Title
Is children's norm learning rational? A meta-analysis

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 43(43)

ISSN
1069-7977

Authors
Partington, Scott
Nichols, Shaun
Kushnir, Tamar

Publication Date
2021

Peer reviewed

Is children's norm learning rational? A meta-analysis

Scott Partington[1,2] **(ssp95@cornell.edu)**
Shaun Nichols[1] **(sbn44@cornell.edu)**
Tamar Kushnir[2] **(tk397@cornell.edu)**
[1]Cornell University, Department of Philosophy, Goldwin Smith Hall, Ithaca, NY 14853 USA
[2]Cornell University, Department of Human Development, Beebe Hall, Ithaca, NY 14853 USA

Abstract

A good deal of recent research has examined children's norm learning across a wide range of novel contexts. The typical interpretation of these findings is that children's norm learning is driven by group-based biases. In this paper, we present an alternative interpretation and corresponding meta-analyses that cast the current body of evidence in a rather different light. First, we argue the extant literature uses an ill-suited standard for assessing bias. Rather than comparing children's judgments to what is expected under random chance (a 'random standard'), bias is better assessed by comparing children's judgments to what is most probable, given their total evidence (an 'evidential standard'). Next, we report a meta-analysis of the known findings to date (k = 40 effect sizes; N = 1,369 in total; ages 4- to 13-years-old) to compare children's norm learning against an appropriate evidential standard. Meta-analytic estimates reveal that children's norm learning is not *restrictively* biased toward narrow-scope inferences on account of group-based factors. Rather, the findings to date are consistent with children's norm learning being *rational* (i.e., statistically appropriate, given their evidence) or even *inclusively* biased toward making the wide-scope inference that a novel norm applies to everyone in a population. We conclude with brief discussion of implications for current understanding and future research on norm acquisition.

Keywords: norms; social learning; cognitive development; rational learning

Introduction

Across many domains in cognitive development, children are *evidentially rational* learners: they make appropriate statistical inferences, given their total evidence (Xu, 2019; Gopnik & Wellman, 2012; Kushnir & Xu, 2012; Perfors et al., 2011). For those interested in better understanding children's social learning and normative cognition, it seems a natural step to pose the question: Is children's norm learning similarly rational?

Going by current accounts, the answer would be an emphatic *no*. Early emerging, group-based biases in children's social learning and judgment are often documented, both about meaningful groups in the real world (reviewed in Chalik et al., 2019) and arbitrary groups in experimental settings (reviewed in Dunham, 2018). Given the expansive nature of these phenomena, one emerging hypothesis is these biases are partly due to common, deep-rooted processes in social cognition. The proposal is that children rely on a set of coalitional heuristics to assimilate and solidify their abstract, causal-explanatory theories of social group behavior (see, e.g., Chudek & Henrich, 2011, p. 219; Dunham, 2018, p. 788; Chalik et al., 2019, p. 5). Over the past several years this account has gained empirical support due to findings that indicate children's norm learning is driven by group-based biases even under completely novel, third-party contexts (e.g., Roberts et al., 2017a; Chalik & Rhodes, 2018; later reviewed in detail). At face value, then, this literature portrays children's norm learning as fundamentally not rational, with mere representations of social groups eliciting biases in learning (see, e.g., Roberts et al., 2017a, p. 593; Chalik & Rhodes, 2020, p. 80).

In this paper, however, we put forth an alternative interpretation and corresponding meta-analyses that cast the current body of evidence in a rather different light. Currently, the literature assesses bias by comparing children's judgments to what is expected under random chance (i.e., a 'random standard'). We argue that learning biases are better assessed by comparing children's judgments to what is most probable, given the total relevant evidence (i.e., an 'evidential standard'). In tandem, we conduct a meta-analysis of known findings to date in order to compare children's norm learning against an appropriate evidential standard. We use this evidential standard to discriminate between competing hypotheses about potential biases that learners bring to bear on normative inference. Ultimately, we find that children's inferences are not *restrictive*, or unduly narrow-scope on account of group-based factors. Rather, the findings to date are consistent with children's inferences being *rational* (i.e., statistically appropriate, given their evidence) or even *inclusive* (i.e., unduly wide-scope, or a tendency towards believing a novel norm applies to everyone in a population).

To begin, we review the findings that are understood to demonstrate a crucial role for group-based biases in children's norm learning.

Novel groups and the random standard for bias

The evidence in question comes entirely from studies which employ a *novel groups* paradigm. In this paradigm, children are first introduced to two previously unfamiliar social groups that are clearly demarcated by visual characteristics (e.g., body or t-shirt color) and verbally labelled by experimenters (e.g., "Hibbles" and "Glerks"). Note that by introducing novel groups, these studies plausibly examine children's learning from a naïve state, rather than tapping their knowledge about specific real-world groups or the influences of affiliative responses. Next, children experience a learning episode during which they receive evidence about the behavioral tendencies of one or both groups. After, children make a forced-choice judgment about the content of the norm that group members follow.

To date, novel groups research has focused on children's learning about two types of norms: *subject norms* and *agent-patient norms*. Subject norms govern self-directed actions. For example, students are subjects to the rule "Students cannot wear pajamas to school." In contrast, agent-patient norms govern dyadic interactions. For the rule "Don't tease your younger siblings," an older child is the agent, and the younger siblings are the patients. Below, we detail notable findings regarding each type of norm.

Roberts and colleagues (2017; 2018; 2019a; 2019b; Foster-Hanson et al., 2018; see also Riggs, 2019) have most extensively studied children's learning about subject norms. In these studies, children are introduced to two groups of non-humanlike characters labelled "Hibbles" and "Glerks." In the learning episode, children receive evidence about the groups' behavioral regularities (e.g., "Hibbles eat purple berries and Glerks eat orange berries"). After, children make a dichotomous judgment about whether a non-conforming individual is breaking a norm (e.g., "Is it okay or not okay for this Hibble to eat an orange berry?"). The typical finding is that children use the descriptive evidence to make a prescriptive judgment: children judge that non-conformity is "not okay" at a greater rate than random chance (i.e., 50%), and younger children in particular are prone to this pattern of response. Roberts and colleagues propose this tendency may be due to a restrictive, group-based bias according to which "group regularities may exert influence by rather automatically fostering a negative evaluative stance," to non-conformity (2017, p. 593) and understand the findings to highlight "the profound implications groups concepts have for stereotyping and normative reasoning" (2018, p. 149).

Chalik and Rhodes (2014; 2018; see also Roberts et al., 2019a) have used a similar procedure to examine children's learning about agent-patient norms. In this work, children are introduced to two groups of humanlike characters who are wearing different t-shirt colors and labelled accordingly (e.g., the "red group" and "blue group"). In the learning episode, children observe a member from one group directing an action toward another individual (e.g., "A kid from the blue group stole a cookie from a kid from the red group"). After, children make a judgment about to whom an additional action will be directed (e.g., "One day, a kid from the blue group hit someone when they were building a tower. Who did the kid from the blue group hit? Did he hit another kid from the blue group? Or did he hit a kid from the red group?"). The typical pattern of response is that children judge negative-valence actions (e.g., hitting, stealing, excluding) will be directed toward the agent's outgroup at a greater rate than random chance, whereas positive-valence actions (e.g., hugging, sharing, including) will not[1]. Together with findings from Roberts and colleagues, these results are understood to reflect how "children assume that all norms (not just moral norms) are bounded by some type of category. Thus, once they learn that something falls under the scope of a norm… they assume there is a boundary on to whom it applies" (Chalik & Rhodes, 2020, p. 80).

In summary, a good deal of recent research is marshaled to support the notion that children's norm learning is driven by restrictive, group-based biases, even in novel contexts. More specifically, the claim is that group-based biases lead to restrictive inferences about norm-governed behavior such that if one learns about the behavior of one or more group members, then one assumes that other members of the group must conform to the same behavior. However, this interpretation relies crucially on a particular methodological choice: assessing bias by comparing children's pattern of response against random chance (i.e., a 'random standard'). Next, we propose that bias is better assessed by comparing children's judgments against an evidential standard. To do so, first we provide some background from the philosophical literature on epistemic rationality.

Novel groups and the evidential standard for bias

The dominant framework in analytic epistemology promotes and defends an evidentialist notion of rationality (c.f., Talbott, 2016). On this view, an agent's beliefs[2] are rational to the extent that they are supported by the agent's total evidence. The minimal requirements for this sort of rationality are twofold. One: the agent's beliefs cohere to the laws of probability. Two: when the agent receives new evidence, her initial (or 'prior') beliefs are systematically updated according to a principle of conditionalization (often, Bayes' Rule is the appropriate principle). Philosophers mainly use these requirements to describe ideal epistemic agents, in much the same way economists use expected utility theory to describe ideal economic agents. So just as behavioral economists use expected utility theory to compute a baseline for assessing whether human decision-making is rational or biased, cognitive scientists can use the minimal requirements for epistemic rationality to do the same for human learning. In this section, we apply this technique to the learning scenarios in the novel groups paradigm.

First, we need to specify the learner's evidence. During the learning episode, children receive *at least* one sample of evidence about a group norm (e.g., "Hibbles eat purple berries"). Of course, there are also pragmatic factors at play: the generic formulation used in some studies plausibly implies 'more than one' (c.f., Tessler et al., 2020), and the pedagogical context common to all studies may also have a similar effect (c.f., Bonawitz et al., 2011). Further, in Roberts and colleagues' studies, children receive evidence about the other group as well (e.g., "Glerks eat orange berries"). Indeed, a purely descriptive model of children's learning may

[1] Note there is a sense in which valence-infused norms may not be truly 'novel' if children already have prior beliefs about harmful and/or prosocial actions. We return to this point in the Discussion.

[2] Epistemologists draw a distinction between beliefs and *degrees* of belief ('credences'). Since the distinction does not play a large role in cognitive science, we use 'belief' as a catch-all here.

do well to incorporate these factors. For the purposes of computing a rational baseline, however, we will start with a conservative specification that does not presume their influence, as otherwise the baseline would presume the very thing that empirical deviations are meant to demonstrate. Thus, our treatment will focus on the case of one independent sample, but throughout we will also mention how the baseline for epistemic rationality shifts if one considers the pragmatic features of the novel groups paradigm to imply stronger confirmatory evidence.

Next, we need to specify the learner's hypothesis space. For the subject cases, when children make the judgment about a non-conforming individual (e.g., a Hibble who eats an orange berry) they are judging which hypothesis is more likely: is the norm *restricted* to only the sampled behavior ($h_{\text{restrictive}}$), or does the norm also *include* behaviors that were not sampled ($h_{\text{inclusive}}$)? For example, can Hibbles only eat purple berries ($h_{\text{restrictive}}$), or can Hibbles also eat orange berries ($h_{\text{inclusive}}$)? The same goes for the agent-patient cases. For example, after learning "A kid from the red group stole a cookie from a kid from the blue group," children judge whether the patient of the next theft will also be from the blue group ($h_{\text{restrictive}}$), or whether an agent from the red group will steal from a member of the red group ($h_{\text{inclusive}}$). Note the candidate norms may differ in type (i.e., injunctive or descriptive), but the learning problem does not change. Rational learners should update their beliefs in accordance with the available evidence, no matter whether those beliefs concern injunctive or descriptive norms.

Across both the subject cases and agent-patient cases, then, if children disproportionately favor $h_{\text{restrictive}}$ this would constitute evidence for a group-based bias, since judgments about the norm's content would be disproportionately tied to the previously observed behavior of group members. Recall that under the random standard if the pattern of responses differs from the 50% baseline, then this constitutes evidence for a restrictive bias in learning.

In order to compute an analogous baseline for the evidential standard, we can leverage the size principle (c.f., Tenenbaum & Griffiths, 2001), which dictates the probability of the evidence given a hypothesis ('likelihood') is (1) inversely proportional to hypothesis size and (2) increases exponentially with each additional piece of evidence. In the novel groups paradigm, naïve learners should assume that observing any one behavior is just as likely as observing any other behavior. We can also assume that when naïve learners are presented with two possible behaviors, they believe there are *only* two possible behaviors[3]. From these assumptions, we can assign values to the likelihood of each hypothesis on the basis of the size principle. It is easy to see that with two possible behaviors, $h_{\text{restrictive}}$ is consistent with half of the possible evidence consistent with $h_{\text{inclusive}}$. Thus, a single piece of confirmatory evidence gives $h_{\text{restrictive}}$ twice as much likelihood as $h_{\text{inclusive}}$, since likelihood is inversely proportional to hypothesis size[4].

Now, we have everything we need to compute an evidential baseline. Recall the minimal requirements for rationality dictate the probabilities assigned to each hypothesis update in response to new evidence E via Bayes' Rule:

$$P(h_{res}|E) = \frac{P(E|h_{res})P(h_{res})}{P(E|h_{res})P(h_{res}) + P(E|h_{inc})P(h_{inc})}$$

Since the learners are naïve, they should assign equal prior belief c to each hypothesis. As we discussed, naïve learners should also assign twice as much likelihood L to $h_{\text{restrictive}}$ than $h_{\text{inclusive}}$. Inserting these values back into Bayes Rule allows us to compute the baseline posteriors for $h_{\text{restrictive}}$ and $h_{\text{inclusive}}$:

$$P(h_{restrictive}|E) = \frac{2Lc}{2Lc + Lc} = \frac{2}{3}$$

$$P(h_{inclusive}|E) = \frac{Lc}{Lc + 2Lc} = \frac{1}{3}$$

Thus, naïve rational learners should assign $P \approx 0.66$ to $h_{\text{restrictive}}$ and $P \approx 0.33$ to $h_{\text{inclusive}}$, given the total evidence in the novel groups paradigm. If children are evidentially rational, then on average we should expect children to make inclusive judgments 33% of the time. For the subject cases, this means approving of non-conformity (e.g., saying it is 'okay' for a Hibble to eat an orange berry) 33% of the time. For the agent-patient cases, this means judging that an agent will target a patient from the opposite group as in the learning episode 33% of the time[5] (e.g., a red group member, who stole from a blue group member, will steal from a red group member next). Critically, such patterns of response are biased by definition under the random standard (i.e., differ from 50%), but according to the evidential standard they are epistemically justified.

Altogether, the evidential baseline provides the following definitions of rationality and bias in the novel groups paradigm:

[3] This assumption is already built-in to the agent-patient cases, where agents can direct behavior to one of two groups, and it also plausibly follows from the context provided in the subject cases. All else equal, this assumption also renders the closest evidential baseline to 50%: as the number of behaviors increases, the posterior for $h_{\text{restrictive}}$ increases and the posterior for $h_{\text{inclusive}}$ decreases.

[4] Recall likelihood also increases exponentially with each additional piece of confirmatory evidence. If we consider the pragmatic cues to imply 'more than one', the upshot is the evidential baseline shifts accordingly: restrictive judgments should be more likely, and inclusive judgments should be less likely.

[5] Note: Roberts et al., 2019 (Study 2) is a subtle exception, because the patients of the norm are not members of the opposite novel group (e.g., one norm is "Glerks make babies cry"). The same evidential baseline remains appropriate, but there is a subtle difference in meaning whereby 'inclusive judgment' applies only to the actions of the candidate agent, rather than a complete agent-patient dyad (as in Chalik and Rhodes' agent-patient studies).

Evidentially rational: the average proportion of inclusive judgments does not differ from 33%.

Restrictive bias: the average proportion of inclusive judgments is less than 33%.

Inclusive bias: the average proportion of inclusive judgments is greater than 33%.

In contrast, according to the random standard children's learning is restrictively biased if the average proportion of inclusive judgments is less than 50%. However, note that any evidentially rational, and even some inclusive, patterns of response would be considered biased according to the random standard. Thus, the current interpretation of the literature leaves us in the dark about whether and how exactly children's norm learning is biased (or rational). To shed light on this gap in understanding, next we report a meta-analysis of the known findings to date (c.f., Holland & Phillips, 2020).

Method

Study selection

We have collected 12 studies published between 2014 and 2019, and 2 unpublished studies, containing 40 effect sizes relating to children's norm learning in novel contexts. This meta-analysis is ongoing, and our plan is to extend its scope beyond the studies included here. Out of the current 14 studies in total, 9 studies report effects of children's learning about subject norms (*k* = 26 effect sizes; *n* = 952 in total; ages 4- to 13-years-old, *m* = 7.22 years, *SD* = 2.23 years) and 5 studies report effects of children's learning about agent-patient norms (*k* = 14 effect sizes; *n* = 417 in total; ages 4- to 9-years-old, *m* = 5.03 years, *SD* = 1.22 years). For a complete list of papers in each group, see Appendix 1.

Studies were selected on the basis of the following criteria. First, the studies must involve children's learning about novel, third-party groups. Second, the studies must present children with evidence about a candidate norm's behavioral content. Third, following the learning episode, studies must report a dichotomous measure of children's judgments about the content of the candidate norm. This criterion is necessary to ensure the measures reflect children's learned belief about the norm's contents and allow us to convert the corresponding effect sizes into a common standard (i.e., a raw proportion) across all studies. Studies were selected by searching the literature on children's learning about novel groups and from helpful reviews of related research that were published recently (Chalik et al., 2018; Chalik & Rhodes, 2020).

In summary: all of the included studies involve children making dichotomous judgments about the content of novel, third-party norms on the basis of limited evidence. Thus, the included studies are well-suited for examining whether the mere representation of social groups elicits a restrictive bias in children's norm learning.

Coding and analysis approach

Each study was read and coded by a rater who was not blind to the aims of the meta-analysis (the first author). The studies were coded for: (1) sample size, (2) sample mean age, (3) norm type (i.e., *subject* or *agent-patient*), (4) the valence of the norm's content (i.e., *positive*, *neutral*, or *negative*), (5) the total number of judgments, and (6) the proportion of inclusive judgments (i.e., judgments that extended the norms' behavioral scope beyond the sampled evidence)[6].

For an example, consider Study 1 from Roberts et al., 2017. In this study, children received evidence of the form "Hibbles eat purple berries and Glerks eat orange berries" and made a forced-choice judgment about whether it is okay for a Hibble to eat orange berries (coded: type = *subject*, valence = *neutral*). The authors report collecting judgments from groups of 4- to 6-year-olds ($n = 24$, $m_{age} = 5.37$ years), 7- to 9-year-olds ($n = 31$, $m_{age} = 8.52$ years), and 10- to 13-year-olds ($n = 27$, $m_{age} = 11.39$ years). The authors report each participant making four judgments in total, and, for each group, the corresponding average proportion of inclusive judgments (i.e., proportion of 'It is okay' responses to non-conformity) was $m_{4\text{-}6} = 0.24$, $m_{7\text{-}9} = 0.43$, and $m_{10\text{-}13} = 0.47$.

Our main focus for the meta-analysis is calculating an average point estimate and 95% CI for the proportion of inclusive judgments for each type of norm. To accomplish this, we used a random-effects analysis, which takes into account both within- and between-study variability in the observed effect sizes. In contrast, a fixed-effects analysis only accounts for within-study variability in the observed effect sizes. Accordingly, the random-effects estimates are more appropriate to consider when one's aim is generalizing beyond the already-observed effects, whereas the fixed-effects estimates more narrowly describe the already-observed effects. Here we use the estimates from the random-effects models to anchor our main interpretations, but both types of models yielded similar results (for a summary of the fixed-effects estimates, see: osf.io).

We used the R package *metafor* to conduct all analyses (Viechtbauer, 2010). To satisfy the normality assumption of the models, we used logit-transformed proportions instead of the raw proportions in the analyses. For ease of interpreting, however, we report the back-transformed effect sizes as raw proportions (i.e., simply, the proportion of inclusive judgments).

Results

How does the average proportion of inclusive judgments compare to the evidential baseline of 0.33? For children's

[6] For a complete record of data, code, and analyses, see: osf.io

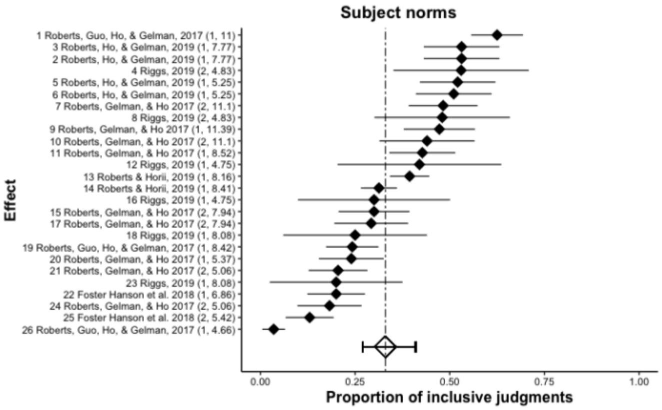

Figure 1: The average proportion of inclusive judgments (diamonds) and 95% CI (lines) for studies involving subject norms. Meta-analytic estimate of the average proportion (diamond) and 95% CI (error bars) is below. The dashed vertical line denotes the evidential baseline for sampled evidence $n = 1$ (note: the baseline shifts downward for all $n > 1$). Effect index (y-axis) includes the publication, study number, and sample mean age.

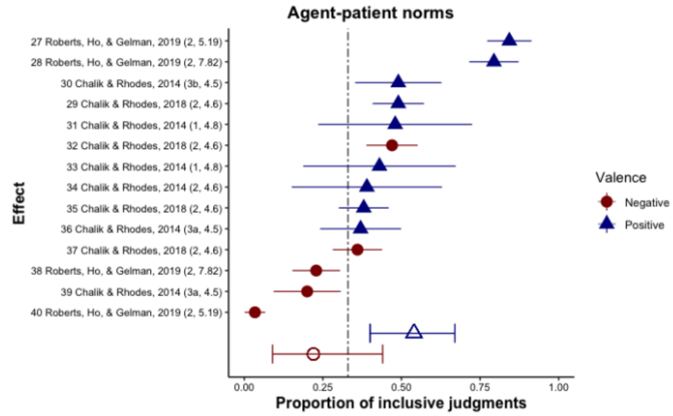

Figure 2: The average proportion of inclusive judgments and 95% CI (lines) for all studies involving agent-patient norms. Red circles denote negative valence behaviors, and blue triangles denote positive valence behaviors. Meta-analytic estimates of the average proportion and 95% CI (error bars) are below. The dashed vertical line denotes the evidential baseline for sampled evidence $n = 1$. Effect index (y-axis) includes the publication, study number, and sample mean age.

judgments about subject norms, the estimate of the average proportion was 0.34, 95% CI [0.27, 0.41]. For children's judgments about agent-patient norms, the estimate of the average proportion was 0.41, 95% CI [0.28, 0.56]. Thus, overall, the meta-analytic estimates of the average proportion of inclusive judgments were not different from the evidential baseline of 0.33. These results support the notion that children's norm learning is not prone to restrictive, group-based bias. Rather, children's responses on the whole are consistent with the hypothesis that they are learning in an evidentially rational way.

However, both analyses found significant heterogeneity between studies (*subject*: $Q(25) = 236.60$, $p < .0001$; *agent-patient*: $Q(13) = 170.09$, $p < .0001$), which indicates that study-specific moderators do exert an influence on the study-specific effects. To get a better sense of potential moderator influences, we conducted a series of moderator analyses using mixed-effects models (Viechtbauer, 2010, p. 4). The moderators that we analyzed were sample size, sample mean age, and valence of the target behavior (dummy coded: 'negative' = 0, 'positive' = 1).

First, we conducted moderator analyses for studies involving subject norms. Sample mean age accounted for a significant proportion of the observed heterogeneity (R^2 = 15.37%), with older children more likely to make inclusive judgments than younger children ($\beta = 0.143$, $SE = 0.07$, $p = .034$). Since these data are not participant-level, however, we emphasize that a healthy dose of caution is warranted before interpreting this result as indicative of a developmental trend. Sample size did not account for a significant proportion of observed heterogeneity ($\beta = 0.004$, $SE = 0.01$, $p = .67$). All studies involving subject norms examined behaviors with neutral valence (e.g., food choice, toy choice, music choice), and thus valence was not included in the moderator analyses for subject norms.

Next, we conducted moderator analyses for studies involving agent-patient norms. Sample mean age ($\beta = 0.144$, $SE = 0.26$, $p = .59$) and sample size ($\beta = -0.005$, $SE = 0.02$, $p = .79$) did not account for a significant proportion of observed heterogeneity. However, the valence of the target behavior did account for a significant proportion of the observed heterogeneity (R^2 = 33.08%), with children more likely to make inclusive judgments about positive-valence behaviors than negative-valence behaviors ($\beta = 1.382$, $SE = 0.531$, $p = .009$). Correspondingly, the meta-analytic estimates for the average proportion of inclusive judgments was 0.54, 95% CI [0.40, 0.67] for positive-valence behaviors and 0.22, 95% CI [0.09, 0.44] for negative-valence behaviors. Thus, compared to the evidential baseline of 0.33, on average children's learning about positive-valence norms was inclusively biased, whereas children's learning about negative-valence norms was statistically appropriate, albeit in a directionally restrictive manner.

Discussion

Across many domains in cognitive development, children are evidentially rational learners: they make statistically appropriate inferences, given their total evidence (c.f., Xu, 2019; Kushnir & Xu, 2012; Perfors et al., 2011). Yet, a prominent view is that children's norm learning is driven by restrictive, group-based biases, even in novel contexts (c.f., Roberts et al., 2017; Chalik & Rhodes, 2020). Upon a closer examination of the methods and evidence that support this interpretation, however, children's norm learning appears to be more rational than commonly thought.

In the ongoing meta-analysis presented here, we find that children's norm learning is not prone to restrictive, group-based biases. Rather, children infer the scope of subject norms at a statistically appropriate rate, given their total evidence. Likewise, we also find that, overall, children infer the scope of agent-patient norms at a statistically appropriate rate. However, it appears the valence of the norm's behavioral content is a key moderator of children's learning. In comparison to a naïve learner faced with the same evidence, children are more likely to make inclusive, wide-scope judgments about norms that govern positive-valence behaviors (e.g., say it is okay to help everyone in a novel population). We also find suggestive evidence that children are more likely to make restrictive, narrow-scope judgments about norms that govern negative-valence behaviors (e.g., say it is not okay to harm everyone in a novel population).

Note this pattern of response is expected if children have already internalized a common characteristic of moral rules (e.g., it is good to direct prosocial, and wrong to direct harmful, actions toward others). Thus, these results may not reflect a learning bias *per se*. More likely, they indicate rational learning given the evidence and non-naïve prior beliefs. As such, the results are consistent with Chalik and Rhodes' claim that children's sophisticated understanding of harm and prosociality is a key factor in their learning about novel norms (c.f., Chalik & Rhodes, 2020, p. 73).

Generally speaking, meta-analysis is a valuable method for quantitatively summarizing the evidence for an effect and highlighting moderators of potential importance (e.g., age and valence of target behavior). However, it is also worth noting that meta-analysis has limitations. If the included studies have not examined a broad range of populations, stimuli, or measures, then meta-analytic results are subject to corresponding limits on generalizability. Here, only one study included in the meta-analysis examined a non-Western population (Roberts et al., 2018), and only one study specifically manipulated the statistical features of children's evidence (Riggs, 2019). These are areas where future research could provide great value. Likewise, incorporating more fine-grained measures of age (e.g., participant-level data) or future longitudinal studies could illuminate the developmental dynamics at play.

For now, our results indicate that, in novel contexts studied to date, children's norm learning is not prone to restrictive, group-based biases. Rather, the evidence is consistent with children learning the scope of novel norms in a statistically appropriate manner—perhaps, even with an inclusive tendency under certain plausible modeling assumptions, given the pragmatic features of the novel groups paradigm (c.f., Tessler et al., 2020; Bonawitz et al., 2011). If children also treat generic and/or pedagogical evidence as more strongly confirmatory than a single independent sample, then the evidential baseline should shift downwards, and the present results would indicate an inclusive bias in norm learning. However, it is not yet clear whether or how exactly these factors exert an influence. As such, the extent to which children's norm learning is strictly rational or inclusively biased is an intriguing topic for future work to address.

In this way, the Bayesian framework leveraged for the present analyses holds promise for future research on norm acquisition. Building from the minimal model specifications used to compute the evidential baseline (essentially, a null hypothesis), there are many possible ways to build and test descriptive models of norm acquisition. For example, more severe tests of the rational learning account can include varying the amount of evidence available to learners and/or varying the group sizes relative to total population (i.e., manipulating the hypothesis space). In an experiment with adult participants (Partington et al., 2020), we found the rational learning account indeed predicts adults' judgments under such manipulations, and an ongoing study suggests children also display a similarly rational pattern of inference. Going forward, the Bayesian framework can also be extended to model the joint influence of pragmatic, stimuli-induced, and individual/cultural differences (see, e.g., Goodman & Frank, 2016; Tauber et al., 2017). All such factors plausibly play an important role in any given instance of norm acquisition and thus plausibly explain the between-study heterogeneity that we found in the meta-analysis.

Even so, the upshot remains that, in novel contexts studied to date, children appear to make roughly the kinds of inferences that they should make, given the evidence (c.f., Nichols, 2021). As such, this calls into question whether mere representations of social groups automatically elicit restrictive biases in norm acquisition (c.f., Roberts et al., 2017a, p. 593; Chalik & Rhodes, 2020, p. 80). Rather, what is made clear by the present analysis is that the inclusive hypothesis is available to children during norm acquisition. In novel contexts, children can—and often do—acquire inclusive norms when the evidence allows. How and why this tendency is amplified or diminished in real-world norm acquisition is rich grounds for future research.

Appendix 1: Studies included

Subject norms: Roberts et al., 2017 (Study 1, 2); Roberts et al., 2018 (1); Roberts et al., 2019 (1); Roberts & Horii, 2019 (1); Foster-Hanson et al., 2018 (1, 2); Riggs, 2019 (1, 2).

Agent-patient norms: Chalik & Rhodes, 2014 (1-4); Chalik & Rhodes, 2018 (2); Roberts et al., 2019 (2).

Acknowledgments

This research was supported by funding from NSF (DS 1823658) and the Evelyn Edwards Milman grant to T.K. Thanks to four anonymous reviewers for helpful comments.

References

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*(3), 322–330.

Chalik, L., Misch, A., & Dunham, Y. (2019). Social cognitive development: The intergroup context. *PsyArXiv*.

Chalik, L., & Rhodes, M. (2014). Preschoolers use social allegiances to predict behavior. *Journal of Cognition and Development*, *15*(1), 136–160.

Chalik, L., & Rhodes, M. (2018). Learning about social category-based obligations. *Cognitive Development*, *48*, 117–124.

Chalik, L., & Rhodes, M. (2020). Chapter Three - Groups as moral boundaries: A developmental perspective. In J. B. Benson (Ed.), *Advances in Child Development and Behavior* (Vol. 58, pp. 63–93). JAI.

Dunham, Y. (2018). Mere membership. *Trends in Cognitive Sciences*, *22*(9), 780–793.

Foster-Hanson, E., Roberts, S. O., Gelman, S. A., & Rhodes, M. (2018). Categories convey normative information across domains and development. *PsyArXiv*.

Goodman, N. D., & Frank, M. C. (2016). Pragmatic language interpretation as probabilistic inference. *Trends in Cognitive Sciences*, *20*(11), 818–829.

Gopnik, A., & Wellman, H. M. (2012). Reconstructing constructivism: Causal models, Bayesian learning mechanisms, and the theory theory. *Psychological Bulletin*, *138*(6), 1085.

Holland, C., & Phillips, J. S. (2020). A theoretically driven meta-analysis of implicit theory of mind studies: The role of factivity. *Proceedings of the 42nd Annual Conference of the Cognitive Science Society*.

Kushnir, T., & Xu, F. (2012). *Rational Constructivism in Cognitive Development* (Vol. 43). Academic Press.

Nichols, S. (2021). *Rational Rules: Toward a Theory of Moral Learning*. Oxford University Press.

Partington, S., Nichols, S., & Kushnir, T. (2020). When in Rome, do as Bayesians do: Statistical learning and parochial norms. *Proceedings of the 42nd Annual Conference of the Cognitive Science Society*.

Perfors, A., Tenenbaum, J. B., Griffiths, T. L., & Xu, F. (2011). A tutorial introduction to Bayesian models of cognitive development. *Cognition*, *120*(3), 302–321.

Riggs, A. E. (2019). Social statistics: Children use statistical reasoning to guide their inferences about the scope of social behavior. *Developmental Psychology*, *55*(1), 66.

Roberts, S. O., Gelman, S. A., & Ho, A. K. (2017). So it is, so it shall be: Group regularities license children's prescriptive judgments. *Cognitive Science*, *41*, 576–600.

Roberts, S. O., Guo, C., Ho, A. K., & Gelman, S. A. (2018). Children's descriptive-to-prescriptive tendency replicates (and varies) cross-culturally: Evidence from China. *Journal of Experimental Child Psychology*, *165*, 148–160.

Roberts, S. O., Ho, A. K., & Gelman, S. A. (2019). The role of group norms in evaluating uncommon and negative behaviors. *Journal of Experimental Psychology: General*, *148*(2), 374.

Roberts, S. O., & Horii, R. I. (2019). Thinking Fast and Slow: Children's Descriptive-To-Prescriptive Tendency Under Varying Time Constraints. *Journal of Cognition and Development*, *20*(5), 790–799.

Talbott, W. (2016). Bayesian Epistemology. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2016). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/win2016/entries/epistemology-bayesian/

Tauber, S., Navarro, D. J., Perfors, A., & Steyvers, M. (2017). Bayesian models of cognition revisited: Setting optimality aside and letting data drive psychological theory. *Psychological Review*, *124*(4), 410.

Tenenbaum, J. B., & Griffiths, T. L. (2001). Generalization, similarity, and Bayesian inference. *Behavioral and Brain Sciences*, *24*(4), 629–640.

Tessler, M. H., Bridgers, S., & Tenenbaum, J. B. (2020). How many observations is one generic worth? *Proceedings of the 42nd Annual Conference of the Cognitive Science Society*.

Vasilyeva, N., & Lombrozo, T. (2020). Structural thinking about social categories: Evidence from formal explanations, generics, and generalization. *Cognition*, *204*, 104383.

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, *36*(3), 1–48.

Xu, F. (2019). Towards a rational constructivist theory of cognitive development. *Psychological Review*, *126*(6), 841.