, & Tomasello, 2012), cooperatively create novel norms when the need arises (Göckeritz, Schmidt, & Tomasello, 2014), and even negotiate the use of novel norms in cases of conflict (Köymen *et al.*, 2014). And by elementary school age, children can explicitly reason about norms (Nucci & Turiel, 1978; Piaget, 1932; Turiel, 1983, 2002), explicate in language the normative dimensions of social roles and activities

*Correspondence should be addressed to Lucas P. Butler, Department of Human Development and Quantitative Methodology, University of Maryland, 3304 Benjamin Building, College Park, MD 20742, USA (email: lpbutler@umd.edu).

DOI:10.1111/bjdp.12108

(Kalish, 1998), and navigate normative judgement in complex interpersonal contexts (Killen & Rutland, 2011).

Young children clearly understand a great deal about social norms. But this literature has focused primarily on children's capacities for behaving and reasoning normatively, rather than the learning mechanisms underlying how those capacities play out as children acquire social norms in the first years of life. As a consequence, the developmental literature has by and large overlooked a key truth about normativity: That learning and acquiring norms – and differentiating them from mere idiosyncratic actions or preferences – is a complex learning problem fraught with inductive challenges and uncertainties. This truth has been long recognized in the study of concepts and categories as the classic inductive problem, in which children (and adults) must evaluate, on the basis of sparse evidence, what to generalize, and to what extent (Goodman, 1983). Decades of work in conceptual development from a variety of perspectives have investigated and debated how children tackle this inductive problem, building a picture of children's early inductive reasoning as reliant on a diverse array of cognitive and social-cognitive learning mechanisms (Carey, 2009; Gelman, 2003; Keil, 1989; Markman, 1989; Rumelhart & McClelland, 1986; Sloutsky & Fisher, 2004; Tenenbaum & Griffiths, 2001; Tenenbaum, Kemp, Griffiths, & Goodman, 2011). But in learning novel social norms, children face an analogous (and equally challenging) learning problem: Inferring which of the many actions they encounter in their first years of life truly represent important, normative knowledge or forms of action that others in general ought to adhere to, and which may simply be rational, affordant, or even common ways of doing something, but do not carry much normative weight.

This work sought to gain initial purchase on the question of how children bring to bear their well-established inductive inference capacities on this challenging and important social learning problem. There are two key parts to this question. First, what is necessary for children to view a novel action as potentially normative? And second, what factors influence the *strength* of those normative judgements? There is some evidence that begins to answer the first question. Even on the basis of simply observing a knowledgeable individual use an object in a particular way, children view this particular use as what the object is 'for' (Phillips, Seston, & Kelemen, 2012), and moreover expect others to behave this way and protest normatively (e.g., saying 'You must do it this way!') when others behave differently – Additional cues such as normative language and ostensive demonstration do not seem to be necessary (Schmidt, Rakoczy, & Tomasello, 2011). But how do children modulate the strength of those inferences? Although evidence certainly suggests some flexibility in children's norm enforcement (Rakoczy, Brosche, Warneken, & Tomasello, 2009; Schmidt *et al.*, 2012), this evidence concerns the *scope* of children's norm enforcement – who should adhere to a norm, and in what contexts – rather than how strongly children expect a given other to adhere to a particular norm. The ability to modulate the strength of normative expectations is critical for developing the ability to navigate the social world. Making an initial generalization, forming an initial normative expectation, is likely to be relatively low-cost. As long as it is quickly relinquished in the face of clear counterevidence – that is, as long as children stop trying to enforce a potential norm upon realizing that it is not one – their initial attempt at enforcement is unlikely to have any major consequences. But making a strong normative generalization, especially on the basis of relatively minimal evidence, is potentially more problematic. Continuing to demand that others behave in a particular way that does not accord to the actual norm might result in a number of negative consequences, from mild

rebuke to potential ostracism from peers. The ability to make wide-ranging and strong generalizations about the world on the basis of minimal evidence is a critical component of human cognition, allowing us to make predictions, construct explanations, and develop a rich causal understanding of the world. But it is only powerful insofar as it is selectively and judiciously applied. Thus, it seems critical that children be able to modulate the strength of their generalizations, especially about something that is as crucial to social functioning as norms.

How might children tackle this inductive challenge? Recent work on children's generalization of novel artefact information may provide a clue. Even in early infancy, children recognize communicative cues such as eye gaze and joint attention, which indicate that someone may intend to teach them something important (Csibra, 2010), and may even have a tendency to treat information that is ostensively communicated as kind-relevant (Futó, Téglás, Csibra, & Gergely, 2010) and even widely shared or generic (Egyed, Király, & Gergely, 2013). Moreover, by preschool age, children use that sensitivity to guide their inductive inferences. They take pedagogically demonstrated information as conveying all one needs to know about a novel toy (Bonawitz *et al.*, 2011), and are more likely to pass along that knowledge to a naïve third party (Vredenburgh, Kushnir, & Casasola, 2014). They make stronger generalizations about a novel artefact function – persisting longer in trying to perform that function using identical but inert objects of the same kind – when those functions were ostensively demonstrated for their benefit by a knowledgeable adult, rather than produced in an intentional, but non-ostensive manner (Butler & Markman, 2012, 2013). Children also use that information to guide their inferences about what defines a novel category – Judging a particular function as critical to category membership (over and above more salient features) when it was demonstrated ostensively (Butler & Markman, 2014). Critically, those experiments found no differences in children's *initial* generalizations. Most children across all conditions initially expected the additional objects to share the newly learned property. But upon discovering that some of the objects failed to share that property, they persisted in trying to get those objects to work and used possession of that property as an indicator of category membership more when the function had been demonstrated, indicating a boost in the strength of those inferences. This suggests that children use ostensive demonstrations not to judge whether or not information is generalizable at all, but to judiciously modulate the strength of those generalizations.

In this experiment, we aimed to bridge the two literatures discussed here – On normative reasoning and understanding, on the one hand, and on children's inductive inferences and social learning on the other. We asked whether the same learning mechanisms that seem crucial in inductive inference about categories may also play an important role in children's learning and generalization of novel social norms. Children observed novel goal-directed actions carried out with novel objects. They either observed this in an ostensive context, in which an adult pedagogically demonstrated the novel action for the child's benefit, or in a non-ostensive context, in which an unknown adult performed the action in a deliberate, intentional, but non-pedagogical manner. Children then saw a series of others perform the action in an effective, but markedly different manner. If children use pedagogical cues to modulate the strength of their normative inferences, we might expect them to persist in enforcing those norms longer when the novel actions were pedagogically demonstrated for them.

Method

Participants

Forty-eight 3-year-old children (24 girls, 24 boys; M_{age} = 3.21 years, range = 3.00–3.50) participated in the study. An additional four children were not included in the final sample, two due to external noise or interruptions during the study, and two because they did not want to finish the study. Children were recruited from local preschools in a medium-sized German city. Children were all native German speakers, and came from diverse socioeconomic backgrounds. We specifically targeted 3-year-old children, as this is the youngest age at which children consistently exhibit protest behaviour that can be clearly coded as normative (cf., Rakoczy *et al.*, 2008). Given our interest in strength of normative inferences specifically, we wanted to ensure that the children were old enough to make clearly normative responses (see below for more details on coding of children's verbal responses).

Materials

Five hand puppets, on average 30 cm tall (polar bear, rabbit, cat, cow, and mouse), a rubber ball, two instrumental tasks, and two target tasks (Table 1 for details) were used to conduct the study. Figure 1 provides an overview of the experimental setting.

Design and procedure

The overall task structure and key manipulation (pedagogical versus non-pedagogical) were closely modelled on prior research on social influences on children's normative judgement (Schmidt *et al.*, 2011). Children were randomly assigned to one of two between-subjects conditions (pedagogical and non-pedagogical), controlling for age and gender across conditions. In each condition, children received the same warm-up and target tasks. The order of the target tasks was varied across children (in each condition via Latin squares). The order of the warm-up tasks was fixed. Children were always introduced to the polar bear puppet, with whom they played the warm-up game and did the two instrumental tasks. The polar bear was also always the first puppet to carry out the alternative novel action. The order of the additional four puppets in the target tasks was counterbalanced (again by Latin squares within each condition). Two (pedagogical

Table 1. Overview of the two target tasks

Task	Material	Procedure
'Daxing'	Styrofoam board with gutter at one side, wooden building block, wooden stick with black paddle attached	A1: Put building block on board, use stick with paddle to push building block across the board into the gutter A2: Put building block on board, lift board so that building block slides into the gutter
'Meeking'	Metal can, plastic stick with hook, jingle bell with pipe-cleaner ring	A1: Place can face up on table, use hook to pick up jingle bell by the ring and place it into the can A2: Place can on its side, then turn can again and place over the bell so it covers it

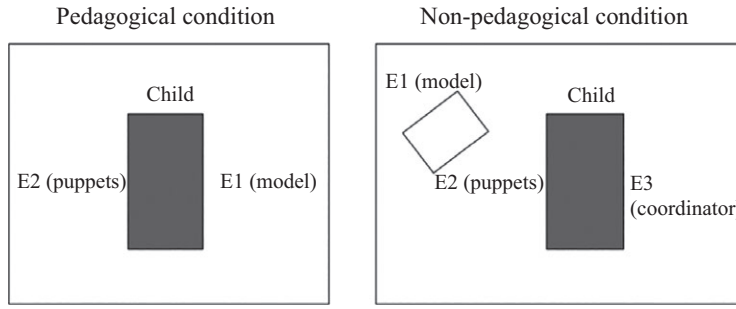

Figure 1. Overview of the experimental set-up in the two conditions of the experiment.

condition) or three (non-pedagogical condition) trained experimenters conducted the experiment.

Warm-up tasks

Children were first introduced to the polar bear puppet (controlled by E2) and then played the warm-up game, tossing the ball back and forth between themselves, the polar bear, and the experimenter leading the session (E1 in the pedagogical condition, E3 in the non-pedagogical condition). This process was then repeated with each of the 4 additional puppets. This was done to familiarize the child with the puppets, and to make them feel comfortable and at ease during the experiment.

Children then did the two instrumental warm-up tasks, designed to familiarize participants with the hand puppet and the fact that mistakes can happen and children may intervene. In each instrumental task, the experimenter leading the session (E1 in the pedagogical condition, E3 in the non-pedagogical condition; see below) performed an instrumental action (without using any language). The child then had the opportunity to reproduce the action. The polar bear then had a turn, and he made a mistake by failing to use a (causally) necessary means to perform the action correctly.

Target tasks

Children then participated in the two target tasks. Each target task consisted of four phases: A model phase, an action phase, a test phase, and a post-test action phase.

In the model phase, the experimenter performed an action (A1), while declaring that they were performing that particular action (e.g., 'Now I'm daxing'; Table 1). The polar bear was absent for this phase, having 'gone to sleep'. In the action phase, the experimenter gave the child the objects, saying 'now you can have that', giving the child the opportunity to act on the objects themselves (e.g., imitate A1) in the action phase. In the test phase, the polar bear returned and performed an alternative action (A2) with the objects. Finally, each of the additional four puppets returned, and each performed the same alternative action (A2). This served to provide children with counterevidence that the initial action (A1) might not be a social norm that is followed by others. Only the puppet currently acting was 'awake', and all others were 'sleeping'.

Although the action and test phases were the same in both conditions, the model phase differed importantly (Table 1). In the pedagogical condition, E1, the polar bear, and the child sat together at a table, and E1 looked at the child, called the child by her name, and said, 'Look!' before performing A1. Upon completing the action, she said, 'Great!' before

giving the child the objects. In the non-pedagogical condition, E1 had never met the child (or, ostensibly, E3 or any of the puppets), never established eye contact with anyone, was busily working, and sat at a separate table performing A1 'for herself'. Nevertheless, children's attention to the model's action A1 was drawn by E1 producing noise when getting the objects ready to use, by E1's soliloquizing to himself that he was performing the action, and by E3's looking curiously at E1's actions. During the test phase, E1 remained at the separate table, pretending to write something down on a piece of paper for her own purposes. She did not watch either the puppets or the child at any time.

After the test phase, the polar bear returned. He then gave the objects to the child, saying, 'Now you can have that again'. This was done to keep children engaged in the study by giving them something active to do before moving on to the second target task.

Coding and reliability

All sessions were videotaped and children's verbal and behavioural responses were transcribed and subsequently coded by a single observer. A second independent observer coded a random sample of 25% of all sessions for reliability.

Children's verbal and behavioural responses in the test phase of each target task were coded into one of three hierarchically ordered categories according to their level of normativity. The highest category was normative protest, which included verbal and/or behavioural protest, critique, and correction (including teaching) that made use of normative vocabulary (e.g., 'You should (not) do this!'). The second category was imperative protest, comprising verbal and/or behavioural protest without normative vocabulary, but which used imperative phrases (e.g., 'Take the thing!' or 'Don't destroy it!') that were related to the puppets' actions with the materials. The third category was implicit protest, comprising behaviours suggestive of protest (e.g., pointing or giving objects to the puppets), but which were not explicit enough to be coded as clear imperatives. All other behaviours were coded as irrelevant. For each trial on each target task, children received the code for the highest category of behaviour seen on that trial. For the purposes of our main analyses, these categories were collapsed into normative (only the highest code) and non-normative (either imperative or implicit). Reliability was good, Cohen's $\kappa = .82$.

Predictions and data analysis plan

As described above, this work is largely (although not solely) based on theoretical work about the role of pedagogical demonstration in children's learning, specifically the third prediction of 'natural pedagogy' theory, which holds that children tend to interpret pedagogical demonstrations as conveying important, generic information about the world (Csibra & Gergely, 2009). Based on previous empirical work on learning from pedagogical demonstration (Butler & Markman, 2012) and its potential role in normative inference (Schmidt *et al.*, 2011), our prediction was one of different inferential strength in the two conditions. Specifically, although Schmidt *et al.* (2011) found that pedagogical demonstration played no essential role in whether or not children initially protested against a possible violation of a novel game rule, Butler and Markman (2012) found that children made stronger inferences, ones that required more negative evidence to override, about properties that had been pedagogical demonstration for them. Based on this, we predicted that children would make stronger normative inferences about actions they saw pedagogically demonstrated for their benefit, and thus they would persist longer

in making normative protests against individuals carrying out that action in a different manner, requiring more counterevidence (evidence that actually people do it this alternative way) than when they had seen the action done in an intentional, but non-pedagogical manner.

Thus, our specific prediction was that there would be no difference between conditions, for either normative or non-normative protest, on trial 1 (consistent with prior work), but that children's level of normative protest specifically would be higher on trial 2 in the pedagogical condition than in the non-pedagogical condition. We included five trials in total (i.e., five instances of counterevidence), because this was the first study on children's persistence in normative protesting and it might have been the case that – against our hypothesis – children's normative protest would remain stable at high levels even on further trials. This enables us to provide a more comprehensive picture of children's reactions to repeated counterevidence, which will also be helpful for designing future research studies on this topic.

To test our predictions, we first used a generalized linear mixed model (GLMM). Although GLMMs have not been common practice in developmental psychology until recently, they are now used when the research design and data require such techniques (cf., McAuliffe, Jordan, & Warneken, 2015). In our case, the response variable was dichotomous (protest: Yes or no), so an ANOVA was not appropriate. For our main question, we had to account for repeated observations per child (i.e., non-independent data) and therefore used mixed models that allow for the inclusion of both fixed and random effects (this is analogous to using a repeated-measures ANOVA).

We first conducted a binomial GLMM on children's responses (protest: Yes or no) with condition, trial, and protest type (including their 3- and 2-way interactions) as predictors, task as a fixed-effect control variable, random effect of subject, and random slopes for the effects of trial and protest type. In mixed models, random effects are included for variables that are measured or observed repeatedly, most importantly for the subject (participant ID as an intercept term, allowing intercepts to vary between participant) and for test predictors (in our case: Trial and protest type; Barr, Levy, Scheepers, & Tily, 2013). As for the fixed effects, we wanted to test whether our main predictors were significant regardless of which specific action the children were responding to, and we thus included task in the model. It should be noted that there was no effect of task on children's protest behaviour and that omitting task as a fixed factor on this or the prior GLMM did not substantially change the results.

Our first step in testing our model was to compare the full model to a null model that included only the intercept and random slopes. In large-scale simulations, this has been shown to be as good a method for protecting against type I error rates (Forstmeier & Schielzeth, 2011) and, interestingly, even superior (i.e., no inflation of type I error) to the alternative approach of using *post-hoc* tests and p-value corrections, such as Bonferroni or others. This allowed us to then specifically look at trials 1 and 2 without needing to protect against type I error rates a second time, as the overall full-null model comparison was significant.

Our next step was to run separate binomial GLMMs on children's responses on trial 1 and trial 2, with condition and type of protest, as well as their interaction, as predictors, fixed effect of task, and a random intercept of subject. To test for the interaction between type of protest and condition on each trial, we compared the full models to reduced models that included the main effects of condition and type of protest, but not their interaction, and tested for significance using a likelihood ratio test.

Our final step was to conduct follow-up analysis looking at protest specifically on trial 2. Due to the distribution of the data, with many children not protesting, we conducted Wilcoxon rank-sum tests for independent samples to test for differences across conditions on children's normative and non-normative protest.

Results

The overall GLMM on children's protest (Figure 2) was significant compared to a null model including only these fixed-effect control variables and random effects and slopes (likelihood ratio test, $\chi^2 = 80.64$, df = 7, $p < .001$), helping to ensure that any subsequent tests would not inordinately inflate the type I error rate (Forstmeier & Schielzeth, 2011).

Follow-up GLMMs on trials 1 and 2 revealed a significant condition × type of protest interaction on trial 2 ($\chi^2 = 5.79$, df = 1, $p = .016$), but no such interaction on trial 1 ($\chi^2 = 0.010$, df = 1, $p = .907$).

Follow-up Wilcoxon rank-sum tests for independent samples of children's responses on trial 2 (responses summed across both tasks) confirmed that children showed more normative protest in the pedagogical condition ($M_{\text{pedagogical}} = 0.42$, $SD = 0.72$) than in the non-pedagogical condition ($M_{\text{instrumental}} = 0.13$; $SD = 0.45$; $p = .038$), but showed no such effect for non-normative protest ($M_{\text{pedagogical}} = 0.58$, $SD = 0.83$; $M_{\text{instrumental}} = 0.54$, $SD = 0.84$, $p = .825$).

There were no other further significant effects of condition on children's protest on trials 3, 4, or 5 (all p's > .20).

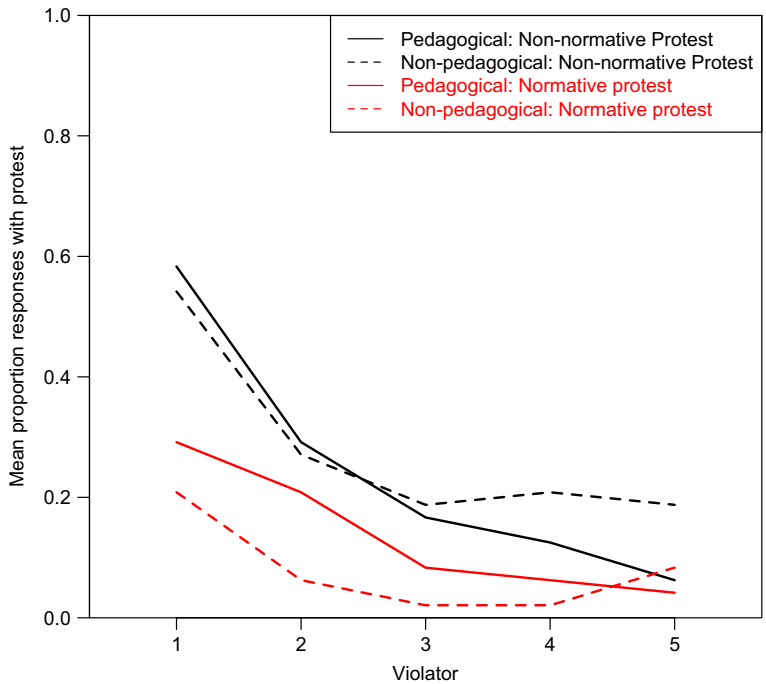

Figure 2. Children's normative and non-normative protest behaviour across the five puppet violators.

Discussion

This finding presents evidence that children's sensitivity to whether or not knowledge is being conveyed pedagogically for their benefit (Csibra, 2010; Csibra & Gergely, 2009) plays a role not only in their reasoning about objects and their properties (Butler & Markman, 2012, 2013, 2014; Futó *et al.*, 2010) or about whether knowledge is widely shared (Egyed *et al.*, 2013), but in their normative reasoning about how others *ought* to behave. Children made significantly stronger normative inferences about novel actions when they saw those actions pedagogically demonstrated, relative to seeing the identical actions carried out in a deliberate and intentional but non-pedagogical manner. Consistent with prior work (Schmidt *et al.*, 2011), 3-year-old children showed a general tendency to jump to a normative interpretation from simply observing an intentional action performed by an unknown knowledgeable adult, protesting when a third party used the same objects to perform a similar, but markedly different action. But when they saw additional individuals continue to perform this markedly different action, children persisted in their normative protest significantly more when the original action had been explicitly demonstrated for them, compared to seeing the identical action performed in a non-pedagogical manner.

The current research adds further insight into our growing understanding of children's social learning and normative reasoning more broadly and relates importantly to several key literatures. First, it is clear that children do not only learn normative information from pedagogical demonstration or direct instruction. Indeed, a great deal of work has illustrated very young children's robust ability to learn from overhearing or overseeing. Much of this work has been focused on toddlers' ability to learn words from overheard speech (Akhtar, 2005; Floor & Akhtar, 2006), although this capacity has also been shown in imitative learning of novel actions, showing that imitative learning from third parties is grounded in developing capacities for self-recognition and perspective taking (Herold & Akhtar, 2008). Moreover, preschoolers not only imitate third-party actions but show evidence of overimitation of causally irrelevant actions, even when not directed at them (Nielsen, Moore, & Mohamedally, 2012). And as mentioned previously, children do make some level of normative inferences simply from watching a confident actor carry out a novel action for their own purposes (Schmidt *et al.*, 2011), which is also consistent with the results from trial 1 of the current experiment. What the current findings add, then, is further depth to our understanding of the role pedagogical demonstration might play in normative learning, namely one of strengthening the initial normative inferences that children make on the basis of any intentional action.

Further, the current work relates in important ways to the growing body of literature on children's overimitation. The classic overimitation finding is that children regularly reproduce clearly irrelevant and unnecessary actions to achieve a simple goal when they observe an ostensibly knowledgeable adult demonstrating these actions (Horner & Whiten, 2005). In recent years, this literature has been greatly expanded, showing that this phenomenon exists across different cultures (Nielsen & Tomaselli, 2010), that it appears to develop very early in childhood and only becomes more robust over the course of development (McGuigan & Whiten, 2009; McGuigan, Whiten, Flynn, & Horner, 2007), and that the social status and age of the model influence the extent to which overimitation occurs (McGuigan, 2013; McGuigan & Robertson, 2015). There are differing perspectives on what explains this phenomenon. One argument is that children may automatically encode intentional action as causally necessary to achieve a broader goal (Lyons, Damrosch, Lin, Macris, & Keil,

2011; Lyons, Young, & Keil, 2007), thus leading them to imitate causally irrelevant actions even though they can explicitly distinguish them from relevant actions when asked to do so. Another perspective is that overimitation stems from children's motivation to affiliate with others. That is, children recognize that actions are irrelevant, but imitate them in order to be like or relate to the model, because they are generally motivated to carry out actions that boost affiliation (Over & Carpenter, 2012). Finally, overimitation has been argued to be intimately related to the capacity for normativity. On this view, children's recognition or lack thereof of specific actions as causally irrelevant is separable from their imitative actions. That is, they may recognize specific actions as irrelevant, but because they are instead focused on the overarching action sequence and goal (Buchsbaum, Gopnik, Griffiths, & Shafto, 2011), perhaps as part of a conventional practice (Keupp, Behne, & Rakoczy, 2013), they see all of the elements of the overarching action as obligatory or necessary from a normative standpoint, and enforce them in much the same way they do explicitly normative acts (Kenward, 2012; Kenward, Karlsson, & Persson, 2011; Keupp *et al.*, 2013). Although the current research was not specifically about overimitation per se, some of the model's actions could be seen as causally unnecessary (although not irrelevant), especially once children see the alternative action. Thus, one possible interpretation of the current results is that pedagogical demonstration led children to see the individual elements of the actions as more relevant (whether causally or normatively), leading them to persist more in insisting that additional actors reproduce them exactly. Given that we found an effect of pedagogical demonstration specifically for the strength of normative protest may favour the latter, normative account of overimitation, but future work is necessary to pull together these as yet disparate lines of research. An intriguing future direction that the current work opens up would be to investigate potential interplays and interactions between mode of presentation, causal relevance or irrelevance, and normative learning.

Stepping back, the current findings have several important implications for our understanding of early normative reasoning. First, it provides further evidence that although children have a strong tendency to view intentional actions as carrying normative weight for others, they are actively evaluating the social context in which those actions are carried out to modulate their inferences. Not only are children sensitive to whether a model is acting knowledgably and intentionally (Schmidt *et al.*, 2011), but they appear to also be sensitive to *why* an adult is carrying out that action, making stronger inferences when the adult's acts are framed for the child within an ostensive context rather than a context of incidental observation. Second, it further illuminates our understanding of the role of pedagogy in children's early learning. Although pedagogical demonstration can have a powerful impact on children's reasoning about the importance and generalizability of novel actions and object properties (Butler & Markman, 2012, 2013, 2014; Egyed *et al.*, 2013; Futó *et al.*, 2010; Yoon, Johnson, & Csibra, 2008), it clearly does not always play such a key role (Phillips *et al.*, 2012; Schmidt *et al.*, 2011). The current research helps resolve this apparent tension in the literature by pinpointing the nuanced role that pedagogy may play – Not in affecting *whether* children see novel actions as generic or normative, but in how strong an expectation that generalization licenses (see also Butler & Markman, 2012). This opens up a number of avenues for future research investigating how children modulate the strength of their normative inferences across various contexts and types of norms, and how this process affects children's norm acquisition in their everyday lives.

Acknowledgements

This research was supported by a postdoctoral fellowship from the Alexander von Humboldt Foundation to the first author. We are especially thankful to Lou Haux, Susanne Mauritz, and Katrin Schumann for their coordination and assistance throughout this project. We also thank Roger Mundry for assistance with data analysis, and Philipp Berger for help with data coding.

References

Akhtar, N. (2005). The robustness of learning through overhearing. *Developmental Science*, *8*, 199–209. doi:10.1111/j.1467-7687.2005.00406.x

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*, 255–278. doi:10.1016/j.jml.2012.11.001

Bonawitz, E., Shafto, P., Gweon, H., Goodman, N. D., Spelke, E., & Schulz, L. (2011). The double-edged sword of pedagogy: Instruction limits spontaneous exploration and discovery. *Cognition*, *120*, 322–330. doi:10.1016/j.cognition.2010.10.001

Buchsbaum, D., Gopnik, A., Griffiths, T. L., & Shafto, P. (2011). Children's imitation of causal action sequences is influenced by statistical and pedagogical evidence. *Cognition*, *120*, 331–340. doi:10.1016/j.cognition.2010.12.001

Butler, L. P., & Markman, E. M. (2012). Preschoolers use intentional and pedagogical cues to guide inductive inferences and exploration. *Child Development*, *83*, 1416–1428. doi:10.1111/j.1467-8624.2012.01775.x

Butler, L. P. & Markman, E. M. (2013). *Preschoolers' ability to navigate communicative interactions in guiding their inductive inferences*. Proceedings of the 35th Annual Meeting of the Cognitive Science Society.

Butler, L. P., & Markman, E. M. (2014). Preschoolers use pedagogical cues to guide radical reorganization of category knowledge. *Cognition*, *130*(1), 116–127. doi:10.1016/j.cognition.2013.10.002

Carey, S. (2009). *The origin of concepts*. New York, NY: Oxford University Press.

Csibra, G. (2010). Recognizing communicative intentions in infancy. *Mind & Language*, *25*(2), 141–168. doi:10.1111/j.1468-0017.2009.01384.x

Csibra, G., & Gergely, G. (2009). Natural pedagogy. *Trends in Cognitive Sciences*, *13*(4), 148–153. doi:10.1016/j.tics.2009.01.005

Egyed, K., Király, I., & Gergely, G. (2013). Communicating shared knowledge in infancy. *Psychological Science*, *24*, 1348–1353. doi:10.1177/0956797612471952

Fehr, E., & Fischbacher, U. (2004). Third-party punishment and social norms. *Evolution and Human Behavior*, *25*(2), 63–87. doi:10.1016/S1090-5138(04)00005-4

Floor, P., & Akhtar, N. (2006). Can 18-month-old infants learn words by listening in on conversations? *Infancy*, *9*, 327–339. doi:10.1207/s15327078in0903_4

Forstmeier, W., & Schielzeth, H. (2011). Cryptic multiple hypotheses testing in linear models: Overestimated effect sizes and the winner's curse. *Behavioral Ecology and Sociobiology*, *65*(1), 47–55. doi:10.1007/s00265-010-1038-5

Futó, J., Téglás, E., Csibra, G., & Gergely, G. (2010). Communicative Function Demonstration induces kind-based artifact representation in preverbal infants. *Cognition*, *117*(1), 1–8. doi:10.1016/j.cognition.2010.06.003

Gelman, S. A. (2003). *The essential child: Origins of essentialism in everyday thought*. New York, NY: Oxford University Press.

Göckeritz, S., Schmidt, M. F. H., & Tomasello, M. (2014). Young children's creation and transmission of social norms. *Cognitive Development*, *30*, 81–95. doi:10.1016/j.cogdev.2014.01.003

Goodman, N. (1983). *Fact, fiction, and forecast* (4th ed.). Cambridge, MA: Harvard University Press.

Herold, K. H., & Akhtar, N. (2008). Imitative learning from a third-party interaction: Relations with self-recognition and perspective taking. *Journal of Experimental Child Psychology*, *101*(2), 114–123. doi:10.1016/j.jecp.2008.05.004

Horner, V., & Whiten, A. (2005). Causal knowledge and imitation/emulation switching in chimpanzees (Pan troglodytes) and children (Homo sapiens). *Animal Cognition*, *8*, 164–181. doi:10.1007/s10071-004-0239-6

Kalish, C. (1998). Reasons and causes: Children's understanding of conformity to social rules and physical laws. *Child Development*, *69*, 706–720. doi:10.1111/j.1467-8624.1998.tb06238.x

Keil, F. C. (1989). *Concepts, kinds, and cognitive development*. Cambridge, MA: MIT Press.

Kenward, B. (2012). Over-imitating preschoolers believe unnecessary actions are normative and enforce their performance by a third party. *Journal of Experimental Child Psychology*, *112*, 195–207. doi:10.1016/j.jecp.2012.02.006

Kenward, B., Karlsson, M., & Persson, J. (2011). Over-imitation is better explained by norm learning than by distorted causal learning. *Proceedings of the Royal Society of London B: Biological Sciences*, *278*, 1239–1246. doi:10.1098/rspb.2010.1399

Keupp, S., Behne, T., & Rakoczy, H. (2013). Why do children overimitate? Normativity is crucial. *Journal of Experimental Child Psychology*, *116*, 392–406. doi:10.1016/j.jecp.2013.07.002

Killen, M., & Rutland, A. (2011). *Children and social exclusion: Morality, prejudice, and group identity*. Hoboken, NJ: John Wiley.

Köymen, B., Lieven, E., Engemann, D. A., Rakoczy, H., Warneken, F., & Tomasello, M. (2014). Children's norm enforcement in their interactions with peers. *Child Development*, *85*, 1108–1122. doi:org/10.1111/cdev.12178

Lyons, D. E., Damrosch, D. H., Lin, J. K., Macris, D. M., & Keil, F. C. (2011). The scope and limits of overimitation in the transmission of artefact culture. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *366*, 1158–1167. doi:10.1098/rstb.2010.0335

Lyons, D. E., Young, A. G., & Keil, F. C. (2007). The hidden structure of overimitation. *Proceedings of the National Academy of Sciences*, *104*, 19751–19756. doi:10.1073/pnas.0704452104

Markman, E. M. (1989). *Categorization and naming in children: Problems of induction*. Cambridge, MA: MIT Press.

McAuliffe, K., Jordan, J. J., & Warneken, F. (2015). Costly third-party punishment in young children. *Cognition*, *134*, 1–10. doi:10.1016/j.cognition.2014.08.013

McGuigan, N. (2013). The influence of model status on the tendency of young children to over-imitate. *Journal of Experimental Child Psychology*, *116*, 962–969. doi:10.1016/j.jecp.2013.05.004

McGuigan, N., & Robertson, S. (2015). The influence of peers on the tendency of 3- and 4-year-old children to over-imitate. *Journal of Experimental Child Psychology*, *136*, 42–54. doi:10.1016/j.jecp.2015.03.004

McGuigan, N., & Whiten, A. (2009). Emulation and "overemulation" in the social learning of causally opaque versus causally transparent tool use by 23- and 30-month-olds. *Journal of Experimental Child Psychology*, *104*, 367–381. doi:10.1016/j.jecp.2009.07.001

McGuigan, N., Whiten, A., Flynn, E., & Horner, V. (2007). Imitation of causally opaque versus causally transparent tool use by 3- and 5-year-old children. *Cognitive Development*, *22*, 353–364. doi:10.1016/j.cogdev.2007.01.001

Nielsen, M., Moore, C., & Mohamedally, J. (2012). Young children overimitate in third-party contexts. *Journal of Experimental Child Psychology*, *112*(1), 73–83. doi:10.1016/j.jecp.2012.01.001

Nielsen, M., & Tomaselli, K. (2010). Overimitation in Kalahari bushman children and the origins of human cultural cognition. *Psychological Science*, *21*, 729–736. doi:10.1177/0956797610368808

Nucci, L. P., & Turiel, E. (1978). Social interactions and the development of social concepts in preschool children. *Child Development*, *49*, 400–407. doi:10.2307/1128704

Over, H., & Carpenter, M. (2012). Putting the social into social learning: Explaining both selectivity and fidelity in children's copying behavior. *Journal of Comparative Psychology*, *126*, 182–192. doi:10.1037/a0024555

Phillips, B., Seston, R., & Kelemen, D. (2012). Learning about tool categories via eavesdropping. *Child Development*, *83*, 2057–2072. doi:10.1111/j.1467-8624.2012.01827.x

Piaget, J. (1932). *The moral judgment of the child*. London, UK: Routledge & Kegan Paul.

Rakoczy, H., Brosche, N., Warneken, F., & Tomasello, M. (2009). Young children's understanding of the context-relativity of normative rules in conventional games. *British Journal of Developmental Psychology*, *27*, 445–456. doi:10.1348/026151008X337752

Rakoczy, H., & Schmidt, M. F. H. (2013). The early ontogeny of social norms. *Child Development Perspectives*, *7*(1), 17–21. doi:10.1111/cdep.12010

Rakoczy, H., Warneken, F., & Tomasello, M. (2008). The sources of normativity: Young children's awareness of the normative structure of games. *Developmental Psychology*, *44*, 875–881. doi:10.1037/0012-1649.44.3.875

Rumelhart, D. E., & McClelland, J. L. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition*. Cambridge, MA: MIT Press.

Schmidt, M. F. H., Rakoczy, H., & Tomasello, M. (2011). Young children attribute normativity to novel actions without pedagogy or normative language. *Developmental Science*, *14*, 530–539. doi:10.1111/j.1467-7687.2010.01000.x

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

Schmidt, M. F. H., & Tomasello, M. (2012). Young children enforce social norms. *Current Directions in Psychological Science*, *21*, 232–236. doi:10.1177/0963721412448659

Searle, J. R. (1995). *The construction of social reality*. New York, NY: Simon and Schuster.

Sloutsky, V. M., & Fisher, A. V. (2004). Induction and categorization in young children: A similarity-based model. *Journal of Experimental Psychology: General*, *133*, 166–188. doi:10.1037/0096-3445.133.2.166

Tenenbaum, J. B., & Griffiths, T. L. (2001). Generalization, similarity, and Bayesian inference. *Behavioral and Brain Sciences*, *24*, 629–640. doi:10.1017/S0140525X01780089

Tenenbaum, J. B., Kemp, C., Griffiths, T. L., & Goodman, N. D. (2011). How to grow a mind: Statistics, structure, and abstraction. *Science*, *331*, 1279–1285. doi:10.1126/science.1192788

Tomasello, M. (2009). *Why we cooperate*. Cambridge, MA: MIT Press.

Turiel, E. (1983). *The development of social knowledge: Morality and convention*. Cambridge, UK: Cambridge University Press.

Turiel, E. (2002). *The culture of morality: Social development, context, and conflict*. Cambridge, UK: Cambridge University Press.

Vredenburgh, C., Kushnir, T., & Casasola, M. (2014). Pedagogical cues encourage toddlers' transmission of recently demonstrated functions to unfamiliar adults. *Developmental Science*, *18*, 645–654. doi:10.1111/desc.12233

Yoon, J. M. D., Johnson, M. H., & Csibra, G. (2008). Communication-induced memory biases in preverbal infants. *Proceedings of the National Academy of Sciences*, *105*, 13690–13695. doi:10.1073/pnas.0804388105

Received 22 September 2014; revised version received 6 July 2015