The Role of Mentalizing in Coordinating Cooperative Behavior and Social Norm Cognition

Paul Deutchman

A dissertation

submitted to the Faculty of

the Department of Psychology & Neuroscience

in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

Boston College

Morrissey College of Arts and Sciences

Graduate School

Boston College

Morrissey College of Arts and Sciences

Graduate School

March 2023

The Role of Mentalizing in Coordinating Cooperative Behavior and Social Norm Cognition

Paul Deutchman

Advisor: Katherine McAuliffe, PhD

Human cooperation is unparalleled in the natural world and is a defining feature of human social life—it shapes nearly every social interaction we experience, from geopolitical conflict, to collective bargaining, to team collaboration. However, cooperation also presents a challenge—it is often personally costly or risky to cooperate. How are humans able to overcome these costs and risks in favor of the interest of the group? To address this question, it is important to investigate the cognitive abilities that allow us to successfully cooperate with others. One important ability for cooperation is mentalizing—the ability to represent other agents' beliefs, knowledge, desires, and intentions. The ability to think about other agents' minds in order to predict how they will behave (e.g., whether they will cooperate or free-ride) is an important component of our own cooperative behavior, particularly in the context of coordination—a type of cooperative interaction where cooperation is mutually beneficial but risky. I test the idea that our ability to represent the beliefs of others plays a critical role in successful cooperation. Studies 1 and 2 examine one cognitive ability for representing others' knowledge—common knowledge—that underlies cooperation by reducing uncertainty about others' cooperative behavior. Studies 3 and 4 investigate how we make inferences about others' beliefs from how they behave and how that influences our own cooperative behavior in the context of social norms. Studies 2 and 4 take a developmental approach to

investigate how early emerging mentalizing is for cooperative behavior to better understand how foundational it is in social cognition. Altogether, the results of these studies suggest that the ability to represent other agents' beliefs in order to predict their behavior plays a fundamental role in supporting successful cooperation.

Table of Contents

List of Tables... iii

List of Figures..iv

Acknowledgments..vi

1. **Introduction** 1

2. **Study 1: Common knowledge increases cooperation in the threshold public goods game** 8

  1. **Introduction** 9

  2. **Experiment 1** 15

  3. **Experiment 2** 28

  4. **Experiment 3** 40

  5. **Open Practices** 58

  6. **References**...58

3. **Study 2: Children use common knowledge to solve coordination problems** ...63

  1. **Introduction** 64

  2. **Method**...70

  3. **Results**...76

  4. **Discussion** 78

  6. **References**...83

4. **Study 3: People update their injunctive norm and moral beliefs after receiving descriptive norm information** 88

  1. **Introduction** 90

  2. **Overview of experiments**...97

  3. **Results** 108

  4. **General Discussion** 122

  5. **Open Data Practices**..130

  6. **References** 130

  7. **Appendix** 136

5. **Study 4: Descriptive Norms Influence Children's Injunctive and Moral Norm Beliefs**..152

1. Introduction ... 154
2. Method ... 164
3. Results ... 172
4. Discussion ... 177
6. References ... 185
6. **General Discussion** ... 192
7. **References** ... 203

List of Tables

Study 1

Table 1. Information-Threshold Conditions – Exp. 1…………………………….…..…....19

Table 2. Information-Threshold Condition Contributions – Exp. 1…………..…...……21

Table 3. Model Estimates – Exp. 1…………………………………………….…..…....26

Table 4. Information-Threshold Condition Contributions – Exp. 2..……………..……32

Table 5. Model Estimates – Exp. 2……………..………………………..….………..34

Table 6. Information-Threshold Condition Contributions – Exp. 3…………..……..…44

Study 2

None

Study 3

Table 1. Design Information for All Studies…………………………………...…..…...99

Table 2. Model Estimates for All Studies …………………………..…………...…112

Study 4

None

List of Figures

Study 1

Figure 1. Contributions by Information Condition – Exp. 1……………………....…....23

Figure 2. Contributions by Information-Threshold Condition – Exp. 1…….……….…25

Figure 3. Contributions by Information Condition – Exp. 2………………..……..…..34

Figure 4. Predicted Contributor Mediation Path Model – Exp. 2……………....…….38

Figure 5. Certainty Mediation Path Model – Exp. 2……………..…………...…...…38

Study 2

Figure 1. Diagram of Coordination Task.………………………………..……….…..75

Figure 2. Coordination by Information Condition. ……………………..…...……..78

Study 3

Figure 1. Plot of Descriptive Norm Condition on Dependent Measures……….……..114

Figure 2. Forest Plot of Meta-Analytic Effects of Descriptive Norm Condition…..…..119

Figure 3. Plot of Descriptive Norm Condition on Injunctive Ratings by Vignette Type……………………………………………………………………………….144

Figure 4. Plot of Descriptive Norm Condition on Behavioral Intentions by Vignette Type……………………………………………………………………………….145

Figure 5. Plot of Descriptive Norm Condition on Other-Morality Ratings by Vignette Type…………………………………………………………….....................146

Figure 6. Plot of Descriptive Norm Condition on Self-Morality Ratings by Vignette Type…………………………………………………………………………..............147

Study 4

Figure 1. Diagram of Norm Task. ……………………………………….....……...169

Figure 2. Results for Dependent Measures by Condition and Vignette Type..…..……176

Acknowledgments

This dissertation would not have been possible without the support and guidance of many people.

First, I'm grateful to my advisor, Katie McAuliffe, who has been the best mentor I could ask for, I owe much of my academic success to her. More than just being a brilliant scientist and academic role-model, Katie is a tremendously kind and generous person who has made my time in graduate school not only productive, but genuinely fun and exciting. Without her guidance and mentorship, I would have never been able to shred the academic gnar.

I'd also like to thank my secondary advisor, Liane Young, and the rest of my dissertation committee, Angie Johnston and Jillian Jordan, for their helpful feedback and insight that not only improved my dissertation, but helped me to think about my research more clearly.

I'm similarly grateful to my undergraduate academic mentors from Skidmore College who helped prepare me to get into and thrive in graduate school, including my undergraduate thesis advisor Jessica Sullivan, Leigh Wilton, and Ian Reed.

I'm also indebted to the members of the Cooperation Lab, past and present, for their feedback, support, and friendship throughout graduate school. I'm likewise grateful to the

members of the Morality Lab. I feel privileged to have been able to work with such brilliant and caring people.

I also want to thank the Boston College Psychology & Neuroscience department, whose commitment to its students and trainees allowed me to succeed in the program and provided an academic home.

Lastly, I'm grateful for the love and support I've received from my friends and family, and my parents in particular, over the past five years.

1. Introduction

Cooperation is a core feature of human social life. The ability to cooperate at a large scale with unrelated individuals has allowed our species to spread across the planet and thrive in even the harshest environments. Yet cooperation presents a number of challenges for social groups. Namely, cooperation is often personally costly—benefiting the group comes at the expense of self-interest—and it's risky—if I cooperate when others do not, I'm worse off than if I did not attempt to cooperate in the first place. Given the importance of cooperative behavior to human survival and reproduction, we have likely evolved cognitive abilities that have allowed us to overcome the challenges inherent in cooperation. Much like how evolution has selected for organs to remove toxins from our bodies or supply our blood with enough oxygen, so has it shaped our cognition to allow our species to successfully cooperate (Tooby, Cosmides, & Barkow, 1992). One important ability in this regard is mentalizing—the ability to represent the minds, desires, and knowledge of other people—which has been extensively studied in the context of theory of mind (ToM; Apperly, 2012; Bockler & Zwickel, 2013; Gallagher & Frith, 2003; Frith & Frith, 2005). While much of the work on mentalizing has centered on ToM, there is another important cognitive ability for representing others' minds: common knowledge. Common knowledge is a heuristic belief state for representing recursive knowledge—A knows X, B knows X, A knows that B knows X, B knows that A knows that B knows X, ad infinitum (Rubinstein, 1981; Thomas et al., 2014). Previous work suggests that both ToM and common knowledge play important roles in supporting cooperative behavior (Rubinstein, 1981; Thomas et al., 2014; Thomas et al., 2018; Tsoi et al., 2021; Tsoi & McAuliffe, 2020). This dissertation will focus on the question of how our beliefs about others' knowledge, beliefs, and behavior—and the cognitive mechanisms that underlie this ability—influences our own cooperative behavior.

When does mentalizing influence cooperative behavior?

While mentalizing plays an important role in social cognition and many different kinds of cooperative actions, it likely plays an especially important role in the context of coordinating behavior. Coordination is a subset of cooperative behavior—all coordination is cooperation, but not all cooperation is coordination. Coordination is distinct from cooperation in that it is risky but mutually beneficial—it is in every agents' self-interest to cooperate—whereas cooperation in a social dilemma is personally costly—the self-interest of the individual and the group conflict (Snidal, 1985).[1] Here, I focus on two types of cooperative interactions in which agents need to coordinate their behavior with others: coordination problems and social norms.[2]

Coordination problems are a specific form of cooperation in which there are multiple stable equilibria, meaning that there are multiple strategies that will result in the best payoff for agents such that they cannot increase their payoff by deviating from their strategy (Schelling, 1960). A classic example of a coordination problem is the stag hunt (Skyrms, 2004). In the stag hunt, agents can jointly hunt a stag (high payoff) or individually hunt a rabbit (low payoff). To successfully hunt the stag, both plays need to

[1] I will occasionally refer to coordination as cooperation since it broadly falls under the umbrella of cooperative behavior (Ashley & Tomasello, 1988).

[2] It's important to note that social norms are not cooperative per se—while social norms play a critical role in maintaining cooperative behavior and coordination, not all norms are cooperative.

cooperate—if either hunts the stag alone, they will fail and receive nothing (e.g., no payoff)—whereas if they hunt the rabbit, they are guaranteed a smaller payoff. In the stag hunt, there are two stable equilibrium: one when both agents cooperate (stag, stag) and another when both agents defect (rabbit, rabbit; Fang et al., 2002). The stag hunt illustrates that despite cooperation being mutually beneficial—it's in the best interest of both agents to jointly hunt the stag— coordination still poses a number of challenges. Perhaps the most important of which is that cooperation in a coordination problem is risky—if you attempt to hunt the stag while your partner hunts the rabbit, your payoff will be nothing and you will be worse off than if you played it safe. In order to predict how other agents will behavior (and specifically whether they will cooperate), it's important to have accurate information about others' belief state (Thomas et al., 2014).

How are humans able to solve these problems in light of these challenges? This is where mentalizing come into play—to infer how others will behave, it is helpful to represent what they know and think. If I know that my group knows what is required to successfully cooperate, then I can be more confident they will also cooperate as compared to if I don't know what my group members know. Previous work suggests that common knowledge (sometimes called *common ground* or *mutual knowledge*; Baltag, Moss, & Solecki, 2016; Bohn & Köymen, 2018; Rubinstein, 1989) is an important ability for representing beliefs about other agent's knowledge and their beliefs about your beliefs (henceforth *higher-order beliefs*; Thomas et al., 2014; Thomas et al., 2018). Namely, this work finds that people are more likely to attempt risky coordination when they and their partner have common knowledge about the payoffs for coordinating (Thomas et al., 2014).

However, while previous work in this area has focused on coordination in the context of dyadic interactions, many cooperative interactions we encounter in our lives involve multiple agents and thus are considerably more complex. While less work has examined this topic, it is likely that higher-order belief representation helps us to solve these problems as well by reducing uncertainty about others' behavior, much like it does in simple coordination problems. One prevalent form of multi-agent coordination where this might be most apparent are social norms. Social norms are the informal rules that govern and constrain behavior in social groups and societies (Bicchieri, 2005), and are a foundational part of human culture that pervade nearly every aspect of human social life—from what we eat for breakfast to how we dress. Social norms facilitate coordination by allowing many individuals to align their expectations about how to behave (Young, 2007; Young, 2015). For example, if a crowd of people are all waiting to get into a bakery, there is a norm that you form a line and when you arrive, you go to the back of the queue. In other words, we all share a common expectation that we should get in line—a norm—which allows us to coordinate our behavior—queuing up. Contrast that with a scenario in which there is no line, just a throng of people pushing past each other to get through the door and you can see the importance of social norms for coordinating behavior. Social norms are especially powerful in situations that are ambiguous and where the appropriate behavior is uncertain (Higgs, 2015; Smith et al., 2007): if you're visiting a foreign country and are unsure whether to tip, you will follow the lead of what other people around you in the restaurant are doing. While norms serve as an important tool for coordinating group behavior, their importance extend beyond coordination. That is, social norms are important for maintaining cooperative behavior more broadly—they often encourage behaviors that are beneficial for the group but costly for the individual. Thus, while social norms play an important role in coordinating group behavior, their significance extends beyond coordination to maintaining cooperation in groups more generally.

Mentalizing and higher-order beliefs likely play an important role in social norm cognition, much like they do for coordination problems like the stag hunt. Namely, one influential account of social norms posits that a defining feature of norms is that they are socially conditional, that is, our compliance with them hinges on our expectation that others condone the behavior and do it as well (Bicchieri, 2005; Bicchieri, 2016). Intuitively, it is easy then to see how important mentalizing is for social norms—we are constantly thinking about what others do and approve of when deciding how to behave. If I'm deciding whether to continue to wear a mask outside after I'm vaccinated, I will think about whether other people in my community approve or disapprove of doing so and I will act accordingly. Given the role mentalizing plays in social norms, this lack of research represents a major gap in our understanding of normative psychology.

Overview of the present research

The goal of this dissertation is to investigate the cognitive abilities and mechanisms that underlie cooperative behavior, particularly in the context of social norms. Studies 1 and 2 examine how our beliefs about others' knowledge and behavior influence cooperative behavior in coordination problems while Studies 3 and 4 explore how we make inferences from others' behavior and how that in turn influences our own beliefs and behavior in the context of social norms.

In **Study 1**, we used the Threshold Public Goods Game (TPGG) in several online experiments with adults to examine the role of common knowledge in underlying cooperative behavior and whether it does so by increasing certainty in the belief that other agents will contribute, thus increasing the likelihood of successful cooperation. We found that common knowledge of the threshold promoted contributions to the public good by decreasing uncertainty around agents' cooperative behavior. In **Study 2**, we used a developmental approach to examine when in ontogeny children begin to use common knowledge to solve coordination problems to better understand how foundational common knowledge is to coordination. We found that by 6-years of age, children are able to use common knowledge to solve coordination problems and that like with adults, it does so by increasing certainty in other agents' cooperative behavior.

In **Study 3**, we used vignettes in a series of online experiments with adults to examine how descriptive norm information about what others commonly do influences our injunctive norm beliefs about what others approve of, moral judgements, and behavioral intentions, and whether this varies depending on the social context. We found that people updated their normative beliefs and behavioral intentions in response to descriptive norm information and that the extent to which they did so varied depending on the behavior such that they generally updated more for fairness and conventional behaviors than harm behaviors and preferences. In **Study 4**, we used a developmental approach to investigate how regularities in the social environment influence different kinds of normative beliefs: are we capable of flexibly tuning those beliefs depending on the frequency and type of behavior? We found that children's injunctive and moral beliefs are influenced by how common or uncommon a behavior is but that this influence does not generalize to all kinds of behaviors, pointing to a special role of social influence on beliefs for behaviors with social consequence. Altogether, these studies highlight the foundational role of mentalizing in underlying cooperative behavior in coordination problems and social norm cognition.

1. Study 1: Common knowledge increases cooperation in the threshold public goods game

Recent work suggests that an important cognitive mechanism promoting coordination is common knowledge—a heuristic for representing recursive mental states. Yet, we know little about *how* common knowledge promotes coordination. We propose that common knowledge increases coordination by reducing uncertainty about others' cooperative behavior. We examine how common knowledge increases cooperation in the context of a threshold public goods game, a public good game in which a minimum level of contribution—a threshold—is required. Across two preregistered studies (N = 5,580), we explored how varying (1) the information participants had regarding what their group members knew about the threshold and (2) the threshold level affected contributions. We found that participants were more likely to contribute to the public good when there was common knowledge of the threshold than private knowledge. Participants' predictions about the number of group members contributing to the public good and their certainty ratings of those predictions mediated the effect of information condition on contributions. Our results suggest that common knowledge of the threshold increases public good contributions by reducing uncertainty around other people's cooperative behavior. These findings point to the influential role of common knowledge in helping to solve large-scale cooperation problems.

This paper is co-authored with Dorsa Amir, Matthew Jordan, and Katherine McAuliffe.

1. Introduction

Cooperation is a key aspect of human social life. While adaptations supporting cooperation are found in many organisms (Clutton-Brock, 2009), cooperation among humans stands out in both its scale and scope. Humans are unique in the extent to which we cooperate with unrelated individuals (Axelrod & Hamilton, 1981; Fehr & Fischbacher, 2004; Rand & Nowak, 2012) and we cooperate in group sizes that are unmatched in the animal kingdom (Clutton-Brock, 2009). While these features of human cooperation have undoubtedly contributed to our success as a species, they also point to a key question: what are the psychological mechanisms that enable humans to solve cooperation problems of this scale?

1.1. Common knowledge

One cognitive mechanism that plays an important role in cooperative behavior is common knowledge (sometimes called mutual knowledge; Baltag et al., 2016; Clark & Marshall, 1981; Halpern & Moses, 1990; Rubinstein, 1981). Common knowledge is the recursive belief state in which A knows X, B knows X, A knows that B knows X, B knows that A knows that B knows X, ad infinitum. Recent work suggests that common knowledge is an important mechanism for coordinating group behavior (Thomas et al., 2014; Thomas et al., 2018). For example, past work has found that people were more willing to attempt risky coordination when there was common knowledge about the mutually beneficial joint payoff for coordination compared to when there was only shared knowledge (such as secondary and tertiary knowledge states; Thomas et al., 2014). These results suggest that common knowledge is likely a distinct cognitive state that may have evolved to solve recurrent problems in human social life (DeFreitas et al., 2019). While previous work suggests that common knowledge plays a role in coordinating behavior (Thomas et al., 2014), no work has explored exactly *how* it does so. What are the mechanisms that underlie the effect of common knowledge on cooperation?

Much of the work on economic games that model cooperation assumes that actors have complete information about the task, that is, they have common knowledge. However, in many circumstances, including more ecologically valid contexts, decision makers lack access to complete information regarding the boundaries and payoffs of the cooperative interaction. Uncertainty about the structure of the task, and importantly, uncertainty about others' knowledge about the task, tends to negatively affect cooperative behavior (Marks & Cronson, 1999; McBride, 2010; Wit & Wilke, 1998). Therefore, a possible mechanism that might explain why common knowledge increases cooperation is that it decreases uncertainty about the social interaction, and specifically, uncertainty about other agents' cooperative behavior. In other words, common knowledge may increase cooperation because it increases certainty that other group members will also contribute when doing so is mutually beneficial, thereby reducing the chance that a cooperative actor will be exploited by others who act selfishly.

1.2. Threshold PGGs

An economic game that is well-suited for studying whether common knowledge increases cooperation is the threshold public goods game. The threshold public goods game is a variant of the public goods game (PGG). In the standard PGG, participants are given an endowment and placed into groups that can vary in size across different instantiations of the game (Fehr & Gachter, 2000; Fischbacher et al., 2001). Participants can contribute any portion of their endowment to a common pot. All contributions to the common pot are then multiplied by the experimenter by a value greater than one and divided equally amongst all group members regardless of their contributions. This game thus captures an important and recurrent dilemma when it comes to cooperation: the conflict between what is best for the individual—freeriding by not contributing anything while others contribute—and what is best for all the members of the group—everyone contributing the entirety of their endowment, resulting in the largest group payoff.

In the threshold PGG, groups must reach a certain level of collective contributions—a threshold—in order for the common pot to be multiplied by the experimenter (Fischbacher et al., 2001). The threshold PGG captures an important feature of many real-life coordination problems—namely, that in many cases, a certain level of contributions must be reached before there are benefits to the initial investment. Take, for example, the case of barbasco fishing, a subsistence practice found among many indigenous Amazonian groups (Heizer, 1953). Barbasco fishing involves the diffusion of a piscicide made from local plants into a river or stream to poison and catch fish. This practice involves multiple contributors who play discrete yet complementary roles, such as building a dam, preparing the barbasco poison, spreading it into the river, herding the fish, and spearing or scooping them. If too few people join in to fulfill the necessary roles, the enterprise will likely be unsuccessful, in which case the initial investment of time and energy will have been wasted. However, it seems probable that if a critical mass of contributors is reached, such that there are enough contributors to fulfill all necessary roles, the chances of success are likely to dramatically increase. This example demonstrates that for certain recurrent social problems, and especially those relevant to our fitness in the evolutionary past, the threshold PGG can be considered a more ecologically valid game than the standard PGG.

Introducing a threshold alters the structure of the game, changing it from a social dilemma—where the interest of the individual is in conflict with the interests of the group—to more of a coordination problem (such as the stag hunt or assurance game; Jansson & Eriksson, 2015; Skyrms, 2004)—where there are multiple stable equilibria (Archetti & Scheuring, 2012). In the case where all members have to contribute to meet the threshold, it is in every actor's best interest to contribute to the public good, but only if the other players contribute as well. However, when the threshold is at an intermediate level, for example when half of the group must contribute to meet the threshold, the game becomes an anti-coordination problem. In an anti-coordination problem, the stable equilibrium is a mixed strategy, such that the best strategy for an actor is to anti-coordinate with other members by withholding their contribution if the other group members contribute, and investing in the public good if other groups members withhold their contributions (Hauert & Doebeli, 2004). Thus, when the threshold is at an intermediate level, the game resembles the volunteer's dilemma, in which there is a strong incentive to free ride but if everyone defects, all players lose (Diekmann, 1985).[3]

The threshold PGG is an ideal economic game to study common knowledge because contributing in the task more closely models coordination than the standard PGG. Since we know that common knowledge increases coordination, it is expected that

[3] Importantly, contributing in the threshold PGG, as it more closely models coordination or anti-coordination problems, is distinct from pure cooperation—in which an actor contributes a benefit at a cost to themselves (West et al., 2007). However, for ease of comprehension and continuity, we describe contributions in the threshold PGG as cooperation in the sense that they confer a benefit to the group.

common knowledge of a threshold will also increase contributions in this task. This is supported by the foundational work of Schelling on coordination problems which suggests that people often rely on focal points—salient features in a coordination problem—to help solve coordination problems (Schelling, 1960). When there is common knowledge, thresholds might constitute a type of focal point that facilitates coordination by reducing uncertainty about others' behavior. In relation to the example of barbasco fishing above, common knowledge that a certain number of contributors is necessary to catch fish might reduce uncertainty about whether other people will contribute, allowing individuals to coordinate on the mutually beneficial outcome.

To date, no work has explicitly investigated common knowledge in the threshold PGG, nor the mechanism through which common knowledge increases coordination. We predict that common knowledge will increase cooperation in this task by reducing uncertainty about whether your group members will also contribute, and thus whether players will succeed in reaching the threshold. Putting ourselves in the mind of a player, the logic is as follows: when I know that we all know the threshold (and that everyone knows that everyone knows), I can be more confident that everyone will contribute, which will in turn make me more likely to contribute myself. Alternatively, it is possible that our beliefs about others' cooperative behavior, and our certainty in those beliefs, will not influence our own cooperative behavior. In other words, people may behave cooperatively or selfishly without regard to other agents' knowledge about the threshold or beliefs about how they will behave. This question has important implications for real-world cooperation problems as it is often the case that we are uncertain about what others know and will do in cooperative endeavors. That we do not yet know whether certainty mediates the effect of common knowledge on coordination represents an important gap in our understanding of social cognition and cooperation. Furthermore, while previous work has examined common knowledge in the context of a 2-player stag hunt game (Thomas et al., 2014)—a coordination problem in which the stable strategy is mutual cooperation or defection—no work has examined common knowledge in *n*-player cooperation problems which more closely model the kinds of cooperation problems we encounter in everyday life.

1.3. Present study

In three studies, we explored how common knowledge and threshold levels influenced contributions in a threshold PGG and whether the effect of common knowledge of the threshold on contributions is mediated by certainty about others' cooperative behavior. In Experiment 1, we tested this by manipulating 1) the information group members knew regarding the threshold, and 2) the level of threshold needed to receive the public good. Because past work has found mixed results regarding the effect of threshold size on contributions (Andrews et al., 2019; Cadsby & Maynes, 1999), we varied the threshold level in our studies to examine if threshold size predicts contributions and to explore whether the effect of common knowledge varies by threshold size or is robust across different sized thresholds. If thresholds promote cooperation, as found previously, we would expect contributions to be higher in all threshold conditions than in the baseline PGG that lacks a threshold. If common knowledge allows individuals to coordinate contributions in the PGG, then we would expect the highest levels of contributions in the common knowledge condition. In our second preregistered study (Experiment 2), we aimed to replicate our findings from Experiment 1 and to test whether the effect of common knowledge on contributions in the threshold PGG is mediated by certainty about the predicted number of group members contributing to the public good. If common knowledge of the threshold increases contributions by reducing uncertainty around the cooperative behavior of others, then we expect to find that contributions in the PGG are mediated both by the predicted number of group members contributing, and certainty about those predictions. Lastly, in a third preregistered study, we aimed to replicate the results of Experiment 2 and to better understand contribution behavior under common ignorance, in which participants know there is a threshold but do not know what it is, to determine whether certainty about the presence of a threshold might explain the contributions levels.

2. Experiment 1

2.1. Method

Participants

We tested N = 2,252 participants (52.35% female), aged 18-77 (M = 36.93) from Amazon's Mechanical Turk in a preregistered study (http://aspredicted.org/blind.php?x=63ct9y). Our sample size was based on previous work on threshold public good games (Jordan et al., 2017) using the same platform. While we initially recruited 3,365 participants, 475 were excluded from analysis for failing to complete the study in its entirety, 618 were excluded for failing the comprehension questions, and an additional 20 were excluded for responding with "three or more" to a question assessing the number of questions they answered without reading or thinking about them carefully. The high rates of exclusions here reflect our stringent exclusion criteria for the several comprehension checks participants had to answer (see preregistration for exclusion criteria).

Design

Study 1 was a 3 × 3 between-subjects design in which participants were randomly assigned to one of three information conditions (*common knowledge, common ignorance, and private knowledge*) and one of three threshold levels (*low, high, and maximum*) or a baseline condition with a standard PGG with no threshold. Thus, in total, participants were assigned to one of ten conditions. We used ex-post matching to randomly assign participants to groups of four and determine their group contributions after data collection (Horton, Rand, & Zeckhauser, 2011; Jordan, McAuliffe, & Rand, 2016).

Procedure

Participants provided informed consent and were given instructions for the task that detailed the rules of the PGG. Participants were assigned to groups of six and allocated an endowment of $0.30 each for the PGG. They were informed of their group size, that all contributions to the public pot would be multiplied by two and divided evenly amongst group members, and they would receive this endowment as a bonus after completing the game.

After reading the instructions, participants answered three comprehension questions to ensure their understanding of the game (see **Supplement**). Participants who failed the comprehension checks after two attempts were excluded from analyses. After answering these questions, participants then read the specific instructions for the threshold information condition to which they were assigned. The instructions explaining the threshold were identical but differed in two key respects: 1) the level of threshold needed to receive the public good, and 2) the information their group members knew regarding the threshold. The threshold level manipulation had three levels, a *low threshold, a high threshold, and a maximum threshold*. In the *low threshold* condition, 2 out of 6 participants in a group needed to contribute their $0.30 endowment in order for contributions to be multiplied and split amongst their group members. In the *high threshold* condition, 4 out of 6 participants in a group needed to contribute. In the *maximum threshold* condition, all 6 out of 6 participants in a group needed to contribute. Contribution decisions were binary, participants could either contribute the entirety of their $0.30 endowment or not contribute. Critically, in our threshold PGG, if participants failed to reach the threshold then all contributions to the common pot were destroyed, leaving them with only the portion of their endowment they did not contribute.

The threshold information manipulation had three conditions: *common knowledge*, *common ignorance*, and *private knowledge*. In the *common knowledge* condition, participants were told that everyone in their group saw the same instructions they did, and that everyone in their group knew that the participant had seen the same instruction as well. Thus, everyone in their group knew that at least 4 out of their 6 group members needed to contribute (or 2 and 6, in the low and maximum threshold treatments, respectively) or the total common pot would be destroyed. This established something at least broadly consistent with a recursive belief state regarding the threshold such that the participant knew the threshold, knew everyone in their group knew the threshold, and knew that everyone in their group knew that they knew the threshold, ad infinitum. The *common ignorance* condition was identical to the *common knowledge* condition but, in this condition, the threshold was unknown (e.g. “the amount you must contribute is unknown to your group”). Thus, the participant did not know the threshold, everyone in their group did not know the threshold, and the participant knew that everyone in their group knew that they did not know the threshold, ad infinitum. We included this condition to investigate whether the presence of a threshold, even when unknown, would be enough to promote contributions in the PGG. In the *private knowledge* condition, participants were told the threshold level but that they could not be certain that their group members saw the same instructions as them, thus resulting in a lack of common knowledge (e.g. “only you know for certain that at least 4 out of your 6 group members must contribute”). Thus, the participant knew the threshold, but they were not sure if everyone in their groups knew the threshold, and everyone in their group did not know if they knew the threshold. After reading the information threshold instructions, participants answered three more comprehension questions for their specific information and threshold condition. After answering these comprehension questions, participants then made their contribution decision.

Participants could contribute their entire endowment or nothing (0 to 30 cents) to their group pot. After making their contribution decision, participants were asked to predict how many other group members would contribute to the public good collectively (0 to 5 contributors). We included this question to assess whether participants’ predictions of their group members’ contributions influenced their own contribution decisions. In order to compare the level of cooperation observed in the information threshold conditions to a standard PGG, we also ran a baseline condition in which participants played an identical PGG without a threshold. We report all measures, manipulations, and exclusions (see supplementary materials for measures not reported here).

Table 1. Table displaying all combinations of threshold level and information conditions.

Information Condition	**Low Threshold**	**High Threshold**	**Maximum Threshold**
Common Knowledge	Everyone knows that everyone knows the threshold is 2 out of 6 people	Everyone knows that everyone knows the threshold is 4 out of 6 people	Everyone knows that everyone knows the threshold is 6 out of 6 people
Common Ignorance	Everyone knows that everyone knows there is a threshold of unknown size		
Private Knowledge	Only I know with certainty the threshold is 2 out of 6 people	Only I know with certainty the threshold is 4 out of 6 people	Only I know with certainty the threshold is 6 out of 6 people
Baseline	No threshold		

Analysis

We ran three pre-registered logistic regression models with contribution (binary: 0 = did not contribute, 1= did contribute) as the response term. To determine whether information type influenced cooperation, the first model included information (baseline, common knowledge, common ignorance, private knowledge) as the predictor variable (see Information column in Table 3). To determine whether there was an interaction between threshold level and information type, the third model included the interaction between information (private knowledge, common knowledge) and threshold (low, high, max), as well as information condition (private knowledge, common knowledge) and threshold level (low, high, max) as the predictor terms (see the Information × Threshold column in Table 3). We left out the common ignorance and baseline conditions from this model because threshold level did not vary across baseline or common ignorance conditions, preventing us from examining the interaction between threshold and information. We also ran a third preregistered model predicting contribution decision by threshold level which we report in the Supplement. To determine whether our predictors explained more variance than a null model, we compared the model with the threshold-information condition interaction, plus age and gender terms, to a model only including age and gender. The model with the interaction term explained significantly more variance than the model without it ($\chi2$ (5) = 17.03, *p* = .004).

For all models, we made specific comparisons within information and threshold conditions by using a series of pre-registered pairwise comparisons using estimated marginal means adjusted using the multivariate t method (MVT) to correct for multiple comparisons. These pairwise tests allowed us to make the critical comparison between the common knowledge and private knowledge conditions in order to determine whether common knowledge increased contributions. We next ran two exploratory models that were not preregistered. To test whether participant's predictions about the number of their group members contributing predicted their own contributions, we ran a logistic regression model with contribution as the response term and predicted number of other group members contributing (continuous: 0-5) as a predictor. We also ran an identical model but included information condition (private knowledge, common knowledge, common ignorance) as a predictor to determine whether it would predict contributions when controlling for predicted group member contributions.

Lastly, in an exploratory model, we examined whether predicted contributors mediated the effect of information condition on contributions by creating a path analysis model with contribution (binary: 0, 1) as the endogenous variable, information condition (private knowledge, common knowledge) as the exogenous variable, and predicted contributors (continuous: 0-5) as the mediator. We used bootstrapping with 5,000 iterations to find standard errors, bias-corrected bootstrapped confidence intervals with 5,000 samples, and diagonally weighted least squares (DWLS) to estimate the model parameters.

Table 2. Proportions and standard deviations (in parentheses) of contributions in the PGG by information and threshold level.

	Low	**High**	**Maximum**	
Baseline	0.5 (0.5)			
Common Knowledge	0.78 (0.41)	0.72 (0.45)	0.78 (0.41)	0.76 (0.42)
Common Ignorance	0.74 (0.44)	0.73 (0.45)	0.71 (0.45)	0.73 (0.45)

Private Knowledge	0.70 (0.46)	0.72 (0.45)	0.65 (0.48)	0.69 (0.46)
	0.74 (0.44)	0.72 (0.45)	0.71 (0.45)	

2.2. Results

Planned Analyses

Overall, we found that information level predicted contributions: participants in the common knowledge (B = 1.15, SE = .17, *p* < .001, OR: 3.15, 95% CI: 2.27, 4.37), common ignorance (B = 0.96, SE = .16, *p* < .001, OR: 2.61, 95% CI: 1.89, 3.61) and private knowledge (B = 0.76, SE = .16, *p* < .001, OR: 2.14, 95% CI: 1.55, 2.96) conditions were all significantly more likely to contribute than participants in the baseline condition (see Table 2 for means). Our findings replicate previous work on thresholds, suggesting that people were more likely to contribute to the public good when there was a threshold. Critically, the comparisons between information conditions revealed that participants were significantly more likely to contribute when there was common knowledge of the threshold than when there was private knowledge of the threshold (B = -0.39, SE = .12, *p* = .009). Participants in the common ignorance condition were not more likely to contribute than participants in the private knowledge condition (B = 0.19, SE = .12, *p* = .35) or those in common knowledge condition (B = -0.19, SE = .12, *p* = .42).

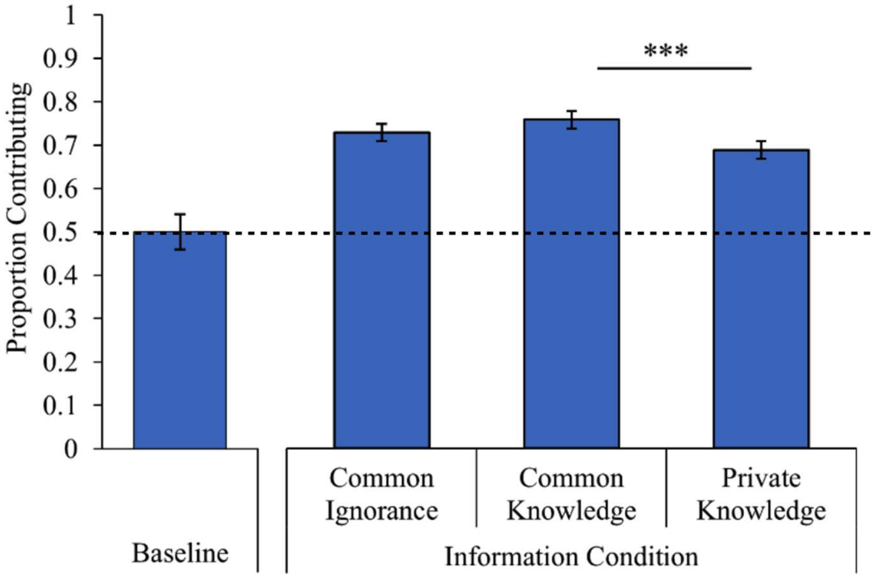


Figure 1. Proportion of participants contributing in the PGG across the three information conditions (common ignorance, common knowledge, private knowledge), and the baseline, non-threshold game. All threshold PGGs elicited more contributions than the baseline condition which is indicated by the dotted line. Error bars indicate standard error. ***p<.001.

When predicting contribution by threshold level, information condition, and their interaction, we found that the interaction between threshold (low and high) and information (common knowledge and private knowledge) was not significant (B = -0.45, SE = .29, p = .13, OR: 0.63, 95% CI: 0.35, 1.14). The interaction between threshold (low and max) and information (private knowledge and common knowledge) was also not significant (B = 0.24, SE = .30, p = .43, OR: 1.27, 95% CI: 0.70, 2.29). However, the significance of this interaction term hinged on the reference condition for the threshold variable: when we set the reference category to the high condition, the interaction between threshold (high and max) and information (private knowledge and common knowledge) was significant (B = 0.69, SE = .29, *p* = .02, OR: 1.99, 95% CI: 1.11, 3.57). The preregistered pairwise comparisons between the private and common knowledge conditions within each threshold level revealed that within the low threshold condition, participants were significantly more likely to contribute in the common knowledge condition than the private knowledge condition (B = -0.46, SE = .22, *p* = .033). Similarly, within the max threshold condition, participants were significantly more likely to contribute in the common knowledge condition than the private knowledge condition (B = -0.69, SE = .21, *p* = .001). Within the high threshold condition, there was no difference in participants' likelihood of contributing between the common knowledge and the private knowledge conditions (B = -0.01, SE = .21, *p* = .98).

Next, we found that the more group members participants predicted would contribute, the more likely participants were to contribute themselves (B =1.30, SE = .06, p < .001, OR: 3.69, 95% CI: 3.27, 4.18). When controlling for predictions about other group members contributing, information condition ceased to predict contributions; participants were not more likely to contribute when there was common knowledge than private knowledge (B = 0.19, SE = .16, p = .21, OR: 1.22, 95% CI: 0.89, 1.66) or when there was common ignorance than private knowledge (B = 0.23, SE = .15, p = .13, OR: 1.26, 95% CI: 0.94, 1.69).

Lastly, we found that the predicted number of group contributors fully mediated the effect of information condition on contributions. The total effect of information condition on contributions was significant (*b* = 0.23, SE = 0.07, *p* <.001), while the direct effect of information condition on contributions was not significant (*b* = .07, SE = .06, *p* = .27). The path from information condition to predicted contributors was significant (*b* = 0.325, SE = .08, *p* < .001), with information condition explaining 11.4% of the variance in the number of predicted group members contributing (see Table S10 in the Supplement for model output). The path from predicted contributors to PGG contribution (*b* = 0.51, SE = 0.01, *p* < .001) was also significant, with the predicted number of contributors explaining 71.8% of the variance in contributions in the PGG. Critically, the indirect effect was significant (*b* = 0.16, SE = 0.04, *p* <.001), explaining 8.2% of the total variance, with the bias-corrected bootstrapped confidence interval with 5,000 samples above zero (95% CI: 0.09, 0.24).

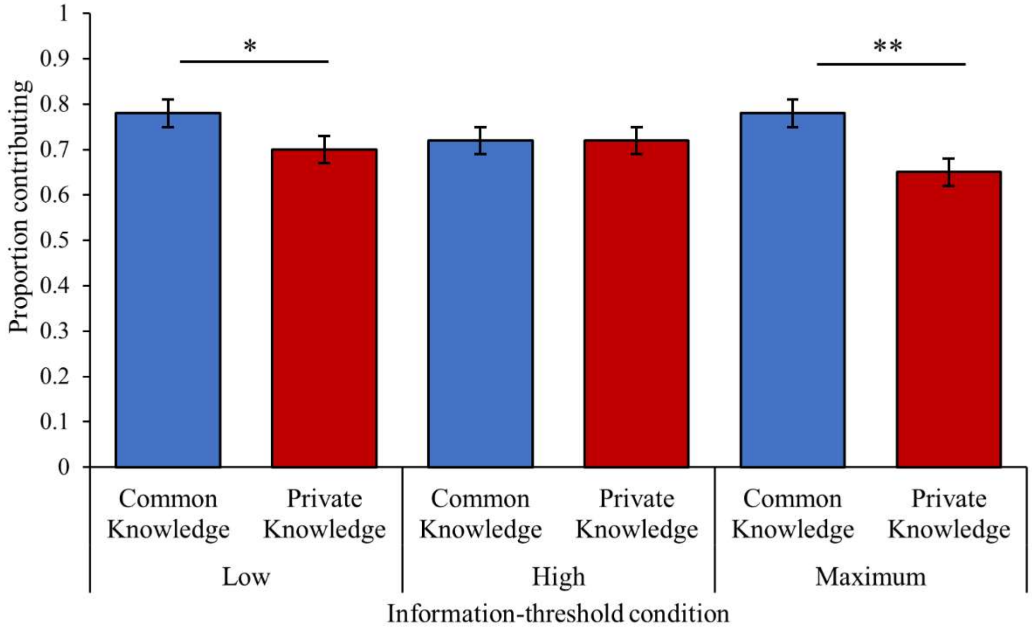


Figure 2. Proportion of participants contributing in the PGG between the common knowledge and private knowledge conditions within the low, high and maximum threshold levels (low: 2 out of 6, high: 4 out of 6, maximum: 6 out of 6). Common

knowledge elicited more contributions only within the low and maximum threshold treatments. Error bars indicate standard error. **p*<.05; ***p*<.01.

Table 3. Estimate and standard error of fixed effects in logistic regression models predicting contribution to the public good. The baseline condition was set as the reference category for the information and threshold models. For the information and threshold interaction model, the reference categories were set as follows: Threshold – Low, Knowledge – Common Knowledge.

	Information	**Information × Threshold**
(Intercept)	0.02 (0.14)	0.82 (0.15)***
Common Ignorance	0.96 (0.17)***	
Private Knowledge	0.76 (0.16)***	
Common Knowledge	1.15 (0.17)***	0.46 (0.22)*
High Threshold		0.11 (0.21)
Max Threshold		-0.22 (0.20)
Low Threshold		
High Threshold × Common Knowledge		-0.45 (0.30)
Max Threshold × Common Knowledge		0.24 (0.30)
AIC	2684.56	1601.93

BIC	2707.44	1633.25
Log Likelihood	-1338.28	-794.96
Deviance	2676.56	1589.93
Num. obs.	2252	1367

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

2.3. Discussion

Consistent with our predictions, we found that participants contributed more to the public good when there was common knowledge of the threshold compared to when there was only private knowledge. We also found tentative evidence that the effect of common knowledge was at least partly due to increased certainty in the number of group members contributing. Additionally, we replicated past work showing that thresholds increase cooperation in the public goods game; contributions were higher across all threshold levels relative to the baseline condition that lacked a threshold (Jordan et al., 2017; Szolnoki & Perc, 2010; van de Kragt et al., 1983). When we examined the interaction between threshold level and information condition, we found that participants contributed significantly more when there was common knowledge, but only in the low and maximum threshold conditions. This potentially suggests that the effect of common knowledge on cooperation depends on the threshold level. However, because we did not predict an interaction between common knowledge and threshold at the outset, we are reluctant to further interpret this finding. We attempt to replicate this interaction in

Experiment 2 to better understand whether the effect of common knowledge does in fact hinge on the threshold level.

A key finding from this study was that the effect of common knowledge on contributions might have resulted from increased certainty regarding whether group members would contribute. Specifically, we found that the expected number of group members contributing predicted participant's own contribution decisions and when controlling for predictions about how many group members would contribute to the public good, information condition no longer predicted contributions. Furthermore, a mediation analysis revealed that the predicted number of contributors fully mediated the effect of information condition on contributions. These findings provide initial evidence that the increased number of contributions in the common knowledge condition may have been a result of decreased uncertainty about group member's cooperative behavior. These results support our prediction that certainty serves as a mechanism underlying the effectiveness of common knowledge on coordination. In Experiment 2, our aim was, first, to extend our findings from Experiment 1 by explicitly investigating whether the effect of common knowledge on coordination is mediated by certainty about group members' cooperative behavior, and second, to replicate our findings from Experiment 1.

3. Experiment 2

3.1. Method

Participants

We tested N = 1859 participants (56.7% female), aged 18-91 (M = 40.7) from Amazon's Mechanical Turk in a preregistered study (https://osf.io/brqky/?view_only=ab7c7982f7454e439ca63c5806c00d52). This sample size was based on that used in previous work in threshold PGGs conducted on Mechanical Turk (Jordan et al., 2017). While we initially recruited 2,866 participants, we excluded 447 from analysis for failing to complete the study in its entirety, 527 for failing the comprehension questions, and 18 for responding with "three or more" to a question assessing the number of questions they answered without reading or thinking about them carefully. Additionally, 15 were excluded for completing the survey more than once (we included first responses) or for completing the survey after an incomplete attempt in which they were exposed to an experimental condition.

Design & Procedure

In Experiment 2, we focused on a subset of the most interesting threshold and information conditions. To replicate the effect of common knowledge on cooperation, we included the two information conditions in which common knowledge of the threshold was present (common knowledge) or absent (private knowledge). Additionally, because the effect of common knowledge differed by threshold level in Study 1, we included a threshold level in which the effect was the strongest (maximum) and one in which it was entirely absent (high). To capture the baseline level of cooperation, we again included a condition without a threshold. In total, participants were assigned to one of five conditions.

The procedure of Experiment 2 was identical to Experiment 1 in every respect (i.e. the game and threshold information manipulations were the same) with the exception of the following differences. After participants read the PGG and threshold information instructions and answered the comprehension questions, they either first made their PGG contribution decision followed by their prediction about the number of group members contributing, or first made their prediction about the number of contributors followed by their PGG contribution. The order of dependent variables was counterbalanced to control for order effects. Unlike Experiment 1, the question assessing participants' predicted number of contributors was incentivized, such that participants received $0.10 for correctly predicting the number of their group members who contributed to the public good. Directly after answering this prediction question, participants answered a question assessing their certainty about their response (0-100 sliding scale, anchored from 0-"not certain at all" to 100-"extremely certain"). Previous work has used a similar self-report approach to measure certainty and confidence (Balakrishnan & Ratcliff, 1996; Bradley, 1981; Thunström et al., 2015). After completing both blocks of dependent variables, participants answered an exploratory question that elicited individual-level certainty distributions for the number of their group members contributing.

Analyses

To replicate our analysis from Experiment 1, we ran three pre-registered logistic regression models with contribution (binary: 0 = did not contribute, 1 = did contribute) as the response term. To determine whether information condition influenced contributions, the first model included information (baseline, private knowledge, common knowledge) as the predictor variable (see the Information column in Table 4). To determine whether there was an interaction between information condition and threshold level, the third model included information (private knowledge, common knowledge), threshold (high, maximum), and the interaction between information and threshold (see the Information × Threshold column in Table 4). We also ran a third preregistered model predicting contributions by threshold which we report in the Supplement. For all three models we made comparisons within information and threshold condition with a series of preregistered pairwise comparisons using estimated marginal means adjusted using the multivariate t method (MVT) to correct for multiple comparisons. For all models, we dummy coded the information condition and threshold level categorical predictors, setting private knowledge and high threshold conditions as the reference categories, respectively. As in Experiment 1, we compared the model with the interaction term, in addition to gender and age, to a null model without the interaction term. The model with the interaction term explained significantly more variance in contributions than the null model ($\chi2(3) = 9.19$, $p = .027$).

To determine whether the distributions of predicted contributors and certainty ratings differed between the private and common knowledge conditions, we ran two, preregistered two-sample bootstrap Kolmogorov-Smirnov (KS) tests. Deviating from our pre-registration, we used the *ks.boot* command in the "Matching" package (Sekhon, 2011) to run bootstrapped KS tests with 5000 iterations in order to handle the ties in our data. We next ran another set of five pre-registered logistic and linear regression models to assess the relationship between information condition (private knowledge, common knowledge), threshold level (high, maximum), predicted contributors (continuous: 0-5), certainty scores (continuous: 0-100), dependent variable order (PGG contribution, PGG contribution prediction), and PGG contributions (binary: 0 = did not contribute, 1 = did contribute). We report the results of these models in the SOM but note here that they conform with our predictions indicating there was no order effect of dependent variable.

To examine whether the participants' predictions about the number of contributors and their certainty about those predictions mediated the effect of common knowledge on contributions, we ran two, pre-registered structural equation models to test for indirect effects. To determine whether predicted contributors mediated the effect of information condition on contributions, we created a path analysis model with contribution (binary: 0, 1) as the endogenous variable, information condition (private knowledge, common knowledge) as the exogenous variable, and predicted contributors (continuous: 0-5) as the mediator. To determine whether certainty ratings for predicted contributors mediated the effect of information condition on contributions, we created a path analysis model with contribution (binary: 0,1) as the endogenous variable, information condition (private knowledge, common knowledge) as the exogenous variable, and certainty ratings (continuous: 0-100) as the mediator. For both models, we used bootstrapping with 5,000 iterations to find standard errors, bias-corrected bootstrapped confidence intervals with 5,000 samples, and diagonally weighted least squares (DWLS) to estimate the model parameters.

Table 4. Proportions and standard deviations (in parentheses) of contributions in the PGG by information and threshold level.

	High	**Maximum**	

Baseline	0.66 (0.48)		
Common Knowledge	0.78 (0.41)	0.81 (0.39)	0.80 (0.40)
Private Knowledge	0.72 (0.45)	0.76 (0.43)	0.74 (0.44)
	0.75 (0.43)	0.78 (0.41)	

3.2. Results

Replication of Experiment 1 Results

Replicating results from Experiment 1, we found that information condition predicted contributions: participants were more likely to contribute in the private knowledge (B = 0.38, SE = 0.14, *p* = .006, OR: 1.46, 95% CI: 1.55, 2.71) and common knowledge conditions (B = 0.72, SE = 0.14, *p* <.001, OR: 2.05, 95% CI: 1.55, 2.71) than in the baseline condition (see Table 4 for means). Critically, the within information condition comparison found that participants were significantly more likely to contribute when there was common knowledge than when there was private knowledge (B = 0.34, SE = 0.12, *p* = .017).

When predicting contributions by threshold level, information condition, and their interaction, we found that participants were significantly more likely to contribute when there was common knowledge than private knowledge (B = 0.35, SE = 0.17, *p* = .038, OR: 1.42, 95% CI: 1.02, 1.98). Participants were not more likely to contribute when the threshold was maximum than high (B = 0.19, SE = 0.17, *p* = .263, OR: 1.21, 95% CI: 0.87, 1.68). The interaction between information condition and threshold level was also not significant (B = -0.04, SE = 0.25, *p* =.86, OR: 0.96, 95% CI: 0.59, 1.55), suggesting that the difference between information condition did not vary by threshold level. The pairwise comparisons revealed that within the high threshold level, participants were significantly more likely to contribute in the common knowledge than private knowledge condition (B = 0.35, SE = 0.17, *p* = .038). Within the maximum threshold level, the difference in contributions between when there was common knowledge compared to private knowledge was trending on significance (B = 0.31, SE = 0.18, *p* = .08).

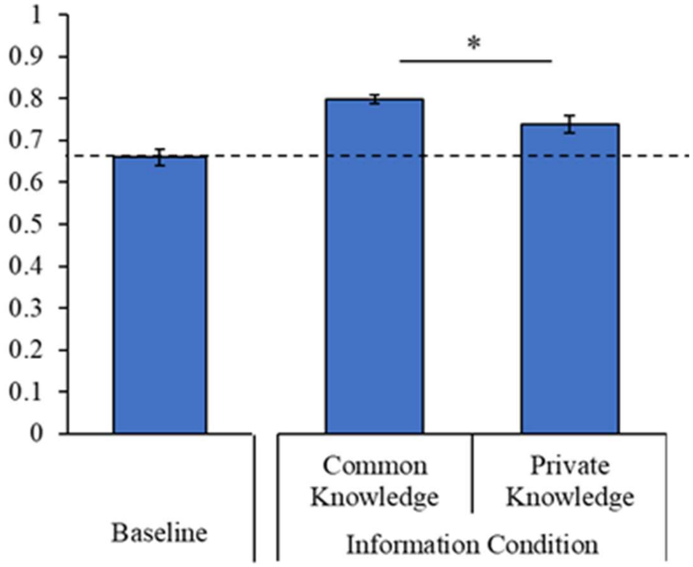


Figure 3. Proportion of participants contributing in the PGG across the two information conditions (common knowledge, private knowledge), and the baseline, non-threshold game. All threshold PGGs elicited more contributions than the baseline condition, which is indicated by the dotted line. Error bars indicate standard error. **p*<.05.

Table 5. Estimate and standard error of fixed effects in logistic regression models predicting contribution to the public good. The baseline condition was set as the reference category for the information and threshold models. For the information and threshold

interaction model, the reference categories were set as follows: Threshold – High, Knowledge – Common Knowledge.

	Information	**Information × Threshold**
(Intercept)	0.64 (0.11)***	0.94 (0.11)***
Common Knowledge	0.72 (0.14)***	0.35 (0.17)*
Private Knowledge	0.38 (0.14)**	
Max Threshold		0.19 (0.17)
High Threshold		
Max Threshold × Common Knowledge		-0.04 (0.25)
AIC	2091.22	1623.84
BIC	2107.80	1645.08
Log Likelihood	-1042.61	-807.92
Deviance	2085.22	1615.84
Num. obs.	1859	1496

Additional analyses

The distributions of predicted contributors differed significantly between the common knowledge and private knowledge conditions (D(1859)=0.15, $p < .001$). A qualitative appraisal of the distributions suggests that the common knowledge distribution was skewed towards a higher number of predicted contributors than the private knowledge distribution (see Supplement for histograms of predicted contributors and certainty ratings). When examining whether the distribution of certainty ratings differed between common knowledge and private knowledge conditions, we found that the distributions differed significantly (D(1859)=0.13, $p < .001$). Again, a qualitative examination of the distributions suggests that the common knowledge distribution was skewed towards higher certainty ratings than the private knowledge condition.

Mediation

We found that the predicted number of group contributors fully mediated the effect of information condition on contributions (see Figure 4 for a path diagram). The total effect of information condition on contributions was significant (b = 0.196, SE = 0.07, p = .007), while the direct effect of information condition on contributions was not significant (b = .038, SE = .055, p = .487). The path from information condition to predicted contributors (b = 0.30, SE = .07, p < .001) was significant, with information condition explaining 10.9% of the variance in the number of predicted group members contributing (see Table S11 in the Supplement for model output). The path from predicted contributors to PGG contribution (b = 0.52, SE = .016, p < .001) was also significant, with the predicted number of contributors explaining 72.5% of the variance in contributions in the PGG. Critically, the indirect effect was significant (*b* = -0.16, SE = 0.04, *p* <.001), explaining 7.8% of the total variance, with the bias-corrected bootstrapped confidence interval with 5,000 samples above zero (95% CI: 0.07, 0.23).

Next, we found that certainty about the number of predicted contributors fully mediated the effect of information condition on contributions (see Figure 5 for a path diagram). The total effect of information condition on contributions was again significant (*b* = .196, SE = 0.07, *p* = .006) but the direct effect of information condition on contributions was not significant (*b* = 0.08, SE = 0.07, *p* = .242). The path from information condition to certainty ratings (*b* = 8.62, SE = 1.37, *p* < .001) was significant, with information condition explaining 16.2% of the variance in certainty ratings (see Table S12 in the Supplement for model output). The path from certainty ratings to PGG contributions (*b* = 0.013, SE = .001, *p* < .001) was also significant, with certainty explaining 34.5% of the variance in contributions to the public good. Importantly, the indirect effect was significant (*b* = 0.11, SE = .02, *p* < .001), explaining 5.6% of the total variance, with the bias-correct bootstrapped confidence interval with 5,000 samples above zero (95% CI: 0.08, 0.16). As pre-registered, we replicate the results of these mediation analyses using a different approach (see supplement). We note here that the results do not differ in interpretation.

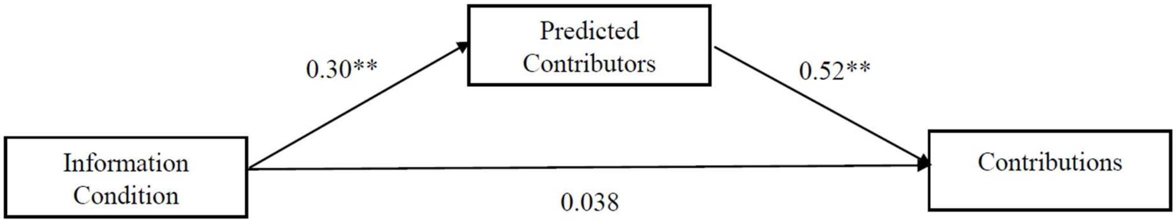


Figure 4. Diagram of the path analysis model with predicted contributor as a mediator. **$p < .01$.

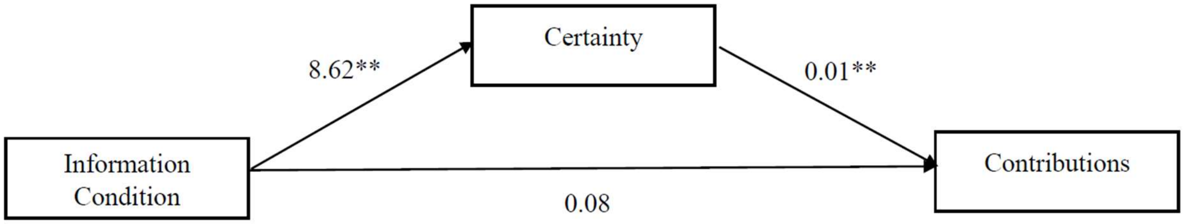


Figure 5. Diagram of the path analysis model with certainty as a mediator. **$p < .01$.

3.3. Discussion

The central aim of Experiment 2 was to test the mediating role of certainty and predicted group contributors on PGG contributions. Results of our path analyses provide compelling evidence that the effect of common knowledge on increased contributions was mediated by the predicted number of group members contributing and certainty ratings about these predictions. Additionally, the KS tests suggested that the distributions of predicted contributors and certainty ratings were significantly different and were more negatively skewed in the common knowledge condition compared to the private knowledge condition. This suggests that when there was common knowledge, participants were more likely to think that more of their group members would contribute and were more certain of those predictions. Overall, the results of these analyses provide strong evidence that common knowledge increases contributions by decreasing uncertainty about group members' cooperative behavior.

In Experiment 2, we also sought to replicate and extend our findings from Experiment 1. Just as in Experiment 1, we found an effect of information condition on contributions: participants were more likely to contribute when there was common knowledge compared to private knowledge. However, unlike Experiment 1, we did not see an interaction between threshold level and information condition: there was not a larger effect of common knowledge on contributions within the maximum threshold level compared to the high threshold level. In fact, the effect of common knowledge was actually stronger in the high than the maximum threshold games. Overall, our results suggest that threshold level is not a strong determinant of the effect of common knowledge on cooperation. Interestingly, we observed much higher baseline levels of cooperation, and a relatively smaller effect of common knowledge, in Experiment 2 than Experiment 1, perhaps reflecting post-COVID changes to the Mechanical Turk participant pool (Arechar & Rand, preprint).

One as yet unaddressed question is why we observed intermediate levels of contributions in the common ignorance condition in Experiment 1. That is, even when participants did not know the threshold level, and knew that their group did not either, they still contributed at levels in-between those observed in the common and private knowledge conditions. On its face, this result could be construed as problematic for our account—if uncertainty mediates contribution decisions, why are people contributing in the common ignorance condition at all? However, participants in this condition still possessed common knowledge that there was *a* threshold, and in the absence of threshold level information, participants might have simply assumed that there was a threshold of intermediate size. Furthermore, because there is common knowledge that there is a threshold, we expect that certainty in others' contributions underlies cooperation in this condition, much like it does in the common knowledge condition: Participants who predict that more of their group members will contribute, and are more certain in those predictions, will be more likely to contribute, even if there is uncertainty about the specific threshold level. In Experiment 3, we sought to to replicate the mediating role of certainty on cooperation and more conclusively determine whether the effect of common knowledge varies by threshold size. Additionally, we aimed to explain the contribution levels observed in the common ignorance condition in Experiment 1, and test our prediction that uncertainty about others' cooperative behavior is the mechanism that underlies contributions. We also introduced a new common ignorance condition in which there is uncertainty regarding whether there is a threshold at all (and if there is, what level it is) to explore whether the certainty about the presence of a threshold might explain the contribution levels observed in the common ignorance condition from Experiment 1.

4. Experiment 3

4.1. Method

Participants

We tested N = 1469 participants (58.82% female), aged 19-83 (M = 40.02) on Amazon's Mechanical Turk in a preregistered study (https://osf.io/xqyjp/?view_only=7605805b183e470daafc2d7d0a535c88). This sample was based on previous work in the threshold PGGs conducted on Mechanical Turk (Jordan et al., 2016) and a power analysis in G*Power which suggested we'd have 98%

power to detect a small odds ratio for the effect of common knowledge on contributions. We initially recruited 2,509 participants, 346 were excluded for failing to complete the entire study, 681 for failing any of the comprehension questions, and 13 for responding with “three or more” to a question assessing the number of questions they answered without reading or thinking about them carefully. The exclusion rate is higher than in previous studies, partly due to a programming error with a comprehension check in the survey in one of the common ignorance conditions that impacted about half of participants in this condition (see SOM for details). We include these participants in our data set and note that contributions, predicted contributors, and certainty ratings did not differ between impacted and unimpacted participants.

Design & Procedure

This Experiment included the same information (common knowledge, private knowledge) and threshold level (low, high) conditions as in Experiment 2, but with the addition of two common ignorance conditions. To determine whether certainty ratings also mediate contributions in the common ignorance condition, we added the same common ignorance condition from Study 1, but this time including our certainty measures from Experiment 2. Additionally, we included a new common ignorance condition in which there was uncertainty regarding whether there was a threshold or not, and if there was, its size. Thus, participants were assigned to one of six conditions between-subjects.

The procedure of Experiment 3 was identical to Experiment 2 in every respect with the exception that participants in the common ignorance conditions made a threshold level prediction, in which they were asked to predict the size of the threshold (0-6), after answering their predicted contributor and certainty ratings. As in Experiment 2, participants read the game instructions, answered the comprehension questions, and then made their PGG contribution and PGG predicted contribution and certainty decisions in a counterbalanced order.

Analyses

To replicate our results from Experiment 2, we ran four pre-registered analyses. First, to replicate our finding that people are more likely to contribute when there is common knowledge, we ran a logistic regression with contribution (binary: 0 = did not contribute, 1 = did contribute) as the response term and information condition (common knowledge, common ignorance, private knowledge) as a predictor. Second, to determine whether the effect of common knowledge varies across threshold levels, we ran another logistic regression with contribution as the response term and information condition, threshold size (high, max), and their interaction as predictors. Lastly, to replicate the mediation models showing that predicted contributors and certainty ratings mediated the effect of information condition on contributions, we ran two path analysis mediation models. The mediation models included contribution as the endogenous variable, information condition (private knowledge, common knowledge) as the exogenous variable and predicted contributors (continuous: 0-5) and certainty (continuous: 0-100) as mediators. For both models, and all subsequent mediation models, we used bootstrapping with 5,000 iterations to find standard errors, bias-corrected bootstrapped confidence intervals with 5,000 samples, and diagonally weighted least squares (DWLS) to estimate the model parameters.

We next ran seven pre-registered models unique to Experiment 3. To examine whether people are most likely to contribute when there is common knowledge of the threshold as compared to common ignorance of the threshold size, common ignorance of a threshold, or private knowledge, we ran a logistic regression with contribution as the response term and information condition as the predictor. We also examined whether the predicted number of contributors and certainty in those predictions within the two common ignorance conditions predict contributions with four logistic regression models. Two models predicted contribution by predicted contributors, one with the common ignorance-old data and another with just the common ignorance-new data, while the other two models predicted contribution by certainty ratings, one with the common ignorance-old data, and another with the common ignorance-new data. We ran a series of bootstrap KS tests to compare the distribution of predicted contributors and certainty ratings between the common ignorance conditions and the common knowledge and private knowledge conditions. To determine whether expected threshold level in the common ignorance conditions predicted contributions, predicted contributors, and certainty ratings, we ran six logistic regression models, three with the common ignorance-old data and three with the common ignorance-new data, including contributions, predicted contributors, and certainty ratings as the response terms.

To determine whether participants are more likely to contribute and predicted a higher threshold level when there is common ignorance of the threshold size (common ignorance-old condition) as compared to when there is common ignorance of whether there is a threshold or not (and if so, what the threshold is; common ignorance-new condition) we ran two logistic regression model with common ignorance condition (common ignorance-old, common ignorance-new) as the predictor and either contribution or expected threshold level as the response term. To explore whether predicted contributors and certainty ratings mediated the difference between the common ignorance conditions on contributions, we ran two mediation path analysis models. These models included common ignorance condition (common ignorance-old, common ignorance-new) as the exogenous variable and either predicted contributors (continuous: 0-5) or certainty (continuous: 0-100) as mediators.

Table 6. Proportions and standard deviations (in parentheses) of contributions in the PGG by information and threshold level.

	High	**Maximum**	
Common Knowledge	0.82 (0.39)	0.85 (0.36)	0.83 (.37)
Common Ignorance-New	0.75 (0.44)		
Common Ignorance-Old	0.77 (0.42)		
Private Knowledge	0.77 (0.42)	0.79 (0.41)	0.78 (0.42)
	0.79 (0.40)	0.82 (0.38)	

4.2. Results

Replication of Experiment 2

Replicating the results from Experiment 2, we found that participants were more likely to contribute when there was common knowledge than when there was private knowledge of the threshold (B = -0.36, SE = 0.16, *p* = .02, OR: 0.69, 95% CI: 0.51, 0.95). Participants were significantly more likely to contribute when there was common knowledge than common ignorance (B = -0.43, SE = 0.20, *p* = .03, OR: 0.65, 95% CI: 0.44, 0.97). A planned pairwise comparison revealed that participants were as likely to contribute when there was private knowledge as common ignorance (B = 0.07, SE = 0.19, *p* = .93). When predicting contributions by threshold level, information condition, and their interaction, we found that participants were no longer more likely to contribute when there was common knowledge than private knowledge (B = -0.31, SE = 0.22, *p* = .15, OR: 0.74, 95% CI: 0.48, 1.12). The interaction between information condition and threshold level was also not significant (B = -0.12, SE = 0.32, *p* = .71, OR: 0.89, 95% CI: 0.47, 1.66), replicating the null effect found in Experiment 2. Planned pairwise comparisons revealed that within the high threshold level, participants were not more likely to contribute when there was common knowledge than private knowledge (B = 0.31, SE = 0.22, *p* = .15). Within the maximum threshold level, participants were not significantly more likely to contribute when there was common knowledge compared to private knowledge, although the effect was trending on significance (B = 0.43, SE = 0.24, *p* = .07).

Replicating the previous mediation effects of predicted contributors found in Experiment 2, we found that the predicted number of group contributors fully mediated the effect of information condition on contributions. The total effect of information condition on contributions was significant (*b* = 0.20, SE = 0.09, *p* = .02), while the direct effect of information condition on contributions was not significant (*b* = .04, SE = .07, *p* = .58). The path from information condition to predicted contributors (*b* = 0.30, SE = .08, *p* < .001) was significant, with information condition explaining 11.5% of the variance in the number of predicted group members contributing (see Table S13 in the Supplement for model output). The path from predicted contributors to PGG contribution (*b* = 0.55, SE = .02, *p* < .001) was also significant, with the predicted number of contributors explaining 72% of the variance in contributions in the PGG. Critically, the indirect effect was significant (*b* = 0.17, SE = 0.05, *p* < .001), explaining 8.3% of the total variance, with the bias-corrected bootstrapped confidence interval with 5,000 samples above zero (95% CI: 0, 0.).

We also replicated the effect of participants' certainty on contributions from Experiment 2: our mediation model again found that the number of predicted contributors fully mediated the effect of information condition on contributions. The total effect of information condition on contributions was again significant (b = 0.11, SE = 0.02, *p* <.001) but the direct effect of information condition on contributions was not significant (b = 0.09, SE = 0.09, *p* = .29). The path from information condition to certainty ratings (b = 10.14, SE = 1.62, *p* < .001) was significant, with information condition explaining 19.1% of the variance in certainty ratings (see Table S14 in the Supplement for model output). The path from certainty ratings to PGG contributions (b = 0.01, SE = .002, *p* < .001) was also significant, with certainty explaining 28.4% of the variance in contributions to the public good. Importantly, the indirect effect was significant (b = 0.11, SE = .02, *p* < .001), explaining 5.4% of the total variance, with the bias-correct bootstrapped confidence interval with 5,000 samples above zero (95% CI: 0.07, 0.16).

Additional Analyses

Participants were more likely to contribute when there was common knowledge of the threshold than when there was private knowledge (B = -0.36, SE = 0.16, p = .02, OR: 0.69, 95% CI: 0.51, 0.95), common ignorance of the threshold size (B = -0.43, SE = 0.20, p = .03, OR: 0.65, 95% CI: 0.44, 0.97), or common ignorance of the presence of a threshold (B = -0.54, SE = 0.19, *p* = .004, OR: 0.58, 95% CI: 0.40, 0.85). In the common ignorance-old condition, in which the threshold size was unknown, participants' predicted number of group contributors (B = 1.36, SE = 0.22, *p* <.001, OR: 3.88, 95% CI: 2.63, 6.15) and certainty in those predictions (B = 0.02, SE = 0.006, *p* = .002, OR: 1.02, 95% CI: 1.01, 1.03) significantly predicted contributions. We find the same pattern for the common ignorance-new condition in which the presence of a threshold was unknown: predicted number of group contributors (B = 1.00, SE = 0.15, *p* <.001, OR: 2.73, 95% CI: 2.08, 3.71) and certainty in those predictions (B = 0.02, SE = 0.006, *p* < .001, OR: 1.02, 95% CI: 1.01, 1.03) both significantly predicted contributions.

When comparing the distributions of predicted contributors, we found that the distributions differed significantly between the common knowledge and common ignorance-old conditions (D(734) = 0.23, *p* < .001). A qualitative appraisal of the distributions shows that the common knowledge distribution skewed towards a greater number of expected contributors (see Supplement for histograms of predicted contributors and certainty ratings). The distributions of predicted contributors between private knowledge and common ignorance-old conditions was significant (D(699) = 0.09, *p* = .047): the common ignorance-old distribution was considerably more uniform than the left skewing distribution in the private knowledge condition. The common ignorance-new and common ignorance-old (D(438)= 0.08, *p* = .15) did not differ significantly.

Comparing the distribution of certainty in predicted contributors, we found that the distribution of certainty ratings differed significantly between the common knowledge and common ignorance-old conditions (D(734) = 0.15, *p* = .001). A qualitative appraisal here suggests that the common knowledge distribution was skewed towards higher certainty ratings than the common ignorance-old condition. The certainty distributions did not differ significantly between the private knowledge and common ignorance-old conditions (D(699) = 0.05, *p* = .81) or between the common ignorance-old and common ignorance-new conditions (D(438) = 0.04, *p* = 0.97).

Predicted threshold size within the common ignorance-old did not significantly predict PGG contributions (B = 0.004, SE = 0.14, *p* = .98, OR: 1.00, 95% CI: 0.76, 1.34). Predicted threshold size significantly predicted the expected number of group members contributing (B = 0.21, SE = 0.08, *p* = .006, OR: 1.24, 95% CI: 1.06, 1.44) but it did not predict certainty ratings in those predictions (B = 2.24, SE = 1.66, *p* = .18, OR: 9.43, 95% CI: 0.36, 248.2). When looking at these same models within the common ignorance-new condition, we find that the predicted threshold size predicted contributions (B = 0.22, SE = 0.09, *p* = .01, OR: 1.25, 95% CI: 1.05, 1.50) and the number of predicted group contributors (B = 0.13, SE = 0.05, *p* = .02, OR: 1.13, 95% CI: 1.02, 1.26) but did not significantly predict certainty in those predictions (B = 1.56, SE = 1.05, *p* = .14, OR: 4.74, 95% CI: 0.59, 37.67). Participants were no more likely to contribute to the public good when there was common ignorance of the threshold size as compared to common ignorance of the presence of a threshold (B = -0.11, SE = 0.22, *p* = .64, OR: 0.90, 95%

CI: 0.58, 1.39), nor were they more likely to predict a higher threshold size between common ignorance conditions (B = -0.03, SE = 0.14, *p* = .86, OR: 0.97, 95% CI: 0.74, 1.28).

When exploring whether predicted contributors mediates the effect of common ignorance condition (old vs. new) on contributions, we found that neither the total effect of common ignorance condition on contributions (b = -0.06, SE = 0.13, *p* = .64), nor the direct effect were significant (b = -0.08, SE = 0.11, *p* = .46). The path from common ignorance condition to predicted contributors was not significant (b = 0.04, SE = 0.13, *p* = .75), with common ignorance condition explaining 1.5% of the variance in the number of predicted contributors (see Table S15 in the Supplement for model output). The path from predicted contributors to contributions was significant (b = 0.48, SE = 0.03, *p* < .001), with predicted contributors explaining 64.5% of the variance in contributions. Critically, the indirect effect was not significant (b = 0.02, SE = 0.06, *p* = .75), explaining 1.0% of the variance, with bias-corrected bootstrapped confidence intervals with 5,000 samples spanning zero (95% CI: -0.10, 0.14).

When exploring whether certainty of predicted contributors mediates the effect of common ignorance condition on contributions, we found that the total effect of common ignorance condition on contributions (b = -0.06, SE = 0.13, *p* = .99) and the direct effect (b = -0.06, SE = 0.13, *p* = .63) were not significant. The path from common ignorance condition to certainty ratings was also not significant (b = -0.02, SE = 2.61, *p* = .99), with common ignorance condition explaining 0% of the variance in certainty (see Table S16 in the Supplement for model output). The path from certainty ratings to contributions was significant (b = 0.01, SE = 0.002, *p* < .001), with certainty explaining 30.1% of the variance in contributions. Critically, the indirect effect was not significant (b = 0.00, SE = 0.03, *p* = .99), explaining 0% of the variance, with bias-corrected bootstrapped confidence intervals with 5,000 samples spanning zero (95% CI: -0.06, 0.06).

Discussion

In line with our predictions, we found that participants' predictions of the number of group members contributing and their certainty in those predictions predicted contributions when there was common ignorance of the threshold. In other words, when participants had common knowledge that there was a threshold of unknown size, certainty about other group members' cooperative behavior supported the effect of common knowledge on contributions. Furthermore, we found that, in the absence of a specific threshold level, participants inferred that the threshold was at an intermediate level. However, in contrast to Experiment 1, we found that participants were significantly more likely to contribute when there was common knowledge than private knowledge or common ignorance, a finding that suggests a unique effect of common knowledge of the threshold size. In line with Experiment 1, participants were as likely to contribute when there was common ignorance as private knowledge. Thus, the fact that contributions, and the number of expected contributors, in the common ignorance condition were not significantly different from private knowledge is likely due to participants inferring an intermediate threshold level and were guided by their certainty that their group members knew there was a threshold. While we initially predicted that participants would be more likely to contribute when there was common ignorance of the threshold size compared to common ignorance of whether there was a threshold at all, we failed to find differences in cooperation between these two conditions. However, we believe this is due to the fact that very few participants in the common ignorance-new condition thought there was no threshold: participants made similar predictions about the threshold size between conditions, suggesting they treated these conditions very similarly. Indeed, participants made nearly identical predictions about the number of contributors and certainty ratings between the common ignorance conditions, results which likely explain why we failed to find to evidence of mediation from the common ignorance conditions on contributions.

We also replicated the effect of common knowledge on contributions found in Experiments 1 and 2 as well as the mediation models from Study 2 that found that the number of predicted group contributors, and certainty in those predictions, mediated the effect of common knowledge on cooperation. This provides stronger support that uncertainty about other agents' cooperative behavior underlies the prosocial effect of common knowledge on cooperation. Consistent with the results from Experiment 2, we again failed to replicate the interaction between information condition and threshold level observed in Experiment 1, suggesting that initial interaction effect might have been spurious.

5. General Discussion

The goal of this project was to investigate how common knowledge promotes cooperation, testing the hypothesis that common knowledge increases cooperation by reducing uncertainty about others' cooperative behavior. Introducing thresholds to the PGG transforms the game from a pure social dilemma to an anti-coordination or coordination problem, and because common knowledge increases coordination (Thomas et al., 2014; Thomas et al., 2018), we predicted that common knowledge would increase contributions by decreasing the uncertainty surrounding whether other group members will contribute. In three studies, we manipulated the information participants had regarding what their group members knew about the threshold, as well as the level of threshold needed to receive the public good. We found that common knowledge of the threshold increased cooperation in the PGG and that the effect of common knowledge was mediated by the predicted number of group members contributing and certainty about the predicted number of contributors.

Overall, our finding that common knowledge increased contributions supports recent work suggesting that common knowledge is an important mechanism for coordinating behavior (Thomas et al., 2014; Thomas et al., 2018; De Freitas et al., 2019). Our work builds upon this literature by showing that common knowledge not only increases cooperation in two-player coordination games, but that it can also increase cooperation in *n*-person coordination games that more closely model the kinds of cooperation problems we encounter in everyday life. More generally, this finding provides some support for the special role of common knowledge in human cooperation, and its function as a potentially distinct cognitive mechanism that may have evolved to help us solve coordination problems and for social strategizing (De Freitas et al., 2019; Thomas et al., 2016). Future work should continue to investigate the role common knowledge plays in coordination problems, as well as what role, if any, common knowledge plays in social dilemmas, where there is a conflict between an actor's self-interest and the interest of the group. If common knowledge evolved as an adaptation for social strategizing (Thomas et al., 2014), then it might actually reduce cooperation in social dilemmas where actors' interests are diametrically opposed to one another. Indeed, Thomas et al., (2016) found that common knowledge reduces prosocial helping in a bystander intervention task that models the volunteer's dilemma, a type of anti-coordination game. However, in our intermediate threshold PGGs, which model anti-coordination problems, we find a prosocial effect of common knowledge This tension suggests that more work is needed to better understand whether common knowledge will reduce prosocial behavior in a prisoner's dilemma or other social dilemma where cooperation is personally costly.

It is important to note that the focus of our knowledge conditions differed from the Thomas et al. (2014) study which examined common knowledge of the joint-payoff for coordinating. In our study, participants had common knowledge of the threshold, rather than the payoff. However, we believe our findings would extend to other important features of the PGG, such that common knowledge of the contribution multiplier, for example, would also increase contributions relative to private knowledge. Additionally, while we only compared common knowledge to private knowledge here, it would be interesting to examine shared or asymmetric knowledge (such as secondary or tertiary knowledge) of a threshold to see how it influences cooperation relative to common knowledge. Future work should compare common knowledge to shared knowledge of a threshold and explore whether common knowledge of other aspects of the game increases contributions.

One of our most important findings was that certainty about others' cooperative behavior mediated the effect of common knowledge about the threshold on contributions. In other words, when deciding whether to contribute, participants incorporated information about others' mental states in order to infer the likelihood that they would contribute and the threshold would be met. That is, common knowledge increased cooperation because it increased certainty that other group members would contribute and the threshold would be met. To the best of our knowledge, this finding provides the first evidence that common knowledge promotes coordination primarily by reducing uncertainty about the behavior of other social agents. This is consistent with past work showing that uncertainty about the threshold or game structure is generally detrimental to the provision of the public good and coordination (Dannenberg et al., 2011; McBride, 2010; Rubinstein, 1989). More generally, this finding highlights the important role our beliefs about other agents' beliefs and behavior play in our own cooperative behavior.

One question that arose after Experiment 1 is why contributions were not significantly lower when there was common ignorance of the threshold as compared to common knowledge. In Experiment 3 we sought to address this question: we found that participants were significantly more likely to contribute when there was common knowledge than common ignorance of the threshold. Additionally, we found that—as with common knowledge—contributions in the common ignorance conditions were predicted by predicted contributors and certainty in those predictions. That contributions were similar when there was common ignorance or private knowledge of the threshold likely reflects the fact that participants in the common ignorance condition possessed common knowledge that there was *a* threshold, just one of unknown size. So, although they lacked the specific threshold information, participants could still be reasonably certain that their group members knew there was a threshold and that they would contribute, which in turn motivated their own contribution decisions. Overall, our finding here that contributions were higher when there was common knowledge than common ignorance supports past work finding that uncertainty about the threshold size is generally detrimental to the provisioning of the public good (Barrett and Dannenberg, 2014; Dannenberg et al., 2015).

Our results also support past work that has found that thresholds promote cooperation in the PGG (Jordan et al., 2017; Szolnoki & Perc, 2010; van de Kragt et al., 1983). Indeed, in both of our studies, participants contributed significantly more when there was a threshold, regardless of the threshold size or beliefs about others' knowledge of it. Our finding from Experiment 1, that participants contributed more to the public good even when the threshold was unknown to themselves and their group, further underscores the effectiveness of thresholds: merely the presence of a threshold was enough to increase contributions. Interestingly, while we did not find an effect of threshold size on contributions across all three studies, we did find that threshold size predicted the expected number of group members contributing: the higher the threshold, the more group members were predicted to contribute. It is surprising that predicted contributors and contribution decisions dissociated when looking at the effect of threshold size, as we otherwise found a strong predictive relationship between predicted contributors and contribution decisions. Future work should examine this finding in more depth. Overall, these findings demonstrate the important role thresholds have in bolstering cooperation, even when there is uncertainty about the size of, or information other individuals know about, the threshold.

There were a few limitations in our studies and opportunities for future work. One concern with our approach is that contribution decisions were binary, all-or-nothing decisions rather than continuous contributions (e.g., any amount from $0 to $0.30). While our approach is not novel, previous work has used binary contribution decisions in threshold PGGs (Rapoport & Eshed-Levy, 1989; van de Kragt et al., 1983), it is possible that people might behave differently when contributions are continuous. Data from our group supports this idea; in a previous study we found that threshold level influences contributions when they are continuous (Deutchman et al., in prep). We note that across both studies a number of participants were excluded for failing comprehension checks. We believe this high exclusion rate reflects the relatively complex nature of the threshold PGG and our inclusion of extensive comprehension checks—three on the dynamics of the threshold PGG and four on the information conditions (see the Supplement for the specific comprehension checks). We wish to note, however, that rates of exclusions were roughly similar across conditions (SOM – S11-13), suggesting that while the task itself was relatively complex, these exclusions are unlikely to have contributed to our reported effects.

Lastly, we would like to note that the effects of common knowledge on contributions found across our studies constituted a relatively small effect. We believe this could be due to several non-mutually exclusive reasons. First, our common knowledge manipulation was relatively subtle compared to how common knowledge is likely established in daily life as it was constrained by collecting data online; given a stronger, more naturalistic common knowledge manipulation (such as establishing common knowledge through eye contact) we would expect a significantly larger effect of common knowledge on cooperation. Second, we observed overall very high levels of cooperation (74% across all studies) that may have constituted something of a ceiling effect, reducing movement between conditions. Third and finally, it is possible that the small stake size used here contributed to the overall high levels of cooperation and, by reducing the risk associated with contributing, lessened the effect of common knowledge. However, previous work has found that there is not a meaningful difference between large and small stake sizes (Amir et al., 2012). Future work should examine the boundary conditions of the effect of common knowledge on cooperation by manipulating common knowledge in a more naturalistic way, such as through eye contact and increasing the riskiness of contributions by varying the stake size and other factors.

In sum, we investigated whether common knowledge increases contributions in the PGG by reducing uncertainty about others' cooperative behavior. Across two studies, we found that common knowledge increased contributions in the PGG, and that this effect was mediated by the predicted number of group contributors and certainty about those predictions. These findings provide strong evidence that the effect of common knowledge on coordination is mediated by certainty about others cooperative behavior. Lastly, our studies provide the important insight that common knowledge can increase cooperation in *n*-person coordination problems. More generally, our results are consistent with theories that common knowledge is an evolved, cognitive mechanism for solving coordination problems. At the broadest level, our findings reveal the potential that common knowledge holds for promoting cooperation in the large-scale coordination problems that predominate our social lives.

5. Open Practices

This project was pre-registered prior to data collection [Study 1: http://aspredicted.org/blind.php?x=63ct9y; Study 2: https://osf.io/brqky/?view_only=ab7c7982f7454e439ca63c5806c00d52; Study 3: https://osf.io/xqyjp/?view_only=7605805b183e470daafc2d7d0a535c88]. All of our data and code are available in an online repository here: [https://osf.io/wkcrf/?view_only=6be9f35572bd44dc8927d75fd63292b0].

6. References

Almaatouq, A., Krafft, P., Dunham, Y., Rand, D. G., & Pentland, A. (2020). Turkers of the world unite: Multilevel in-group bias among crowdworkers on amazon mechanical turk. *Social Psychological and Personality Science, 11*(2), 151-159.

Amir, O., & Rand, D. G. (2012). Economic games on the internet: The effect of $1 stakes. *PloS One*, *7*(2), e31461.

Andrews, T. M., Delton, A. W., & Kline, R. (2018). High-risk high-reward investments to mitigate climate change. *Nature Climate Change*, *8*(10), 890.

Archetti, M., & Scheuring, I. (2012). Game theory of public goods in one-shot social dilemmas without assortment. *Journal of Theoretical Biology, 299*, 9-20.

Arechar, A. A., & Rand, D. (Preprint). Turking in the time of COVID.

Axelrod, R., & Hamilton, W. D. (1981). The evolution of cooperation. *Science*, *211*(4489), 1390-1396.

Balakrishnan, J. D., & Ratcliff, R. (1996). Testing models of decision making using confidence ratings in classification. Jo*urnal of Experimental Psychology: Human Perception and Performance, 22*(3), 615.

Baltag, A., Moss, L. S., & Solecki, S. (2016). The logic of public announcements, common knowledge, and private suspicions. In *Readings in Formal Epistemology* (pp. 773-812). Springer, Cham.

Bradley, J. V. (1981). Overconfidence in ignorant experts. *Bulletin of the Psychonomic Society, 17*(2), 82-84.

Clark, H. H., & Marshall, C. R. (1981). Definite knowledge and mutual knowledge.

Clutton-Brock, T. (2009). Cooperation between non-kin in animal societies. *Nature*, *462*(7269), 51.

Dannenberg, A., Löschel, A., Paolacci, G., Reif, C., & Tavoni, A. (2015). On the provision of public goods with probabilistic and ambiguous thresholds. *Environmental and Resource Economics*, *61*(3), 365-383.

De Freitas, J., Thomas, K., DeScioli, P., & Pinker, S. (2019). Common knowledge, coordination, and strategic mentalizing in human social life. *Proceedings of the National Academy of Sciences*, *116*(28), 13751-13758.

Deutchman, P., Amir, D., Jordan, M., McAuliffe, K. (In prep). The Effect of Social and Non-Social Thresholds in the Public Goods Game.

Diekmann, A. (1985). Volunteer's dilemma. *Journal of Conflict Resolution, 29*(4), 605-610.

Fehr, E., & Fischbacher, U. (2004). Social norms and human cooperation. *Trends in Cognitive Sciences*, *8*(4), 185-190.

Fehr, E., & Gachter, S. (2000). Cooperation and punishment in public goods experiments. *American Economic Review*, *90*(4), 980-994.

Fischbacher, U., Gächter, S., & Fehr, E. (2001). Are people conditionally cooperative? Evidence from a public goods experiment. *Economics Letters*, *71*(3), 397-404.

Halpern, J. Y., & Moses, Y. (1990). Knowledge and common knowledge in a distributed environment. *Journal of the ACM (JACM)*, *37*(3), 549-587.

Hauert, C., & Doebeli, M. (2004). Spatial structure often inhibits the evolution of cooperation in the snowdrift game. *Nature*, *428*(6983), 643.

Heizer, R. F. (1953). Aboriginal fish poisons. *Bureau of American Ethnology Bulletin*.

Horton, J. J., Rand, D. G., & Zeckhauser, R. J. (2011). The online laboratory: Conducting experiments in a real labor market. *Experimental economics*, *14*(3), 399-425.

Jansson, F., & Eriksson, K. (2015). Cooperation and shared beliefs about trust in the assurance game. *PloS One, 10*(12), e0144191.

Jordan, J., McAuliffe, K., & Rand, D. (2016). The effects of endowment size and strategy methodon third party punishment. *Experimental Economics*, *19*(4), 741-763.

Jordan, M. R., Jordan, J. J., & Rand, D. G. (2017). No unique effect of intergroup competition on cooperation: non-competitive thresholds are as effective as competitions between groups for increasing human cooperative behavior. *Evolution and Human Behavior*, *38*(1), 102-108.

Marks, M. B., & Croson, R. T. (1999). The effect of incomplete information in a threshold public goods experiment. *Public Choice*, *99*(1-2), 103-118.

McBride, M. (2010). Threshold uncertainty in discrete public good games: An experimental study. *Economics of Governance*, *11*(1), 77–99.

Rand, D. G., & Nowak, M. A. (2012). Evolutionary dynamics in finite populations can explainthe full range of cooperative behaviors observed in the centipede game. *Journal of Theoretical Biology*, *300*, 212-221.

Rapoport, A., & Eshed-Levy, D. (1989). Provision of step-level public goods: Effects of greed and fear of being gypped. *Organizational Behavior and Human Decision Processes*, *44*(3), 325-344.

Rubinstein, A. (1989). The electronic mail game: Strategic behavior under "almost common knowledge." The American Economic Review, 79, 385–391

Sekhon, Jasjeet S. 2011. Multivariate and Propensity Score Matching Software with Automated Balance Optimization: The Matching package for R. *Journal of Statistical Software. 42*(7): 1-52.

Skyrms, B. (2004). The stag hunt and the evolution of social structure. New York, NY: Cambridge University Press.

Szolnoki, A., & Perc, M. (2010). Impact of critical mass on the evolution of cooperation in spatial PGGs. *Physical Review E*, *81*(5), 057101.

Thomas, K. A., DeScioli, P., Haque, O. S., & Pinker, S. (2014). The psychology of coordination and common knowledge. *Journal of Personality and Social Psychology, 107*, 657-676

Thomas, K. A., De Freitas, J., DeScioli, P., & Pinker, S. (2016). Recursive mentalizing and common knowledge in the bystander effect. *Journal of Experimental Psychology: General, 145*, 621– 629.

Thomas, K. A., DeScioli, P., & Pinker, S. (2018). Common knowledge, coordination, and the logic of self-conscious emotions. *Evolution and Human Behavior*, *39*, 179 190.

Thunström, L., Nordström, J., & Shogren, J. F. (2015). Certainty and overconfidence in future preferences for food. *Journal of Economic Psychology, 51*, 101-113.

West, S. A., Griffin, A. S., & Gardner, A. (2007). Social semantics: altruism, cooperation, mutualism, strong reciprocity and group selection. *Journal of Evolutionary Biology*, *20*(2), 415-432.

Wit, A., & Wilke, H. (1998). Public good provision under environmental and social uncertainty. *European Journal of Social Psychology*, *28*(2), 249–256.

Van de Kragt, A. J., Orbell, J. M., & Dawes, R. M. (1983). The minimal contributing set as a solution to public goods problems. *American Political Science Review*, *77*(1), 112-122.

2. Study 2: Children use common knowledge to solve coordination problems

Recent work suggests that common knowledge is an important cognitive mechanism for coordinating prosocial behavior, in part because it reduces uncertainty about others' cooperative behavior. However, it remains unclear whether children also rely on common knowledge to solve coordination problems. Here we examined whether 6-9-year-old children (N = 133) from the US were more likely to attempt to coordinate when they had common knowledge about a joint payoff. Participants saw three vignettes that modeled the structure of a two-player coordination problem and were provided with common knowledge, secondary knowledge, or private knowledge about the mutually beneficial, but risky, joint payoff. By 6-years of age, participants were more likely to attempt to coordinate when they had common knowledge than secondary knowledge, and secondary knowledge than private knowledge. Participants were also most likely to expect the other player to coordinate, and were most certain in their predictions, when there was common knowledge. Results indicate that, by middle childhood, children are able to solve coordination problems by relying on common knowledge, in part because it likely increases their certainty in others' cooperative behavior. Overall, findings suggest that common knowledge is an important cognitive mechanism for coordinating behavior and that it does so by reducing uncertainty about others' cooperative behavior.

This paper is co-authored with Katherine McAuliffe.

1. Introduction

Cooperation is a critical aspect of human social life, yet it presents a number of challenges. We must coordinate behavior with other people, even when doing so goes against self-interest. Furthermore, even when cooperation is mutually beneficial, it is often a risky endeavor: if you cooperate and others do not, you risk wasting your contribution or being exploited by others. Given these challenges, how has cooperation been so successful in our species? What cognitive abilities allow us to overcome these possible barriers and the inherent risk of cooperative interactions?

Recent work suggests that one important cognitive mechanism that promotes cooperation is common knowledge. Common knowledge is the recursive belief state in which A knows X, B knows X, A knows that B knows X, B knows that A knows that B knows X, ad infinitum (Thomas et al., 2014; Rubinstein, 1989). Importantly, because infinite recursion is cognitively impossible to represent, researchers believe that common knowledge is a unique cognitive state that serves as a heuristic for recursive beliefs which is activated by any public signal that reliably establishes that everyone knows that everyone knows X (Thomas et al., 2014; De Freitas et al., 2019). Research with adults suggests that common knowledge is an important mechanism for coordinating group behavior (Thomas et al., 2014; Thomas et al., 2018). Specifically, people are more willing to attempt risky coordination[4] when there is common knowledge about the benefits of coordinating compared to when there is only shared knowledge (such as secondary knowledge states—A knows that B knows X but nothing else—and tertiary

[4] Note that cooperation and coordination are conceptually distinct—coordination is risky but mutually beneficial whereas cooperation in a social dilemma is personally costly (Snidal, 1985). Here we occasionally refer to coordination as cooperation since it broadly falls under the umbrella of cooperative behavior (Ashley & Tomasello, 1988).

knowledge states—A knows that B knows that A knows X but nothing else; Thomas et al., 2014).

Building on these studies, recent work has found that common knowledge promotes cooperation in the threshold public goods game—an economic game that captures the dynamics of cooperation problems experienced in everyday life—by reducing uncertainty around group members' cooperative behavior (Deutchman et al., 2022). Specifically, under common knowledge, people were more willing to cooperate than under private knowledge (where they lacked information about their group members' beliefs), they expected that other members of their group would cooperate, and they were more certain in these expectations. Together, these results suggest that common knowledge plays an important role in human cooperation by reducing uncertainty in others' cooperative behavior (henceforth **social uncertainty**).

While work with adults suggests that common knowledge plays an important role in cooperative interactions by reducing social uncertainty, thus reducing the riskiness of cooperating, it remains unclear whether children similarly rely on common knowledge to help solve risky coordination problems. This is an important, under-explored question because children frequently encounter coordination-like problems throughout development, such as how to take turns in games during free play (e.g., follow-the-leader, playing catch, etc.) or work together to solve a common goal (Ashley & Tomasello, 1998; Eckerman et al., 1989). Addressing this question will inform our understanding of how cooperative behavior develops in childhood and elucidate our understanding of the cognitive mechanisms critical for successful cooperation. Namely, if children can use common knowledge to successfully coordinate, this would suggest that common knowledge is one of the core social-cognitive abilities supporting cooperation.

Existing developmental work on common knowledge (sometimes called *common ground*; Bohn & Köymen, 2018) suggests that the foundations of this ability might be present early in development, becoming increasing sophisticated through the preschool years. For example, infants as young as 14-months-old can use joint attention to guide their behavior (Liebal et al., 2009; Moll et al., 2007), an important ability for establishing common knowledge. Three-to-five-year-olds can reason with a peer based on common ground assumptions (Köymen et al., 2014) and adapt their justifications depending on what information is common knowledge (Köymen et al., 2016), while 5- but not 3-year-olds understand that common ground establishes an implicit joint-commitment (Kachel et al., 2019).

Communicative eye contact, which can establish common knowledge, creates expectations of collaboration in 5-year-olds (Siposova et al., 2018) and 6-year-olds are more likely to behave prosocially in a helping task when they shared common knowledge about the experimenter's need compared to when they only had private knowledge (Siposova et al., 2021). Other work suggests that while children are able to use common knowledge to make basic social inferences, they do not make more complex inferences about another individual's group membership until 8 years of age (Soley & Koseler, 2021). Together, this work suggests that the ability to understand common knowledge is likely present early in development, while the ability to use this knowledge to guide appropriate inferences and behavior emerges later during middle-childhood (Bohn & Köymen, 2018). Relatedly, there are also developmental differences in the ability to recognize common knowledge—young children require direct social interaction, such as eye contact and communication, to establish common knowledge—while children might not be sensitive to more indirect cues of common knowledge, such as focal points or shared cultural knowledge, until later in development (Bohn & Köymen, 2018; Goldvicht-Bacon & Diesendruck, 2016). Because we frequently rely on these kinds of cues to establish common knowledge in our everyday lives, it is important to study when in development children are able to recognize them. Thus, while past work suggests a rudimentary form of common knowledge might be present early in childhood, it remains unclear when children develop the more sophisticated ability to establish common knowledge from indirect cues and then use that knowledge to guide their behavior.

Around the same time that children begin to understand and use common knowledge, they also begin to solve coordination problems. For example, children can successfully coordinate in simple coordination problems by 5 years of age (Grueneisen et al., 2015; Grueneisen et al., 2015b), and begin to solve more complicated coordination problems by 7 or 8 years of age (Grueneisen & Tomasello, 2019; Grueneisen & Tomasello, 2020). That children's ability to use common knowledge emerges around the same time in development as their ability to solve coordination problem raises the exciting question of whether these maturational processes are related. Namely, while common knowledge representation emerges by early-middle childhood, it remains unclear whether children use this knowledge to solve coordination problems or whether children's newly acquired ability to solve coordination problems is supported by other emerging developmental abilities such as higher-order belief representation or executive functioning. Therefore, it is important to explicitly examine the role of common knowledge in coordination to determine whether it is the primary ability allowing children to coordinate successfully.

In the present preregistered study, we explored whether children use common knowledge to solve coordination problems. Moreover, we explored whether common knowledge promotes cooperation by increasing certainty that a mutually beneficial outcome will be achieved. To investigate this, we presented 6- to 9-year-old children with three vignettes that modeled the dynamics and payoff structure of a coordination problem. Because cooperation is mutually beneficial in coordination problems—the best outcome is when both agents cooperate—we expect to see reasonably high levels of overall cooperation in our task. However, because cooperation in coordination problems is risky—the worst outcome is to cooperate if the other agent defects—we expect cooperation to be lower when agents have incomplete or asymmetric information. Cooperation is riskier in such instances because it is harder to predict whether the other agent will cooperate, thus increasing the chances of coordination being unsuccessful. Common knowledge can help solve this problem by providing both agents with greater confidence that the other possess complete information about the cooperative interaction, thereby increasing certainty in other agents' cooperative behavior and decreasing the riskiness of cooperation.

In our coordination scenarios, we manipulated the knowledge states of the agents in the scenarios and measured whether participants decided to cooperate or defect in these hypothetical coordination problems, what they predicted the other agent would do, and how certain they were in their prediction. We hypothesized that children would be more likely to cooperate when there was common knowledge about the mutual benefits of coordination than when there is shared or private knowledge. Additionally, we hypothesized that participants would be more likely to predict that the other agent will cooperate, and be more certain in their predictions, when there was common knowledge than shared or private knowledge. We selected our age range of 6-9 because it represents a development period in which children begin to more consistently solve coordination problems (Grueneisen et al., 2015; Grueneisen & Tomasello, 2019; Grueneisen & Tomasello, 2020). Yet, only toward the end of this age range—around 8—are children able to apply common knowledge to make social inferences (Soley & Koseler, 2021). Given this previous work, we predicted that, while children might understand common knowledge by 6 years of age, they will not be able to consistently apply it in coordination problems until later in development, such that the effect of common knowledge on coordination will increase with age, aligning with adults' performance by 9 years of age.

In addition to common knowledge, other mentalizing[5] abilities may play a role in promoting cooperation. For instance, 5-year-old children modify their behavior in coordination problems depending on their partner's presumed knowledge, suggesting that their cooperative behavior is sensitive to others' knowledge states (Goldvicht-Bacon & Diesendruck, 2016). Furthermore, 6-year-old children can use second-order false beliefs in order to coordinate (Grueneisen et al., 2015), suggesting that representing the beliefs of other agents—second-order belief representation—might be enough to allow children to solve coordination problems. In order to explore whether common knowledge offers a unique benefit for cooperation over second-order belief representation, we included a *secondary knowledge* condition in which participants knew the other agents' beliefs (but

[5] Here we refer to mentalizing very broadly as the ability to make inferences about others' mental states (Frith & Frith, 2006)

where the other agent did not know the participants' knowledge). If participants are more likely to cooperate when they have common knowledge as compared to secondary knowledge, that would suggest that common knowledge is an especially important mechanism for cooperation above that of second-order beliefs.

2. Method

Participants

Participants were N = 133 6-9-year-old children recruited into two age groups: 6-7-year-olds (N = 63, M = 7.04, SD = 0.64, range = 6.03-7.95, 46.03% females) and 8-9-year-olds (N = 70, M = 9.03, SD = 0.59, range = 8.02-9.85, 47.14% females) and were majority Caucasian (Caucasian = 56.4%, Asian = 14.3%, Hispanic = 6.8%, Black = 3%, Biracial = 15.0%, Other = 4.5%). Participants were recruited via a lab database and Facebook ads across the United States. The sample size was determined by a power analyses in G*Power which found we would have 80% power to detect a medium effect of information condition. All participants were from the United States and were tested online via Zoom video conferencing technology. While we initially recruited N = 151, 18 children were excluded from data analysis for meeting our preregistered exclusion criteria, including severe inattention (1), atypical development as reported by guardian (5), parental interference (1), failing the comprehension checks (8), or failing to complete the study in its entirety (1). Two additional children were excluded for living outside the US (1) or outside our age range (1). This study was approved by the IRB (#16.242.04-33) and preregistered prior to data collection: https://aspredicted.org/H4W_3RS.

Materials

Participants saw animated gifs of scenes depicting a coordination problem. Animations were created using Vyond animation software. Animations and questions were embedded in a Qualtrics survey that an experimenter controlled and was shown to participants via Zoom's screen sharing function (for a discussion of the validity of remote data collection, see Sheskin et al., 2020).

Design

Children were presented with three different animated scenarios that modeled coordination problems. Children received the information conditions within-subjects in a randomized order. For each scenario, children received one of the information conditions (common knowledge, secondary knowledge, private knowledge). The pairings of coordination vignettes and information conditions were counterbalanced, such that each vignette-information condition combination was equally likely to be presented.

Procedure

Children were first introduced to the experimenter and provided their consent to participate. They were then presented with an avatar selection choice in which they selected one of two avatars (one male, one female) to represent them in the stories. We included the avatar selection in order to promote engagement with the scenarios. Whichever avatar the child selected was used in the stimuli for the rest of the task.

Participants were then presented with three different first-person scenarios that were designed to model the dynamics of a coordination problem (see Figure 1 for a diagram of the task). In each vignette, participants were presented with an animated story in which they were asked to imagine that they were friends with another character and had to decide whether to attempt risky coordination by meeting the character at the park (cooperating) or play it safe by staying home (defecting). Coordination was risky in the sense that while cooperation is mutually beneficial, failed coordination (where an agent cooperates but their partner defects) results in a worse outcome than defecting. The payoff structure was the same as a simple coordination game: the best outcome is for both players to choose to play together and meet at the park (coordination) and the worst outcome is to try to play together by going to the park while the other player stays home (failed coordination). In between those payoffs, the middle-value outcome is to play it safe by staying at home (see Figure 2 for the payoff matrix). While unincentivized, the payoff units were framed in terms of utility for the expected outcome (i.e., fun). After learning about the payoff outcomes, participants answered three comprehension questions assessing their understanding of the coordination payoffs for each vignette (see supplement for the comprehension questions). If participants answered any question incorrectly, they were reminded of the relevant text from the vignette that provided the answer and were asked again. Participants were given three attempts to answer each question. If they failed on the third attempt, the experimenter provided the correct answer and moved on, and the participant was excluded from the analyses (see preregistration for exclusion criteria).

To manipulate knowledge states, participants were told that the location of the coordination activity (the park) was sometimes closed, thus creating uncertainty about the mutually beneficial joint-payoff. We then manipulated whether the agents knew that the park was open and thus that the mutually advantageous payoff was available (Figure 1). There were three information conditions within-subject: common knowledge, secondary knowledge, and private knowledge. For all the information conditions, we manipulate knowledge state using a messenger—a third animated character—a method which has been used in previous work on common knowledge and coordination (Thomas et al., 2014). In the common knowledge condition, the messenger made an announcement on the loudspeaker that everyone could hear (see supplement for script and stimuli). This established common knowledge that the park was open because everyone heard the announcement, and everyone knew that everyone else heard the announcement. In the secondary knowledge condition, the messenger told participants that the park is open and that they stopped at the other character's house to relay the message. Thus, participants knew that their friend also knew the park was open, but that the friend did not know whether the participant knew the park was open. In the private knowledge condition, the messenger told participants that the park is open today but not whether they also stopped at the other character's house. Thus, participants did not know whether their friend also knew the park was open, or whether their friend knew that they knew the park was open. After each information condition, participants answered three comprehension questions to ensure they properly understood the knowledge states of the corresponding information condition. As with the earlier comprehension checks, participants were given three attempts to answer correctly before moving on with the survey. Participants who failed after their third attempt were excluded from data analysis (5.3% of participants).

After receiving the information condition, participants answered the dependent variables. First, they were asked whether they wanted to try to play together with the other character by going to the park (a cooperation decision), or to play alone by staying home (a defection decision). This allowed us to compare the likelihood of attempting to coordinate across the different information conditions. Second, participants predicted whether their friend would attempt to play together by going to the park (cooperate) or stay home and play alone (defect). To measure their prediction certainty, children then rated how sure they were about their prediction (continuous: 1-5). This allowed us to explore whether common knowledge helps to solve coordination problems by reducing uncertainty surrounding others' cooperative behavior. Lastly, because previous work has found that common knowledge creates expectations of cooperation (Siposova et al., 2018), we measured violation of cooperative expectations by asking participants how surprised they would be if the other character did not try to coordinate (continuous: 1-5). Participants answered these measures three times, once for each information condition. All of our code, materials, and data are publicly available online at the Open Science Framework (https://osf.io/pq42r/?view_only=9c46860b1385463ebe6aca1ba2607cb2).

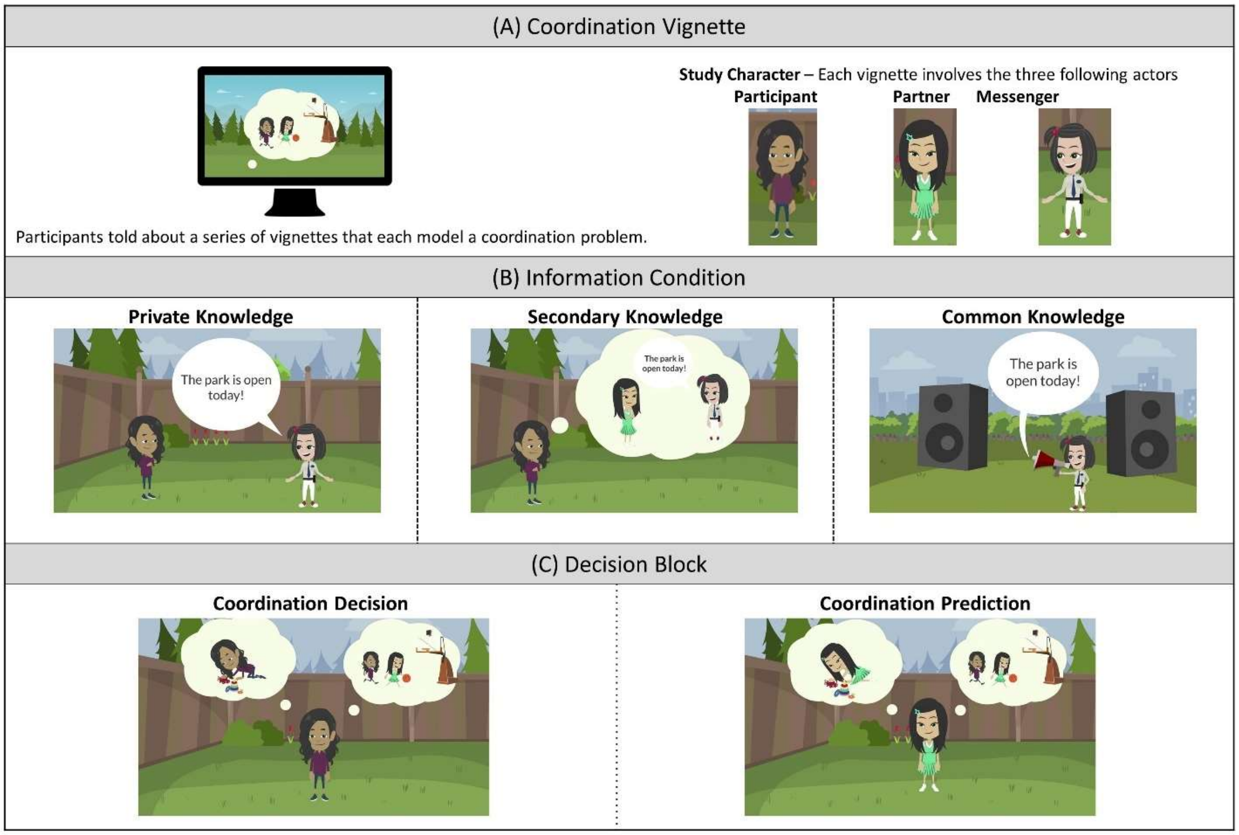


Figure 1. Diagram of the coordination task with images from the stimuli. © 2022 GoAnimate, Inc.[6]

Analyses

We randomly selected 20% of participant videos to be coded by an independent coder who found there was 100% agreement (Cohen's Kappa = 1) on all four dependent variables.

[6] Images are copyrighted by and used by permission of VYOND™. VYOND is a trademark of GoAnimate, Inc., registered in Australia, Brazil, Canada, Chile, Egypt, the European Union, Hong Kong, India, Indonesia, Israel, Japan, Malaysia, Mexico, New Zealand, Norway, OAPI, the Philippines, Russia, Singapore, Switzerland, the United Kingdom and Vietnam."

All analyses were conducted in R version 4.1.2 (R Core Team, 2020). We ran five preregistered generalized linear regression models predicting each of our main dependent variables by information condition (baseline = common knowledge; Models 1-4) or the interaction between information condition and age (Model 5), and including participant ID as a random effect (see supplement for details). To compare the secondary knowledge and private knowledge conditions, we made pairwise comparisons using estimated marginal means adjusted using the multivariate t method (MVT) to correct for multiple comparisons. Lastly, while we had planned to conduct a structural equation mediation model examining whether certainty ratings about the other agents' cooperation mediate the effect information condition (private knowledge, common knowledge) on cooperation, our preregistered model did not account for the repeated nature of our data (information condition was within-subjects) and multilevel mediation was unsuitable, in part because there was too little variance in certainty ratings in the common knowledge condition.[7] Consequently, we do not report the results of the preregistered mediation model here but note that we include it in our supplement in the interest of transparency. We report the results of all other preregistered models.

3. Results

Figure 2 depicts the proportion of participants deciding to cooperate across the three information conditions (see supplement for plots with age breakdown). We first compared a baseline model including the random effect of participant and age to a model

[7] Multilevel mediation was unsuitable because there was no residual at the level 1 data when including the within-subject information condition.

that additionally included the information condition term. The model with the information condition term offered a significantly better fit to the data than the model without it (LRT, $\chi 2(2) = 45.4$, p < .001). We next compared the model predicting cooperation by information condition to the model predicting cooperation by the interaction between information condition and age (continuous). The model including the interaction term did not provide a significantly better fit to the data than the model without it (LRT, $\chi 2(3) = 3.22$, p = .36). Participants were significantly more likely to cooperate when there was common knowledge than secondary (B = -1.13, SE = 0.49, *p* = .02, OR: 0.32, 95% CI: 0.12, 0.80) or private knowledge (B = -2.59, SE = 0.49, *p* < .001, OR: 0.07, 95% CI: 0.03, 0.18), and when there was secondary as compared to private knowledge (B = -1.46, SE = 0.37, *p* = .002). Participants also predicted that the other agent would be more likely to cooperate when there was common knowledge compared to secondary (B = -1.92, SE = 0.61, *p* = .002, OR: 0.15, 95% CI: 0.05, 0.48) or private knowledge (B = -3.52, SE = 0.68, *p* < .001, OR: 0.03, 95% CI: 0.008, 0.11), and secondary compared to private knowledge (B = -1.61, SE = 0.41, *p* = .003).

Participants were more certain that their partner would cooperate when they had common knowledge than secondary (B = -0.63, SE = 0.11, *p* < .001, 95% CI: -0.84, -0.42) or private knowledge (B = -0.99, SE = .12, *p* < .001, 95% CI: -1.23, -0.77), and when they had secondary compared to private knowledge (B = -0.37, SE = .12, *p* = .007). Likewise, participants reported they would be more surprised that their partner failed to cooperate when they shared common knowledge compared to secondary (B = -0.73, SE = 0.15, *p* < .001, 95% CI: -1.02, -0.43) or private knowledge (B = -0.91, SE = 0.15, *p* < .001, 95% CI: -1.21, -0.62). There was no difference in surprise between secondary and private knowledge (B = -0.18, SE = 0.15, *p* = .44). The effect of common knowledge on cooperation did not vary by age: the interactions between age and common knowledge-private knowledge (B = -0.50, SE = 0.39, *p* = .19, OR = 0.61, 95% CI: 0.28, 1.29), and age and common knowledge-secondary knowledge were not significant (B = -0.18, SE = 0.41, *p* = .66, OR = 0.84, 95% CI: 0.37, 1.88).

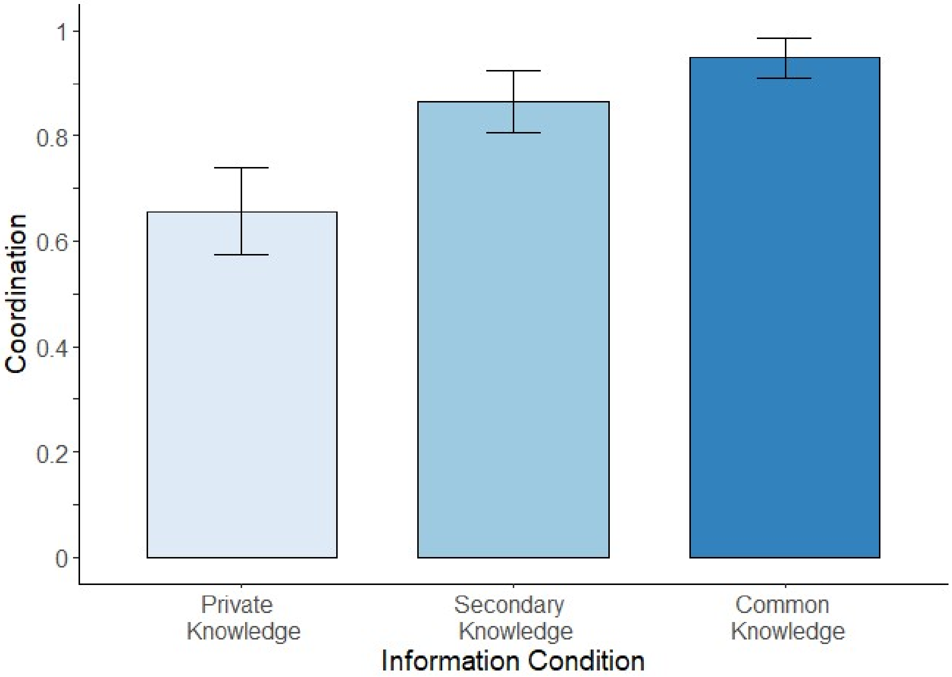


Figure 2. Proportion of participants who attempted to coordinate across the three information conditions (private knowledge, secondary knowledge, common knowledge). Error bars indicate 95% confidence intervals.

4. Discussion

We found that by 6 years of age, when children had common knowledge about the viability of the mutually beneficial joint payoff, they were more likely to cooperate, predict that their partner would cooperate, and were more certain in those predictions than when they had private or secondary knowledge. These results paint a fuller picture of how children solve coordination problems and the role that common knowledge plays in cooperative interactions. More generally, our finding that common knowledge allows children to solve coordination problems—likely by decreasing uncertainty about other agents' cooperative behavior—provides important insight into how cooperative behavior develops during childhood. Namely, this finding suggests that the improvements in cooperative behavior in middle childhood are possibly undergirded by the ability to understand and use common knowledge as it allows individuals to be more certain that others will cooperate, thereby reducing the riskiness of personally cooperating.

Our finding that participants expressed more surprise with their partner when there was common knowledge aligns with work by Siposova and colleagues that communicative looks—which are thought to establish common knowledge—created an expectation of collaboration in children (2018). Our findings extend this previous work, offering evidence for a proximate explanation for *how* common knowledge promotes cooperation. Namely, our results suggest that common knowledge likely promotes cooperation by increasing certainty in the likelihood of other agents' cooperative behavior. That is, when there is common knowledge, an agent can be more confident that their partner(s) will cooperate, which in turn makes the agent more likely to cooperate because it reduces the inherent risk of cooperating. However, because were unable to conduct our planned mediation model, we cannot make strong claims with our data regarding the causal role of certainty underlying the effect of common knowledge on cooperation. That said, this finding shows interesting convergence with adult work which has found that certainty in group members' cooperative behavior mediates the effect of common knowledge on cooperation (Deutchman et al., 2022),

Contrary to our initial predictions, we found no interaction between age and information condition: children were able to use common knowledge to coordinate by 6 years of age, the youngest age we examined. This finding comports with recent work which has found that 6-year-olds understand common knowledge (Siposova et al., 2021), as well as other work which indirectly suggests that it might be present by 5 years of age (Siposova et al., 2018; Grueneisen et al., 2015). Because some work even suggests that the ability to understand common knowledge emerges by 3-years of age (Köymen et al., 2014), it is possible that it is the ability to use common knowledge to guide behavior—rather than the ability *per se*—that emerges later in development. However, as our findings here suggest, this ability to understand cues to common knowledge and apply it to guide their cooperative behavior is already present by 6-years of age.

While the present study shines important light on the role of common knowledge in the development of cooperative behavior, it also raises a number of important questions and new directions for future work. First, while we chose scenarios that would be familiar to our population (sports) and that we believed would be meaningful in terms of their incentives (fun), it would be interesting to expand this line of inquiry in different domains, including incentivized behavioral tasks commonly used with adults. However, while behavior was hypothetical and not financially incentivized, that does not mean that children were not properly motivated in the task—units of "fun" are an ecologically valid motivator for children that, in some respects, are more consistent with their lived experience than more abstract monetary incentives—and work in adults suggests that whether rewards are real or hypothetical often has little impact on cooperative behavior (Ben-Ner et al., 2008; Locey et al., 2011; but see Vlaev, 2012). That said, while we anticipated high levels of cooperation given the nature of the cooperative interaction, it is possible that the unincentivized nature of our task may have additionally contributed to the relatively high levels of cooperation found here. Additionally, while children had to correctly answer comprehension check regarding the relative value of each payoff, given the unincentivized nature of the task, we cannot know for sure that they valued the individual activity ("a little fun") more than going to the park by themselves ("no fun"). However, because the incentives for cooperating were consistent across information conditions, we do not believe our results are dependent on the specific incentives used in the task.

Second, because our sample was drawn from a WEIRD population (i.e., the United States; Henrich et al., 2010), it would be valuable to replicate our findings cross-culturally in non-WEIRD populations in order to increase the generalizability of these findings. Given that recent cross-cultural work suggests that a sensitivity to intent—an ability largely assumed to be a universal feature of human moral psychology—as well as the development of theory of mind in fact varies substantially across societies (Barrett et al., 2016; Stengelin et al., 2020), it is possible that common knowledge does as well. If children from disparate societies use common knowledge to solve coordination problems, that might suggest that common knowledge is a universal ability underlying cooperative behavior. Lastly, our age range of 6-9 was relatively narrow—we chose this age range because past work suggests children cannot make inferences from common knowledge and consistently solve coordination problems until around 8-years of age (Grueneisen et al., 2015; Grueneisen & Tomasello, 2019; Soley & Koseler, 2021). However, given that we did not find an effect of age in our study, future work should extend the question studied here to 4-5-year-olds in order to identify the timepoint in development when children begin to use cues of common knowledge to coordinate.

While our results suggest that common knowledge likely plays a critical role in cooperative interactions, it is not the only mechanism that might do so. It is possible that shared experiences—such as co-attending to the messenger in the common knowledge condition—potentially increased cooperation through a different mechanism than certainty (e.g., increased liking of the target); future work should investigate this relationship between shared experience, common knowledge, and cooperation. Additionally, other forms of mentalizing, such as higher-order belief representation—beliefs about others beliefs—might also play a role as evinced by our finding that children cooperated more when they had secondary knowledge than private knowledge. Future work should examine the relative importance of common knowledge and higher-order beliefs in cooperative behavior, as well as whether they are distinct abilities or whether higher-order beliefs support common knowledge. Future work should also investigate when in development common knowledge emerges relative to higher-order belies—if common knowledge is present in ontogeny before higher-order belief representation, that would support the notion that it is a distinct form of mentalization.

A critical aspect of human social development is acquiring the ability to cooperate and work with others. Throughout their development, children frequently encounter coordination problems in which cooperation is mutually beneficial but risky, requiring the need to anticipate the behavior of other agents, such as taking turns during free play or working together towards a common goal. Our findings suggest that children solve these cooperative problems much like adults, relying on common knowledge to increase their confidence in other agents' cooperative behavior and thereby reduce the riskiness of cooperating. These results showcase common knowledge as an important mechanism for promoting cooperation, even as early as childhood and, broadly, helps refine our understanding of the psychology supporting human cooperation.

6. References

Apperly, I. A. (2012). What is "ToM"? Concepts, cognitive processes and individual differences. *Quarterly Journal of Experimental Psychology*, *65*(5), 825-839.

Ashley, J., & Tomasello, M. (1998). Cooperative problem-solving and teaching in preschoolers. *Social Development, 7*(2), 143-163.

Barrett, H. C., Bolyanatz, A., Crittenden, A. N., Fessler, D. M., Fitzpatrick, S., Gurven, M., Henrich, J., Kanovsky, M., Kushnick, G., Pisor, A., Scelza, B., Stich, S., von Rueden, C., Zhao, W., & Laurence, S. (2016). Small-scale societies exhibit fundamental variation in the role of intentions in moral judgment. *Proceedings of the National Academy of Sciences, 113*(17), 4688-4693.

Ben-Ner, A., Kramer, A., & Levy, O. (2008). Economic and hypothetical dictator game experiments: Incentive effects at the individual level. *The Journal of Socio Economics, 37*(5), 1775-1784.

Bohn, M., & Köymen, B. (2018). Common ground and development. *Child Development Perspectives, 12*(2), 104-108.

De Freitas, J., Thomas, K., DeScioli, P., & Pinker, S. (2019). Common knowledge, coordination, and strategic mentalizing in human social life. *Proceedings of the National Academy of Sciences, 116*(28), 13751-13758.

Deutchman, P., Amir, D., Jordan, M., McAuliffe, K. (2022). Common Knowledge Promotes Cooperation in the Threshold Public Goods Game by Reducing Uncertainty. *Evolution and Human Behavior, 42*(2), 155-167.

Eckerman, C. O., Davis, C. C., & Didow, S. M. (1989). Toddlers' emerging ways of achieving social coordinations with a peer. *Child Development*, 440-453.

Etel, E., & Slaughter, V. (2019). ToM and peer cooperation in two play contexts. *Journal of Applied Developmental Psychology*, *60*, 87-95.

Frith, C., & Frith, U. (2005). ToM. *Current Biology*, *15*(17), R644-R645. https://doi.org/10.1016/j.neuron.2006.05.001

Gallagher, H. L., & Frith, C. D. (2003). Functional imaging of 'ToM'. *Trends in Cognitive Sciences*, *7*(2), 77-83.

Goldvicht-Bacon, E., & Diesendruck, G. (2016). Children's capacity to use cultural focal points in coordination problems. *Cognition, 149*, 95-103.

Grueneisen, S., Wyman, E., & Tomasello, M. (2015). Children use salience to solve coordination problems. *Developmental Science*, *18*(3), 495-501.

Grueneisen, S., Wyman, E., & Tomasello, M. (2015). "I know you don't know I know…" Children use second-order false-belief reasoning for peer coordination. *Child Development*, *86*(1), 287-293.

Grueneisen, S., & Tomasello, M. (2019). Children use rules to coordinate in a social dilemma. *Journal of Experimental Child Psychology*, *179*, 362-374.

Grueneisen, S., & Tomasello, M. (2020). The development of coordination via joint expectations for shared benefits. *Developmental Psychology*, *56*(6), 1149.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world?. *Behavioral and Brain Sciences, 33*(2-3), 61-83.

Kachel, U., & Tomasello, M. (2019). 3-and 5-year-old children's adherence to explicit and implicit joint commitments. *Developmental Psychology, 55*(1), 80.

Köymen, B., Mammen, M., & Tomasello, M. (2016). Preschoolers use common ground in their justificatory reasoning with peers. Developmental Psychology, 52(3), 423 429.

Köymen, B., Rosenbaum, L., & Tomasello, M. (2014). Reasoning during joint decision makingby preschool peers. *Cognitive Development, 32*, 74-85.

Liebal, K., Behne, T., Carpenter, M., & Tomasello, M. (2009). Infants use shared experience to interpret pointing gestures. *Developmental Science, 12*(2), 264-271.

Locey, M. L., Jones, B. A., & Rachlin, H. (2011). Real and hypothetical rewards. *Judgment and Decision making*, *6*(6), 552.

Moll, H., Carpenter, M., & Tomasello, M. (2007). Fourteen-month-olds know what others experience only in joint engagement. *Developmental Science, 10*(6), 826-835.

R Core Team (2021). R: A language and environment for statistical computing. R Foundation forStatistical Computing, Vienna, Austria.

Rubinstein, A. (1989). The electronic mail game: Strategic behavior under "almost common knowledge." The American Economic Review, 79, 385–391.

Sheskin, M., Scott, K., Mills, C. M., Bergelson, E., Bonawitz, E., Spelke, E. S., Fei-Fei, L., Keil, F., Gweon, H., Tenenbaum, J., Jara-Ettinger, J., Adolph, K., Rhodes, M., Frank, M., Mehr, S., & Schulz, L. (2020). Online developmental science to foster innovation, access, and impact. *Trends in Cognitive Sciences, 24*(9), 675-678.

Siposova, B., Tomasello, M., & Carpenter, M. (2018). Communicative eye contact signals a commitment to cooperate for young children. *Cognition*, *179*, 192-201.

Siposova, B., Grueneisen, S., Helming, K., Tomasello, M., & Carpenter, M. (2021). Common knowledge that help is needed increases helping behavior in children. *Journal of Experimental Child Psychology*, *201*, 104973.

Snidal, D. (1985). Coordination versus prisoners' dilemma: Implications for international cooperation and regimes. *American Political Science Review, 79*(4), 923-942.

Soley, G., & Köseler, B. (2021). The social meaning of common knowledge across development. *Cognition, 215*, 104811.

Stengelin, R., Hepach, R., & Haun, D. B. (2020). Cultural variation in young children’s social motivation for peer collaboration and its relation to the ontogeny of Theory of Mind. *PloS One, 15*(11), e0242071.

Thomas, K. A., DeScioli, P., Haque, O. S., & Pinker, S. (2014). The psychology of coordination and common knowledge. *Journal of Personality and Social Psychology, 107*, 657-676.

Thomas, K. A., DeScioli, P., & Pinker, S. (2018). Common knowledge, coordination, and the logic of self-conscious emotions. *Evolution and Human Behavior*, *39*, 179-190.

Tsoi, L., Dungan, J., Waytz, A., & Young, L. (2016). Distinct neural patterns of social cognition for cooperation versus competition. *NeuroImage*, *137*, 86-96.

Tsoi, L., Hamlin, J. K., Waytz, A., Baron, A. S., & Young, L. L. (2021). A cooperation advantage for theory of mind in children and adults. *Social Cognition*, *39*(1), 19 40.

Vlaev, I. (2012). How different are real and hypothetical decisions? Overestimation, contrast and assimilation in social interaction. *Journal of Economic Psychology*, *33*(5), 963-972.

3. Study 3: People update their injunctive norm and moral beliefs after receiving descriptive norm information

Recent work suggests an association between descriptive norms—what we think other people commonly do—and injunctive norms—what we think other people dis/approve of. What role do descriptive norms play in forming injunctive beliefs and what does that tell us about the cognitive processes underlying social norm cognition? Across six studies (*N*=2,671), we examined whether people update their injunctive norm beliefs—as well as their moral judgements and behavioral intentions—after receiving descriptive norm information about how common (or uncommon) a behavior is. Specifically, we manipulated the descriptive normativity of behaviors, describing behaviors as being weakly (20% of people were doing the behavior) or strongly (80% of people were doing the behavior) normative. To measure belief updating, we assessed beliefs prior to and after receiving information about the descriptive norm. We find that participants updated their injunctive norm beliefs, moral judgements, and behavioral intentions after receiving descriptive norm information and did so to a greater extent for strong compared to weak descriptive norms. Descriptive norms also influenced participants' injunctive beliefs more than their moral judgements, indicating that injunctive norms beliefs are distinct from moral judgements. Both injunctive norms and moral judgements partially mediated the effect of descriptive norms on behavioral intentions, suggesting that descriptive norms influence our behavior in part by changing our beliefs about what others approve of and find moral. Together, our findings suggest that descriptive norms play an influential role in shaping our injunctive norm beliefs and moral judgements and help to paint a fuller picture of the social cognition of social norms.

This paper is co-authored with Gordon Kraft-Todd, Liane Young, and Katherine McAuliffe.

1. Introduction

Why is it appropriate to serve chocolate muffins for breakfast in the USA while serving chocolate cupcakes seems odd? Why are eggs a breakfast staple in restaurants across America but are rarely seen on dinner menus? That we accept these unspoken social rules, often without thinking or questioning them, demonstrates the prevalence of social norms in human life. There's no scientific or health-related reason why it's acceptable to eat chocolate muffins for breakfast but eating chocolate cupcakes for breakfast is uncommon, and for the most part, frowned upon. The only difference in this case are social norms: eating muffins for breakfast is acceptable because we have social norms that dictate what is and what isn't acceptable breakfast food. Social norms are a foundational part of human societies and pervade nearly every aspect of human social life—from what we eat for breakfast to how to share resources. However, while social norms are a foundational part of social interactions and have been a central focus of study in social psychology for over 40 years (Axelrod, 1986; Buffalo & Rodgers, 1971), we know surprisingly little about how they arise and change over time. Given the importance of norms in human social life, it is important to study how norms are formed and come to influence our behavior and beliefs.

1.1. Relationship between norms

One prominent account of how norms arise holds that they are socially conditional—we conform to a behavior because other people in our group do it and expect that others will do so as well (Bicchieri, 2005; Bicchieri, 2016). This suggests that key ingredients in the development of social norms are our beliefs about: 1) what others do; and 2) what others think people should do. Indeed, much of the recent research on social norms has examined two related types of normative information: descriptive norms, what we think other people are actually doing (norms of is), and injunctive norms, what we think other people approve or disapprove of (norms of ought; Cialdini et al., 1990). For example, that most people speak quietly or whisper in a library is a descriptive norm (it's what people do) while the belief that most people approve of talking quietly is an injunctive norm (it's what people approve of). Both descriptive and injunctive norms can influence behavior in important ways (Cialdini et al., 1991; Cialdini et al., 2006) and are generally congruent—that is, most people generally approve of what is commonly done (Bear & Knobe, 2017; Blanton et al., 2008). For example, if you see there's a line to get into a store, most people would think that you should go to the back of the line and most people would actually do so as well.

While descriptive and injunctive norms are generally thought of as distinct constructs (Cialdini et al., 2006; Reno et al., 1993), recent work highlights the ways in which they are highly interrelated. Eriksson and colleagues have found evidence for an association between descriptive and injunctive norms (2015). Specifically, using the Implicit Association Test (IAT), they found that people show an automatic association between concepts that are descriptive and injunctive. This association is not just implicit: they found that people also made explicit bi-directional inferences between descriptive and injunctive norms: when told that a behavior is common (or uncommon), people infer it is injunctive (or not injunctive), and vice-versa. Additionally, their results suggest that information about a society's descriptive norm directly influenced participants' own moral judgements of the characters engaging in those behaviors. Other work has more directly explored how descriptive norms intersect with moral judgements of behavior. The "common is moral" (CIM) heuristic is the hypothesis that the frequency or commonness of a behavior influences its perceived moral status (Lindstrom et al., 2018). In support of this theory, researchers have found that both prosocial and selfish behaviors are evaluated as more moral when common than when rare, suggesting that we infer morality from the frequency of behaviors. Taken together, this work indicates that there is a strong association between descriptive norms, injunctive norms, and moral judgements.

Other recent work has found stronger evidence of a directional inference from descriptive to injunctive norm beliefs. Namely, people tend to infer what ought to be (injunctive inferences) from what is typical (descriptive norms; Tworek & Cimpian, 2016). This "ought-to-is" relationship is present early in development—by 4 years of age, children infer injunctive judgements about how a novel social group should behave based on common behavior in the group (Roberts et al., 2017), and this effect replicates across cultures (Roberts et al., 2018). Other work finds that already by age 6, children's injunctive norm beliefs are influenced by descriptive norm information that a behavior is common (Deutchman et al., preprint). Altogether, this work suggests that we implicitly associate and explicitly infer the injunctiveness of a behavior from how common it is, and vice versa. However, while descriptive and injunctive norms are largely congruent, they can also also dissociate (e.g., most people might think you should recycle but many do not always do so consistently) and while uncommon, they can also come in conflict (e.g., most people might think you should conserve energy but crank the AC when it's hot out). That descriptive and injunctive norms can occasionally dissociate has led some researchers to conceptualize them as distinct constructs. Given this relatively mixed evidence about the relationship between injunctive and descriptive norms, it’s important to study how these norms relate to one another and when they do not in order to gain a more nuanced understanding of social norm cognition.

While there is some initial evidence that people infer injunctive normativity from descriptive information—suggesting that descriptive norms might partially contribute to injunctive norm formation—it remains unclear exactly how they might do so and to what extent. For example, although past work suggests that people make basic, explicit binary inferences between descriptive and injunctive norms (e.g., is a behavior injunctive or not given that it is common or uncommon; Eriksson et al., 2015), it’s unclear how the strength of the descriptive norm (e.g., the number or proportion of people actually engaging in the norm) influences the extent to which we think others approve of a behavior. In other words, there might be a meaningful difference in the inferences we make about how injunctive a behavior is given a descriptive norm where 20% vs. 80% of people are engaging in it. Furthermore, little work in this area to date has examined beliefs before and after receiving normative information. Utilizing a repeated measures design can allow us to better understand how individuals adjust their beliefs in response to novel norm information. Here we investigate this updating process in detail to better understand what amount of descriptive norm information is required to change injunctive norm beliefs. A more fine-grained approach to exploring the relationship between descriptive and injunctive norms will allow us to better understand how descriptive norm information specifically changes the strength of injunctive beliefs, and in the process, will help reveal how closely associated these concepts are as well as how they interact to influence behavior, and will inform our understanding of the cognitive processes underlying norm cognition

Additionally, to help generate a clearer picture of the relationship between descriptive and injunctive norms and how they influence beliefs, it is also important to have a better understanding of how they relate to another important feature of social norm cognition—moral judgements. While injunctive norms are generally conceived of as moral (Russell et al., 2021; Lu et al., 2020), it is possible that morality and injunctive norms are dissociable, much like the relationship between descriptive and injunctive norms. While in practice they often overlap, it is possible for injunctive norms to not be perceived as morally good. For example, most people think you should eat dessert after dinner, yet many people would agree it is not necessarily morally wrong to not do so. To date, very little work has examined the relationship between descriptive and injunctive norm beliefs and moral judgments within one experimental design. Doing so will help us to better understand how these constructs are associated with one another and interact to influence behavior. If introducing a descriptive norm influences injunctive norm beliefs and moral judgements to a similar extent, this would suggest that injunctive norms and moral judgements are highly-overlapping constructs. On the other hand, if descriptive norms differentially influence injunctive norm beliefs and moral judgements, such that descriptive norm information more strongly influences injunctive norm beliefs than moral judgements, that would provide some evidence that injunctive norms and morality are related but distinct concepts. Furthermore, because people's personal moral judgements about a behavior might vary from their beliefs about others' moral judgements, here we explore how descriptive and injunctive norm beliefs relate to both first- and second-order moral judgements. If people's beliefs about what others find moral are more influenced by descriptive norms than their personal moral beliefs, that would suggest there is a key distinction between first- and second-order moral judgements in norm cognition.

Lastly, it is also important to understand how descriptive and injunctive norms relate to behavior. A large body of work has found that descriptive and injunctive norms influence behavior in important ways (Cialdini et al., 1990; Cialdini et al., 2006; Elek et al., 2006; Raihani & McAuliffe, 2014; Reno et al., 1993). However, it is less clear whether people are more likely to engage in a behavior after receiving new information about the descriptive norm. Thus, while not the main focus of this paper, we also explored whether people update their behavioral intentions in response to descriptive norms and how behavioral intentions relate to injunctive norm beliefs and moral judgements. While self-report behavioral intentions are an imperfect proxy of actual behavior, measuring intentions can still provide important information about how descriptive and injunctive norm beliefs relate to behavior.

1.3. Present Study

In the present set of experiments, we explored whether people update their injunctive norm beliefs, moral judgements, and behavioral intentions after receiving descriptive norm information about how common (or uncommon) the behavior is to better understand the relationship between descriptive and injunctive norms, moral judgements, and behavior. To investigate this, we presented participants with a series of vignettes detailing different normative behaviors. Specifically, we manipulated whether there was a weak descriptive norm, in which 20% of people in the vignette were doing the behavior, or a strong descriptive norm, where 80% of people were doing the behavior. To measure belief updating, we assessed beliefs prior to and post receiving information on the descriptive norm. Because previous research suggests that people view different kinds of norms as psychologically distinct–such as between moral norms and social conventions, for example (Smetna, 2013; Smetna et al., 2014)—we also manipulated the categories of behaviors participants saw. For all analyses reported here, we collapse across behaviors in order to examine whether belief updating is robust to different behaviors. We report the between-behavior updating findings in detail in a separate paper (Deutchman et al., in prep).

Answering these questions will provide important insight into the social cognition underlying social norms. Namely, if we find that people readily update their injunctive norm beliefs and moral judgements after receiving descriptive norm information, that would provide some of the strongest evidence to date that descriptive norms, injunctive norms, and moral judgements are highly related concepts—possibly speaking to the extent to which they align in everyday life or tap into a latent, underlying norm construct. Furthermore, if we find that people update their beliefs after receiving descriptive information, that would suggest that there is a strong directional effect of descriptive norms on injunctive norm beliefs and moral judgements such that descriptive norm information plays an important role in the formation of injunctive norms and moral judgements. In other words, just seeing that many people are engaging in a certain behavior might lead us to infer that most people approve of this behavior and think it's morally good, which in turn, might influence our own beliefs and decision to comply with the behavior. Alternatively, if we find that people do not update their injunctive norm beliefs or moral judgements, that would suggest that descriptive and injunctive norm beliefs and moral judgements are more distinct constructs than previously thought and indicate that descriptive norm information plays little to no role in shaping injunctive norm beliefs and moral judgements.

2. Overview of experiments

In six experiments we assessed whether people updated their injunctive norm beliefs, moral judgements, and behavioral intentions after receiving novel information about a descriptive norm (Experiments 2-5). We manipulated the descriptive norm such that it was either common (strong descriptive norm) or uncommon (weak descriptive norm) to explore how the relative strength of descriptive norm information influences beliefs (Experiments 1-5). Additionally, to better understand how descriptive norms influence behavior, we explored whether injunctive norm beliefs and moral judgements mediated the effect of descriptive norm condition on behavioral intentions (Experiments 2-5).

2.1. Overview of methods used across experiments

Because the experimental design was similar across studies, we describe all studies in parallel in order to highlight their similarities and differences. See Table 1 for a summary. Across all studies, we randomly assigned participants to one of two descriptive norm conditions. To measure updating, participants read vignettes and answered the dependent measures before and after receiving the descriptive norm information (Studies 2-5) while they only answered once post-descriptive norm information in Experiment 1.

In Experiment 1, we assessed people's injunctive norm beliefs, moral judgements, and their behavioral intentions on six conventional norms when there was either a strong descriptive norm that the behavior was common or a weak descriptive norm that the behavior was uncommon. In Experiment 2, we measured belief updating by assessing normative beliefs prior- to and post-receiving the descriptive norm information (strong or weak).

In Experiments 2-5, we varied the type of norm to understand whether updating differs depending on the category of behavior (Deutchman et al., in prep). Here, they serve as a set of studies testing whether the effect of descriptive norms on updating is robust to different kinds of normative behaviors. Thus, Experiment 3 aimed to replicate the updating results found in Experiment 2 across a greater range of behaviors.

Our goal in Experiments 4a-b was to replicate the updating effects found previously after ruling out potential design effects on participant responses. Specifically, our aim in Experiment 4a was to replicate the updating effects using a between-subjects design to rule out the possibility that updating was influenced by demand characteristics from a within-subject design. Having replicated the updating differences across vignettes with a fully between-subject design, we return to treating vignette type as a within-subjects variable in all remaining studies. In Experiment 4b, we wanted to test whether our injunctive norm updating results were robust to the specific wording of the measure. Because previous work has operationalized injunctive norms in several different ways (Bicchieri, 2016; Cialdini et al., 1990; Lu et al., 2020; Russell et al., 2021), we replaced the previous injunctive measure with a new one assessing the extent to which participants believe that other people think you should engage in the behavior.

We had three goals for Experiment 5: we aimed to (1) replicate belief updating across norm types using a new set of harm behaviors and both injunctive norm measures in one study, (2) validate that participants perceived our norm types as actually falling into the hypothesized categories, and (3) explore whether the vignette rating measures predicted updating (or the lack thereof) in the norm updating task. To that end, we included all four sets of vignettes used in earlier studies plus a new set of harm vignettes (fairness, convention, harm-old, harm-new, preference), both injunctive measures as dependent variables (injunctive-approve, injunctive-should), and a new vignette rating task.

Lastly, we combined the data from Experiments 2–5 to examine whether people updated their beliefs and behavioral intentions after receiving descriptive norm information and whether they update to a greater extent for strong than weak descriptive norms. Examining these questions across all experiments provided a more definitive answer to our hypotheses by allowing us to control for between-study variability.

Table 1. Table with design information on all studies.

Study	Design	Norm Type	Participants
Experiment 1	2 (Descriptive norm: weak, strong) between subject	Conventional	N = 401 participants (M_{age} = 40.09, F = 46.6%)

Experiment 2	2 (Descriptive norm: weak, strong) between subject × 2 (Vignette type) within-subject	Fairness, Conventional	N = 414 participants (M_{age} = 41.65, F = 52.42%)
Experiment 3	2 (Descriptive norm: weak, strong) between subject × 3 (Vignette type) within-subject	Fairness, Harm, Preference	N = 402 participants (M_{age} = 41.73, F = 50.99%)
Experiment 4a	2 (Descriptive norm: weak, strong) between subject × 3 (Vignette type) between-subject	Fairness, Harm, Preference	N = 643 participants (M_{age} = 41.08, F = 53.81%)
Experiment 4b	2 (Descriptive norm: weak, strong) between subject × 3 (Vignette type) within-subject	Fairness, Harm, Preference	N = 411 participants (M_{age} = 43.03, F = 60.83%)
Experiment 5	2 (Descriptive norm: weak, strong) between subject × 5 (Vignette type) within-subject	Conventional, Fairness, Harm-1, Harm-2, Preference	N = 400 participants (M_{age} = 41.33, F = 54.3%)

2.2. Participants

All studies were conducted online using Cloud Research and Amazon's Mechanical Turk (Arechar et al., 2017) with participants from the United States who received between $0.60 - $3.15 to complete a survey of variable length depending on study (see Table 1 in supplement for study specific information). The sample size for Experiment 1 was based on previous work using the same platform (Deutchman et al., 2022; Dungan et al., 2017). In Experiments 2-5, we determined the study sample size and statistical power using the R package simr to conduct power simulations with 5,000 bootstraps. In all power simulations we used the observed effect size from the proceeding study with the exception of Experiment 5 which was based on the observed effect from Experiment 4a[8]. All studies were powered to have a minimum of 98.64% power to detect an effect of the key outcome (Experiments 1-2: descriptive norm condition; Experiments 3-5: descriptive norm condition × vignette type interaction). In total across all studies we recruited N = 2,671 (53.82% female, $M_{Age} = 41.77$). See Table 1 for the sample size and demographic information for each experiment. While we initially recruited N = 2,921, we excluded N = 250 participants across all studies for failing our preregistered exclusion criteria which included an attention check, reporting they answered more than 2-3 questions 'with little or no thought put into them', or failing to complete the study in its entirety (see preregistrations for study specific exclusion criteria and supplement for exclusions per study).

2.3. Design & Procedure

[8] We used the effect size from Experiment 4a because the vignette conditions were also within-subject in Experiment 5 while they were between-subjects in Experiment 4b.

In Experiment 1, participants were presented with a series of vignettes which they rated on a number of dependent variables (see Measures below). In Experiments 2-5, participants answered the dependent variables twice—once prior to receiving the descriptive norm conditions (their *priors*) and again after receiving the descriptive norm information. After answering their priors, but before receiving the descriptive norm information, participants completed a filler task consisting of simple trivia questions. We included this task to serve as a buffer between the two main parts of the experiment to help reduce demand effects. After this filler task, participants received the vignettes again with the norm information in a random order that varied from how they first received them. We created our key dependent measure of belief updating for Experiments 2-5 by subtracting participants prior norm ratings from their post-descriptive norm ratings. The order of all the dependent measures (with the exception of injunctive certainty which was always last) was randomized for each vignette. The binary behavioral intention question always came after the other measures as we did not have *a priori* predictions for this measure, and because it preceded an open-response question asking participants to explain their choice to engage in the behavior or not in Experiments 1 & 2.

In all studies, participants were randomly assigned to one of two descriptive norm conditions between-subjects—a *weak descriptive norm* condition, where 20% of people were engaging in the behavior, or a *strong descriptive norm* condition, where 80% of people were engaging in the behavior. The strength of the descriptive norm was presented as a proportion out of a total number of people in the scenario (e.g., 4 out of 5 people are talking in the library). We varied the denominator across vignettes (strong: 4/5, 8/10, 16/20; weak: 1/5, 2/10, 4/20) in order to increase generalizability and realism of the scenarios. The denominator was consistent between the weak and strong norm vignettes (see supplement for vignette text).

Participants saw a number of vignettes in each study but the exact number of vignettes, and the types of behaviors the vignettes included varied across studies (see Table 1). In Studies 1-3, we included positive- and negatively-valenced behaviors; valence was determined during vignette norming (see below for more information). However, because many of the positive behaviors were at ceiling on our dependent measures, we only included negatively valenced behaviors from Experiment 4a onward. Across all the studies we included four different kinds of vignettes: conventional, fairness, harm, and preferences.

Prior to inclusion in the studies presented here, all the vignettes were normed on Mechanical Turk to ensure they were consistent on several potentially relevant dimensions such as the cost, benefit, frequency, injunctive normativity, descriptive normativity, morality, etc. Only the most closely related behaviors on all norming dimensions were included. See the supplement for details of our norming procedure.

In Experiment 5 only, participants completed a vignette rating task in which they received all ten vignettes in the survey, spanning all four norm types, without descriptive norm information. They rated each vignette on the extent to which the behavior pertained to fairness, harm, convention, preference, severity, self-other impact, and where it fell on a spectrum between fairness and harm (see Measures below). We included this task to 1) validate that the behaviors in our vignettes did indeed fall into our predicted norm categories, and 2) explore whether these ratings predicted belief updating across vignettes in the updating task. The order of presentation of the vignette rating task and the belief updating task was counterbalanced across participants.

2.4. Measures

2.4.1. Injunctive Normativity

Participants rated their beliefs about the extent to which the behavior in each vignette was injunctively normative on a 0 – 100-point sliding scale (0 – Definitely not approve, 100 – Definitely approve). In Experiment 1 only this question was on a 1 – 7-point Likert scale): "In general, would most people approve of X". Studies 4b & 5 additionally included a new injunctive measure with a different common operationalization of injunctive normativity: "To what extent do other people think you should X" (0 – Not at all, 100 – A great deal).

2.4.2. Injunctive Certainty

In Studies 3-5, participants rated how certain they were in their injunctive beliefs after answering the initial dependent variable block containing all other measures. This question was on a 0 –100-point sliding scale (0 – Extremely uncertain, 100 – Extremely certain) and as a reminder, contained an image of the injunctive norm question and their answer to the question: "Previously you were asked the question. You answered [injunctive norm response]. How certain are you in that response?"

2.4.3. Descriptive Normativity

In all studies participants rated their beliefs about the descriptive norm to serve as a manipulation check to ensure that the descriptive norm conditions were in fact influencing descriptive norm beliefs. Studies 2-5 asked this on a 0 – 100-point sliding scale (0 – No one, 100 – Everyone): “In general, how many people would walk on the grass?” The Experiment 1 descriptive norm measure was identical but was on a 1 – 7-point Likert scale (1 – No one, 7 – Everyone).

2.4.4. Behavioral Intentions

Participants rated their behavioral intentions regarding how likely they would be to comply with the norm on a 0 – 100-point sliding scale (0 – Extremely unlikely, 100 – Extremely likely): “How likely would you be to X?” Experiment 1 asked the same question on a 1 – 7-point Likert scale (1 – Extremely unlikely, 7 – Extremely likely). In Experiments 2 – 4b, participants also answered a binary behavioral intention question at the end of the dependent variable block asking “If you were in this situation, what would you do” (0 – I would not X, 1 – I would X). In Studies 1 and 2, participants also answered an open-response question probing their rationale for their decision in the binary intention question which always preceded it: “Why did you choose that option”.

2.4.5. Morality

In Experiment 1, participants answered how morally wrong engaging in the behavior would be on a 1 – 7-point Likert scale (1 – Not wrong at all to 7 – Extremely wrong): “How morally wrong is it for someone to X in this scenario”. Studies 2 – 5 measured morality with two different questions, both on 0 – 100-point sliding scales anchored from

0 – Extremely immoral to 100 – Extremely moral: self-morality and other-morality. Self-morality assessed how moral participants personally found a behavior: “How moral do you personally think it is for someone to X?” Other-morality assessed second-order moral judgements about how moral participants thought others found the behavior: “How moral do you think other people think it is for someone to X?”

2.4.6. Punishment

In Experiment 1, participants rated how much of a reward or punishment they would give someone who engaged in the behavior on a 1 – 7-point Likert scale (1 – Large punishment, 7 – Large reward): “If you were in a position to reward or punish someone for engaging in this behavior, how much of a punishment or reward would you give someone who was X in this scenario?” In Experiment 4b, participants rated how much someone engaging in the behavior should be punished on a 0 – 100-point sliding scale (0 – Not punished at all, 100 – Severely punished): “To what extent should someone be punished for X?”

2.4.7. Rating Task

In Experiment 5, participants completed a vignette rating task in addition to the updating task. Here they rated all ten vignettes included in this study on the extent to which they pertained to the following measures. All measures were on a 0 – 100-point sliding scale. Fairness: “To what extent does X relate to fairness?” Harm: “To what extent does X involve harming others?” Conventionality: “To what extent is X a social convention?” Preference: “To what extent is X a personal preference?” These four items were anchored from 0 – Not at all, to 100 – Entirely. Participants also rated the severity of engaging in the behavior (0 – Not bad at all, 100 – Extremely bad): “How bad is it to X?” Self-other impact (0 – Only impacts yourself, 100 – Only impacts others): “To what extent does X impact others as compared to yourself?” Fair-harm scale (0 – Unfair, 100 – Harmful): “Where does X fall on a scale from unfair to harmful?”

2.5. Analytic approach

In each individual experiment we conducted a series of preregistered linear mixed effects regression models (LMEM) predicting the dependent measures by descriptive norm condition using the lme4 package in R (Bates et al., 2015; R Core Team, 2022). All models included participant and vignette identity as random effects. In all studies with the exception of Experiment 1, the models included the dependent measure difference score which was found by subtracting the prior ratings from the ratings post-descriptive norm information. Experiment 1 did not collect prior ratings and so we could not compute difference scores. For example, if a participant rated their likelihood of talking in the library as 30 prior to receiving the descriptive norm condition, and then rated it as 50 after receiving information that the behavior is common, their behavioral intention updating score would be +20. We also ran a series of unplanned multilevel regression models subsetting the data by descriptive norm condition (weak, strong) and with the un-transformed dependent variable ratings in place of the difference scores to examine whether participants updated their prior beliefs after receiving either the strong or the weak descriptive norm. Due to the design differences between Experiment 1 and the other studies, we only report the results of Experiments 2-5 here and report Experiment

1's results in the supplement. However, we note here that the results of Experiment 1 are consistent with the findings from the other studies: participants' injunctive beliefs, behavioral intentions, and personal and other moral judgements were more influenced by the strong than the weak descriptive norm, suggests that participants' beliefs were sensitive to novel descriptive norm information. For the sake of simplicity, we only include the belief updating results here. Additionally, given the complexity of the between-vignette updating results, we collapsed across vignette type for all analyses reported here to focus solely on the question of whether participants are updating their beliefs after receiving the descriptive norm. Thus, we do not report any vignette comparisons here but analyze them in detail in a separate manuscript in order to fully unpack the complex relationship between belief updating and behavior type (Deutchman et al., in prep). For Experiment 5 only, we ran a series of LMEMs predicting vignette rating (fairness, harm, conventionality, preference, severity, self-other impact) by vignette type. For these models, we set the relevant vignette as the reference category (e.g., the fairness vignette type was set as the baseline when comparing vignettes on fairness) and we set harm as the reference for the severity and self-other impact models. We present all results for the individual studies in detail in the supplement. We summarize the key results—the main effect of descriptive norm condition on updating—in Table 2.

3. Results

3.1. Individual Experiment Results

3.1.1. Do people update their beliefs and behavior after receiving descriptive norm information?

Across all studies we find support for our prediction that people update their beliefs in response to novel descriptive norms. We first report the updating results for the strong descriptive norm. In five out of five studies, we found that participants updated their injunctive beliefs—participants rated the behavior as more injunctive after finding out it was common. In four out of five studies, participants updated their behavioral intentions of engaging in the behavior after receiving the strong descriptive norm. In five out of five studies, we found that, after receiving a strong descriptive norm, participants updated their beliefs about how moral other people found the behavior. In contrast, participants only updated their personal moral beliefs after receiving the strong descriptive norm in one out of five studies.

Turning to the weak descriptive norm updating, we found that in five out of five studies, participants updated their injunctive norm beliefs after receiving the weak descriptive norm that the behavior was uncommon. In five out of five studies participants updated their behavioral intentions of engaging in the behavior after receiving the weak descriptive norm information. Participants updated their beliefs about how moral other people found the behavior in four out of five studies after receiving the weak descriptive norm. Lastly, in four of five studies, participants updated their personal moral beliefs after receiving the weak descriptive norm. On average, we found that participants *positively* updated their beliefs—finding the behavior more injunctive or moral—after receiving a strong descriptive norm that it is common and that they *negatively* updated their beliefs—finding a behavior less injunctive or moral—after receiving weak descriptive norms that it was uncommon. For the injunctive norm, behavioral intention, and other-morality measures, participants positively updated to a larger extent for strong descriptive norms than they negatively updated their beliefs for weak descriptive norms. However, in the case of their personal moral belief ratings, participants were more likely to negatively update their beliefs about the morality of the behavior whereas there was little evidence that they positively updated their beliefs after receiving a strong descriptive norm.

3.1.2. Do people update their beliefs and behavior more for strong descriptive norm information than weak descriptive norm information?

Across all five studies we find evidence to support our hypothesis that people will update their beliefs and behavior more after receiving strong descriptive norm information that a behavior is relatively common (80%) as compared to uncommon (20%). In five out of five studies, we found that participants updated their injunctive beliefs more after receiving a strong descriptive norm than a weak descriptive norm. In all five studies, participants updated their behavioral intentions to engage in the behavior more for strong than weak descriptive norm information. Lastly, we found that in five out of five studies participants updated both their personal moral beliefs and their beliefs about others' morality more after receiving strong descriptive norms than weak descriptive norms.

3.1.3. Do vignette ratings predict belief updating?

Participants perceived that the fairness vignettes as pertaining more closely to fairness than the harm behaviors (B = -10.31, SE = 3.02, *p* = .03, 95% CI: -15.18, -5.45), conventional behaviors (B = -33.08, SE = 3.02, *p* < .001, 95% CI: -37.94, -28.21), and personal preferences (B = -57.01, SE = 3.02, *p* < .001, 95% CI: -61.87, -52.14). Similarly, the harm vignettes were rated as pertaining more to harm than the fairness (B = -35.85, SE = 10.88, *p* = .03, 95% CI: -52.98, -18.71), conventional (B = -53.10, SE = 10.88, *p* = .008, 95% CI: -70.24, -35.97), and preference behaviors (B = -79.95, SE = 10.88, *p* = .002, 95% CI: -97.08, -62.82). The conventional behaviors were rated more highly as concerning social conventions than the fairness (B = -16.18, SE = 3.95, *p* = .01, 95% CI: -22.61, -9.74) and harm behaviors (B = -21.85, SE = 3.95, *p* = .005, 95% CI: -28.29, -15.41), but not more than the preference (B = -0.62, SE = 3.95, *p* = .88, 95% CI: -22.61, -9.74). The preferences were perceived as more closely relating to personal preference than the fairness (B = -31.83, SE = 2.65, *p* < .001, 95% CI: -36.23, -27.42), harm (B = -40.86, SE = 2.65, *p* < .001, 95% CI: -45.26, -36.45), and conventional behaviors (B = -31.25, SE = 2.69, *p* < .001, 95% CI: -35.65, -26.85). The harm behaviors were also perceived as more severe than the conventional (B = -51.14, SE = 5.25, *p* < .001, 95% CI: -59.47, -42.81) and preference behaviors (B = -78.83, SE = 5.25, *p* < .001, 95% CI: -87.16, -70.50), but not more severe than the fairness behaviors (B = -9.70, SE = 5.25, *p* = .13, 95% CI: -18.03, -1.38). Next, looking at the perception that the behavior impacted the self vs. other, we find that the harm behaviors were viewed as impacting others marginally more than the fairness behaviors (B = -33.37, SE = 12.32, *p* = .05, 95% CI: -52.73, -14.00) but not the conventional behaviors (B = -16.09, SE = 12.32, *p* = .26, 95% CI: -35.45, 3.28). Preferences were perceived as affecting the self more than all other behaviors (all *p*’s < .017).

When examining whether vignette ratings predicted injunctive belief updating, we find that the fairness ratings, (B = .008, SE = .01, *p* = .43, 95% CI: -0.013, 0.029), conventional ratings (B = -.02, SE = .01, *p* = .08, 95% CI: -0.045, 0.002), harm ratings (B = .02, SE = .02, *p* = .20, 95% CI: -0.012, 0.055), and self-other impact ratings (B = 0.02, SE = 0.02, *p* = .27, 95% CI: -0.016, 0.052) did not predict injunctive updating. However, the preference (B = -0.04, SE = .01, *p* < .001, 95% CI: -0.066, -0.0173) and severity ratings (B = 0.07, SE = 0.02, *p* < .001, 95% CI: 0.039, 0.116) did predict injunctive norm updating such that perceiving the behaviors as being more like a personal preference led to less injunctive updating while viewing the behaviors as more severe led participants to greater injunctive belief updating.

Table 2. Table showing model estimates and significance levels of the descriptive norm condition term for the five key dependent measures across studies 2-5. All models included the difference score for each of the dependent variables.

	Experiment 2	**Experiment 3**	**Experiment 4a**	**Experiment 4b**	**Experiment 5**
Effect of Descriptive Norm Condition on Updating					
Injunctive - *Approve*	B = 25.71***	B = 43.37***	B = 23.25***		B = 30.27***
Injunctive – *Should*				B = 29.51***	B = 27.81***

Behavioral Intention	B = 9.35***	B = 29.87***	B = 5.59***	B = 4.00***	B = 5.92***
Self-morality	B = 6.09***	B = 30.28***	B = 2.23**	B = 1.47*	B = 2.06**
Other-morality	B = 18.61***	B = 35.59***	B = 15.93***	B = 16.16***	B = 21.04***

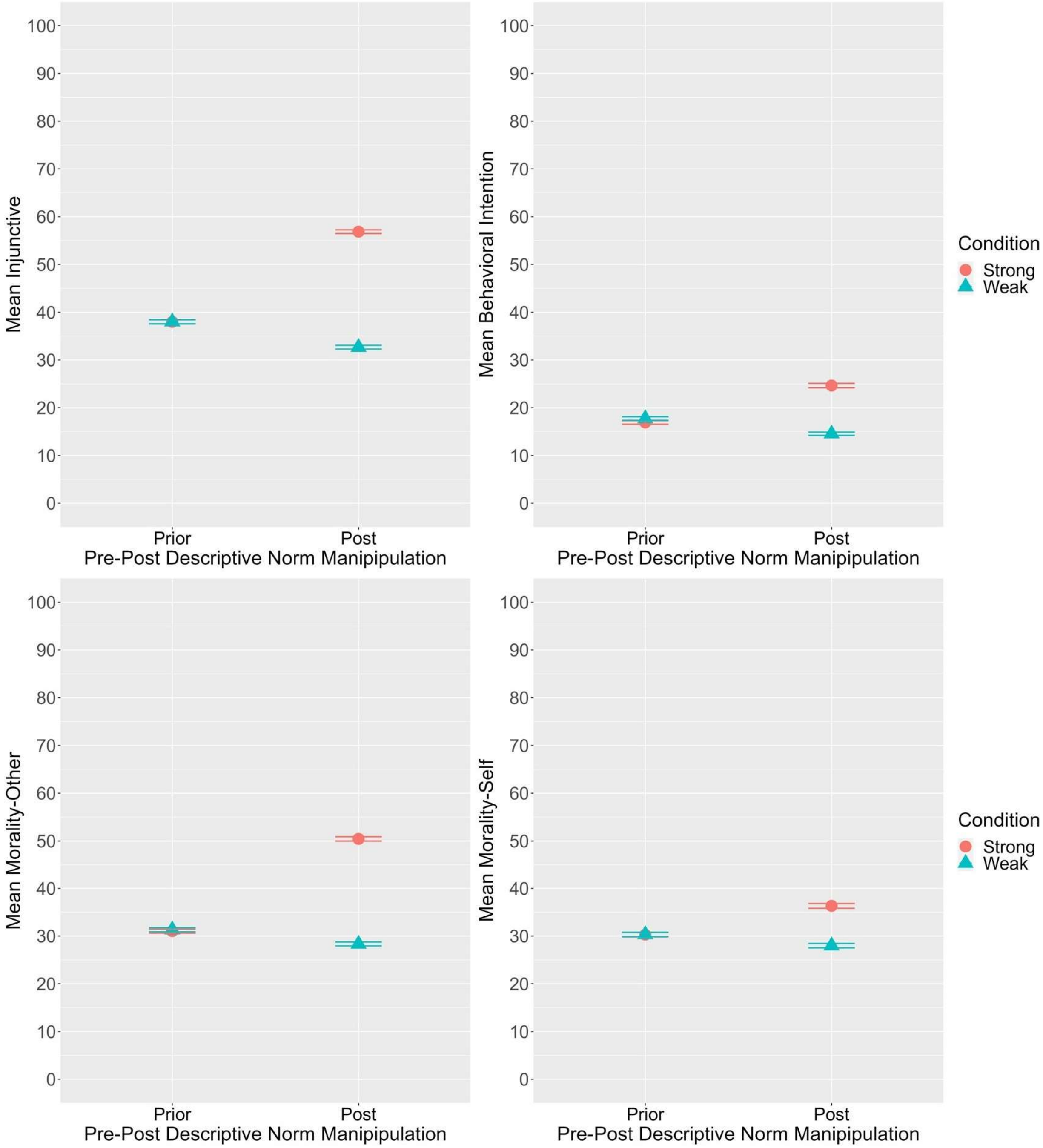


Figure 1. Line plot comparing belief ratings for the weak and strong descriptive norm condition in comparison to prior beliefs for the key dependent measures (injunctive norm beliefs, behavioral intention, self-morality beliefs, other-morality beliefs). This figure

collapses across data from Studies 2 – 5 and only includes the negatively valenced vignette results. Errors bars show standard error.

3.2. Cross-Experiment Results

3.2.1. Analytic Approach

After conducting analyses across all of our individual studies, we found that the effect size for updating varied considerably across studies, perhaps reflecting variability in the samples between studies. In order to evaluate the totality of evidence, we decided to pool our data and run a cross-experiment analysis. While not initially preregistered, taking a meta-analytic approach to analyzing our data offered a more definitive test of whether people update their beliefs after receiving descriptive norms by increasing the number of observations and statistical power.

To that end, we conducted a series of meta-analyses across our five studies that assessed belief updating (Experiments 2-5) following the method described by Harrer et al., (2021) and using the meta package in R (Balduzzi et al., 2019). We conducted a meta-analytic test for each of our main dependent measures—injunctive normativity, self-morality, other-morality, and behavioral intentions. Specifically, we looked at the effect of updating by comparing ratings for the dependent measures prior to receiving the descriptive norm to after receiving the descriptive norm. Thus, for each measure we include two models, one looking at the effect of updating after receiving the strong descriptive norm and another looking at updating after receiving the weak descriptive norm. We also ran a set of meta-analyses for the four key measures comparing updating differences between the weak and the strong descriptive norm conditions.

While our studies had similar demographic information and presumably sampled from the same population (see supplement for demographic breakdown between studies), we observed large between study heterogeneity and so we used random-effect models in our meta-analyses to account for between-study variability in effect sizes. All models used the restricted maximum-likelihood estimator to calculate τ^2 and Knapp-Hartung adjustments to control for between-study heterogeneity. For determining effect sizes, we used Hedges g to correct for small sample bias in calculating the standardized mean difference. We replicate the results of the meta analyses using a series of linear mixed effect models treating experiment identity—along with participant and vignette identity—as random effects. We report the results of these models in the supplement but note that their results are consistent with the results of the meta-analyses reported here.

To determine whether injunctive and moral beliefs mediated the effect of descriptive norm condition on likelihood of engaging in the behavior, we tested for indirect effects using path analysis structural equation models and pooled the data from all studies that measured belief updating. Our models used bootstrapping of 5,000 iterations to find standard errors and bias-corrected bootstrapped confidence intervals. We created three models, the first model was preregistered (Experiments 2 - 5) and included injunctive norm ratings as a mediator while the second and third exploratory models included self-morality and other-morality as mediators, respectively.

3.2.2. Do people update their beliefs and behavior after receiving descriptive norm information?

When looking at the effect of belief updating across studies, we find support for our prediction that people update their injunctive norm beliefs after receiving information about the descriptive norm. There was a significant effect of descriptive norm condition on updating (*d* = 0.91, 95% CI: 0.56, 1.26, *p* = .002), such that people positively updated their injunctive beliefs about how approved the norm is after receiving information that there is a strong descriptive norm. Similarly, we find a significant meta-analytic effect for the weak descriptive norm condition but in the opposite direction: people negatively updated their injunctive beliefs after receiving information that the behavior is uncommon (*d* = 0.16, 95% CI: -0.27, -0.06, *p* = .01).

When looking at the meta-analytic effect of belief updating for behavioral intentions, we find mixed support for our prediction that people update their behavioral intentions after receiving a descriptive norm. Namely, there was not a significant effect of strong descriptive norm condition on behavioral intentions (*d* = 0.22, 95% CI: -0.12, 0.56, *p* = .15), although this effect was in the predicted direction and significant in the common effects model not controlling for random effects of study (*d* = 0.22, 95% CI: 0.18; 0.26, *p* < .001), suggesting there was substantial between study variability in effect size. However, we did find a significant effect of the weak descriptive norm condition on behavioral intentions (*d* = 0.12, 95% CI: -0.15, -0.09, *p* < .001), such that people were significantly less likely to say they would engage in the behavior after finding out it is uncommon.

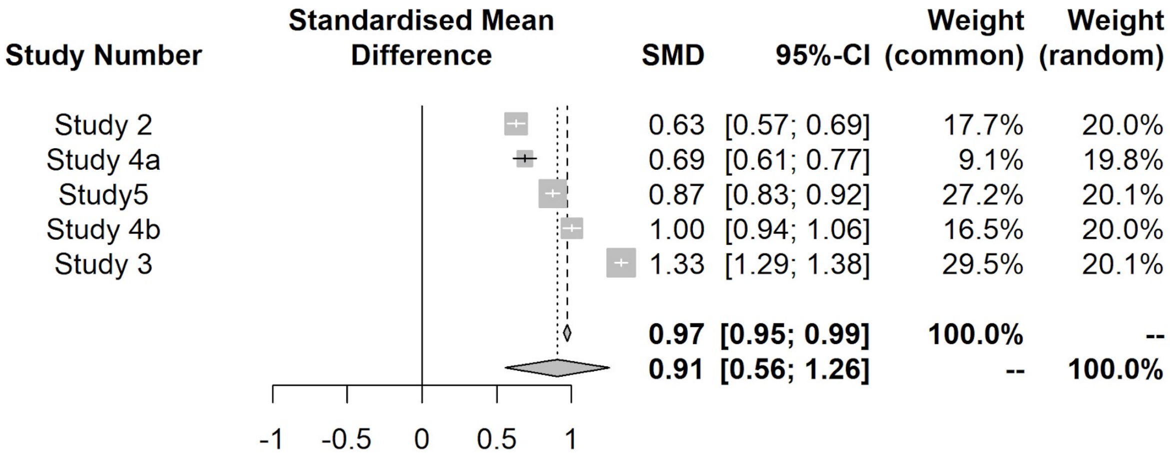
We next examined the meta-analytic effect of belief updating on self-morality—personal beliefs about how moral the behavior is. We find mixed evidence in support of our belief updating prediction: there was no significant effect of the strong descriptive norm condition on self-morality beliefs (*d* = 0.16, 95% CI: -0.24, 0.56, *p* = .33) although this was in the predicted direction. However, we do find a significant, if small, effect of the weak descriptive norm condition on self-morality beliefs (*d* = 0.08, 95% CI: -0.14, -0.01, *p* = .03), such that participants personally thought the behaviors were less moral after finding out that they were uncommon.

Lastly, we turn to the meta-analytic effect of belief updating on other-morality—beliefs about how moral other people think the behavior is. We find some support for this prediction. Namely, there was a significant effect of the strong descriptive norm on other-morality beliefs (*d* = 0.57, 95% CI: 0.27, 0.88, *p* = .006), such that people thought that others would think that the behavior is more moral after finding out it is commonly done. We also found a small negative updating effect of the weak descriptive norm condition on other-morality beliefs that was trending on significance (*d* = 0.11, 95% CI: -0.24, 0.02, *p* = .07), such that people thought that others found the behaviors as marginally less moral after finding out it is uncommon.

3.2.3. Do people update their beliefs and behavior more for strong descriptive norm information than weak descriptive norm information?

When comparing updating between weak and strong descriptive norm conditions across studies, we find evidence that people updated their injunctive norm beliefs. There was a significant effect of descriptive norm condition on injunctive norm updating (*d* = 1.13, 95% CI: 0.89, 1.37, *p* < .001), such that people updated their beliefs more for the strong descriptive norm than the weak descriptive norm. Turning to the meta-analytic effect of belief updating for behavioral intentions, we find a similar pattern: participants updated their behavioral intentions to a larger extent across studies after receiving a strong descriptive norm compared to a weak descriptive norm (*d* = 0.46, 95% CI: 0.12, 0.79, *p* = .019).

Next, when examining the meta-analytic effect of belief updating of personal moral beliefs about the behavior, we find that participants were not significantly more likely to update their personal moral beliefs after receiving the strong than weak descriptive norm (*d* = 0.24, 95% CI: -0.29, 0.77, *p* = .283). Lastly, we find that a significant updating effect across studies for second-order moral judgements about what others find moral. Namely, participants were significantly more likely to update their second-order moral beliefs after receiving a strong descriptive norm than a weak descriptive norm (*d* = 0.85, 95% CI: 0.60, 1.11 , *p* < .001).



Study Number	SMD	95%-CI	Weight (common)	Weight (random)
Study 2	0.63	[0.57; 0.69]	17.7%	20.0%
Study 4a	0.69	[0.61; 0.77]	9.1%	19.8%
Study5	0.87	[0.83; 0.92]	27.2%	20.1%
Study 4b	1.00	[0.94; 1.06]	16.5%	20.0%
Study 3	1.33	[1.29; 1.38]	29.5%	20.1%
	0.97	**[0.95; 0.99]**	**100.0%**	**--**
	0.91	**[0.56; 1.26]**	**--**	**100.0%**

Figure 2. A forest plot of the meta-analytic effects of the strong descriptive norm condition on injunctive norm ratings.

3.2.4. Mediation of descriptive norms on behavioral intentions

3.2.4.1. Injunctive beliefs as mediator

We found that injunctive norm ratings partially mediated the effect of descriptive norm condition on intentions of engaging in the behavior. The total effect of descriptive norm condition on behavioral intentions was significant (b = -10.09, SE = 0.58, p < .001), as was the direct effect of descriptive norm condition on likelihood (b = 6.21, SE = 0.54, p < .001). The path from descriptive norm condition to injunctive norm ratings (b = -31.07, SE = .52, p < .001) was significant, with descriptive norm condition explaining 48.1% of the variance in injunctive ratings (see Table S2 in the Supplement for model output). The path from injunctive ratings to behavioral intentions (b = 0.53, SE = .01, p < .001) was also significant, with injunctive ratings explaining 54.1% of the variance in behavioral intentions. Critically, the indirect effect was significant (b = -16.3, SE = 0.42, p <.001), explaining 26% of the total variance, with the bias-corrected bootstrapped confidence interval with 5,000 samples below zero (95% CI: -17.17, -15.42), suggesting that descriptive norm information influenced behavioral intentions in part by changing beliefs about how injunctive the behavior was.

3.2.4.2. Moral beliefs as mediator

We found that personal moral ratings partially mediated the effect of descriptive norm condition on the likelihood of engaging in the behavior. The direct effect of descriptive norm condition on behavioral intentions was significant (b = -6.17, SE = 0.49, p < .001). The path from descriptive norm condition to self-morality ratings (b = -8.34, SE = .66, p < .001) was significant, with descriptive norm condition explaining 11.6% of the variance in self-morality ratings (see Table S3 in the Supplement for model output). The path from self-morality ratings to behavioral intentions (b = 0.47, SE = .01, p < .001) was also significant, with injunctive ratings explaining 53.9% of the variance in behavioral intention ratings. Importantly, the indirect effect was also significant (b = -3.92, SE = 0.33, p <.001), explaining 6.3% of the total variance with the bias-corrected bootstrapped confidence interval with 5,000 samples below zero (95% CI: -4.57, -3.25), suggesting that the descriptive norms influenced the likelihood of engaging in the behavior by changing beliefs about how moral people found the behavior.

Turning to the other-morality mediation model, we find that beliefs about others' morality also partially mediated the effect of descriptive norm condition on the likelihood of engaging in the behavior. The direct effect of descriptive norm condition on behavioral intentions was statistically significant (b = -1.24, SE = 0.54, p = .02). The path from descriptive norm condition to other-morality ratings (b = -22.1, SE = .44, p < .001) was significant, with descriptive norm condition explaining 41.8% of the variance in other-morality ratings (see Table S4 in the Supplement for model output). The path from other-morality ratings to behavioral intention ratings (b = 0.40, SE = .01, p < .001) was also significant, with injunctive ratings explaining 42.9% of the variance in intention ratings. Critically, the indirect effect was also significant (b = -8.85, SE = 0.29, p <.001), explaining 17.9% of the total variance with the bias-corrected bootstrapped confidence interval with 5,000 samples below zero (95% CI: -9.49, -8.21), suggesting that the descriptive norms influenced behavioral intentions in part by changing participants' beliefs about how moral others found the behavior.

4. General Discussion

Across six studies, we explored the relationship between descriptive norms, injunctive norms, moral judgements, and behavioral intentions. Specifically, we assessed participants' beliefs about normative behaviors both before and after receiving information that there was either a strong or weak descriptive norm. Across all studies, we found that people updated their beliefs about the injunctive normativity of a behavior after receiving descriptive norm information. When there was a strong descriptive norm that the behavior is common, participants thought the behavior was more approved of than before receiving the norm information. When receiving a weak descriptive norm that the behavior is uncommon, participants thought the behavior was less approved of then they initially thought. We found a similar, albeit weaker, pattern for behavioral intentions–participants rated themselves as more likely to engage in the behavior after receiving a strong descriptive norm, and less likely after receiving the weak descriptive norm.

We found somewhat conflicting updating results for the strong descriptive norm on behavioral intentions: while participants updated their behavioral intentions in response to the strong descriptive norm in most individual studies, the meta-analytic effect controlling for the random effect of study was non-significant, suggesting that that effect size varied substantially between studies. We also found somewhat mixed evidence that participants updated their personal moral beliefs—they did not significantly update their beliefs in response to the strong descriptive norm but did so in response to the weak descriptive norm, although this effect was small. In contrast, we found stronger evidence that participants updated their second-order moral beliefs after receiving descriptive norms information. Thus, participants' second-order more beliefs were more sensitive to descriptive norms than their first-order, personal moral beliefs.

For all dependent measures, we found that the effect size of the strong descriptive norm was larger than for the weak descriptive norm. In other words, people positively updated their beliefs to a greater extent for strong descriptive norms than they negatively updated their beliefs for weak descriptive norms. All together, our results support previous work documenting a strong association between descriptive norms, injunctive norms, and moral judgements (Eriksson et al., 2015; Lindstrom et al., 2018).

4.1. Do people update their beliefs and behavior after receiving descriptive norm information?

One of the central goals of this research was to investigate how, and to what extent, people update their injunctive norm beliefs and moral judgements after receiving descriptive norm information. While previous work has found that people make simple, bidirectional inferences between descriptive and injunctive norms (e.g., is a behavior injunctive or not given that it is common or uncommon; Eriksson et al., 2015), it was unclear how the strength of descriptive norm information (e.g., how common a behavior is in terms of the proportion of those engaging in it) influences our beliefs about what others approve of and find moral. Through manipulating the strength of the descriptive norm, we were better able to understand the extent to which descriptive norms influence beliefs and behavior. For example, we found that people *negatively* updated their beliefs after receiving a weak descriptive norm that the behavior was relatively uncommon while they tended to positively update their beliefs after receiving a strong descriptive norm that the behavior was common. Furthermore, our study builds on and extends this work by assessing beliefs before and after receiving descriptive norm information, demonstrating how individuals update their own beliefs about the injunctive normativity and morality of a behavior in the face of new descriptive norm information. By examining descriptive beliefs, injunctive beliefs, moral judgements, and behavioral intentions within one design, our results offer the clearest evidence to date of how descriptive norms shape injunctive norm beliefs, and informs our understanding of the relationship between descriptive and injunctive norms, moral judgements, and behavior.

More generally, these results highlight the important role of mentalizing and belief representation in norm cognition—when we see a lot of people doing something and infer that most people must also approve of it, we are relying on our ability to represent others' minds to infer things about their beliefs based on their behavior. Importantly, as the results of our mediation analyses demonstrate, our inferences about what others approve of and find moral based on how they are behaving influences our own intentions of engaging in the behavior. Representing others' beliefs and behavior is a core aspect of the socially conditional account of norms that has been proposed by Bicchieri and colleagues (2005; 2016). Namely, this account proposes that a fundamental feature of social norms is that they depend on our beliefs about what others do and expect. That we find a strong relationship between descriptive norm information and beliefs and behavior here supports this theory of norms and points to the importance of mentalizing in social norm cognition.

4.2. Relationship between injunctive norms and moral judgements

A secondary goal of this project was to better understand the relationship between descriptive norms, injunctive norms, and moral judgements—how exactly do these concepts relate to one another? Are they tapping related but dissociable concepts? While generally thought of as distinct constructs, previous research has documented a strong relationship between descriptive and injunctive norms and moral judgements, with some work even defining injunctive norms in terms of morality (Goldring & Heiphetz, 2020; Lindstrom et al., 2018; Russell et al., 2021; Smith & Masser, 2012; Zhang et al., 2022). Across studies we found a moderately strong association between injunctive norms and moral judgements, indicating they are measuring related but distinct beliefs. This finding highlights the importance of studying both injunctive norm beliefs and moral judgements and suggests that researchers should be careful to avoid conflating injunctive and moral beliefs.

Our results also reveal a key dissociation between our personal moral beliefs and second-order moral beliefs. Namely, despite being strongly correlated with one another ($r = 0.84$) participants' second-order moral beliefs—beliefs about what others think is moral—was more readily influenced by descriptive norm information than their personal moral beliefs about the behavior. Furthermore, beliefs about others' morality had a stronger association to injunctive norm beliefs ($r = 0.69$) than personal moral beliefs ($r = 0.56$), suggesting that, while likely still distinct, injunctive beliefs are perhaps more closely aligned with second-order moral beliefs than first-order moral beliefs. More generally, this result highlights an interesting paradox—people think that other people's moral judgements are easily swayed by descriptive norm information when in reality, most people's personal moral judgements were largely resilient to descriptive norms, only updating to a small extent if at all. In other words, this finding suggests that people think that other people’s beliefs are more flexible than they really are. This dissociation demonstrates the value of assessing both first-order and second-order moral beliefs in work on social norms and moral psychology.

While we were primarily interested in the relationship between normative beliefs and moral judgements, some researchers propose that there is a distinct category of moral norms. This work defines moral norms as internalized preferences—expectations we hold for ourselves regardless of others’ beliefs or behavior—that are insensitive to the social expectations that are inherent to injunctive norms (Bicchieri, 2006; House, 2018). Relatedly, others view moral norms as personal injunctive norms (or just personal norms; Morris et al., 2015), defined as an “individual’s internalized moral rules” that we follow independent of others’ expectations and influence (Parker et al., 1995; White et al., 2009). Injunctive norms differ from moral norms in that they are not internalized, meaning that they are socially conditional and thus influenced to a greater extent by others’ behavior and expectations. This notion is supported by our finding that injunctive beliefs and moral judgements dissociate: participants in our study consistently and robustly updated their injunctive norm beliefs after receiving descriptive norm information while on the whole, they largely did not update their personal moral beliefs about the behaviors. That moral judgements were less influenced by descriptive norm information compared to injunctive beliefs indicates that injunctive norms are different from moral norms because, unlike injunctive norms, moral norms are internalized and not socially conditional.

While not the main focus of this paper, we also explored whether injunctive norm beliefs and moral judgements mediate the effect of descriptive norms on behavioral intentions. Across all studies, including the combined analyses, we found that injunctive beliefs partially mediated the effect of descriptive norm information on reported likelihood of engaging in the behavior. That is, receiving information about how common a behavior is makes people more likely to engage in it, partly because it influences their beliefs about how much other people approve of or condone the behavior. We found a similar result for the mediation models including personal morality and other morality—both personal beliefs about the morality of a behavior and second-order moral beliefs partially mediated the effect of descriptive norm information on behavioral intentions. Together, these models suggest that descriptive norms influence behavioral intentions in part because they influence our beliefs about what others approve of and find moral. Future work should further investigate the relationship between descriptive norms, injunctive norms, moral judgements, and behavior using experimental mediation to make stronger claims about causation.

4.3. Limitations

The present work was not without its limitations. First, because we studied existing behaviors rather than entirely novel ones, we could not prevent participants from bringing in their own priors about the morality and normativity of the behaviors we studied. In other words, people might have had existing beliefs about how common or approved of a given behavior is based on their personal experiences which could have influenced their behavior in our task. For example, participants' personal moral beliefs might have been largely unaffected by descriptive norm information because they had strong prior beliefs acquired from their lives about the morality of cheating on a test or cutting a line. In order to avoid people's prior experiences influencing their beliefs and behavior, future work should explore injunctive belief updating in the context of totally novel behaviors that people do not have existent priors for. If we find that people continue to update their injunctive beliefs and behavior after receiving descriptive norm information for norms they have no prior experience with or beliefs about, that would provide strong evidence that descriptive norms influence injunctive norm beliefs and behavior.

Second, while we measured behavioral intention ("how likely are you to X") as a proxy for behavior, there are likely differences between what people say they would do and what they would actually do (Blake et al., 2014; Dang et al., 2020). So, while we found that people updated their intentions of engaging in the behavior after receiving the descriptive norm, it is possible this effect would not replicate for actual behavior. That said, work on the relationship between behavioral intentions and behavior finds that intentions have a significant effect on behavior, but the size of the effect is smaller for actual behavior than intentions (Webb & Sheeran, 2006). This might suggest that, while people's actual behavior is influenced by descriptive norms, it may be influenced to a lesser extent than their behavioral intentions. Future work should use behavioral experiments to investigate whether people actually change their behavior—rather than just their self-reported intentions—after receiving descriptive norm information.

Third, because we solely focused on the effect of descriptive norms on injunctive norm beliefs, we cannot necessarily make claims about effects in the opposite direction from injunctive to descriptive (e.g., that injunctive norm information influences people's descriptive norm beliefs). However, given the strong relationship between descriptive and injunctive norms, and previous work showing people make bidirectional inferences between them (Eriksson et al., 2015), we expect that people would be as likely to update their descriptive beliefs from injunctive norm information. Future work should explore whether and to what extent people update their descriptive norm beliefs, moral judgements, and behavior after receiving injunctive norm information.

4.4. Conclusion

Across six experiments, we found that people update their injunctive norm beliefs after receiving descriptive norm information that the behavior was either common or uncommon. Additionally, we found that the effect of descriptive norms on behavior was partially mediated by injunctive beliefs, suggesting that descriptive norms influence our behavior partly by changing our beliefs about what others think is acceptable. These results inform our understanding of the relationship between descriptive norms, injunctive norms, moral judgements, and behavior, shining light on how we form normative beliefs from a common source of social information—descriptive norms. More generally, our findings highlight the important role our beliefs about what others believe and do play in social norm cognition. Given the importance and prevalence of norms in our social world, a better understanding of norm cognition can reveal important insights into social cognition.

5. Open Data Practices

All of the experiments in this project were preregistered prior to data collection [Experiment 1: https://aspredicted.org/blind.php?x=9zq26n; Experiment 2: https://osf.io/e3c6n/?view_only=8d06a95126b64edda82f978b6e858f8a; Experiment 3: https://osf.io/gua3q/?view_only=873ce428e7534e74838fe22c09de90f0; Experiment 4a: https://osf.io/c6r8p/?view_only=3d9c8bd3651445fca1ad332a50c67d30; Experiment 4b: https://osf.io/5pz8m/?view_only=bc9df3f77c5f4fa6bf314629f9bbc0ae; Experiment 5: https://osf.io/pdnyx/?view_only=839ed94ca2854895b05cbad92a1ba043] All of our data and code are available in an online repository here [https://osf.io/sc842/?view_only=83662b719da3479f8241eefce2c4ade4].

6. References

Arechar, A. A., Kraft-Todd, G. T., & Rand, D. G. (2017). Turking overtime: How participant characteristics and behavior vary over time and day on Amazon Mechanical Turk. *Journal of the Economic Science Association*, *3*(1), 1-11.

Axelrod, R. (1986). An evolutionary approach to norms. *American Political Science Review*, *80*(4), 1095-1111.

Balduzzi, S., Rücker, G., & Schwarzer, G. (2019). How to perform a meta-analysis with R: a practical tutorial. *Evidence-based Mental Health*, *22*(4), 153-160.

Baumard, N., André, J. B., & Sperber, D. (2013). A mutualistic approach to morality: The evolution of fairness by partner choice. *Behavioral and Brain Sciences*, *36*(1), 59-78.

Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). Parsimonious mixed models. *arXiv preprint arXiv:1506.04967*.

Bear, A., & Knobe, J. (2017). Normality: Part descriptive, part prescriptive. *Cognition*, *167*, 25-37.

Blake, P. R., McAuliffe, K., & Warneken, F. (2014). The developmental origins of fairness: The knowledge–behavior gap. *Trends in Cognitive Sciences*, *18*(11), 559-561.

Blanton, H., Köblitz, A., & McCaul, K. D. (2008). Misperceptions about norm misperceptions: Descriptive, injunctive, and affective 'social norming' efforts to change health behaviors. *Social and Personality Psychology Compass*, *2*(3), 1379-1399.

Bicchieri, C. (2005). *The grammar of society: The nature and dynamics of social norms*. Cambridge University Press.

Bicchieri, C. (2016). *Norms in the wild: How to diagnose, measure, and change social norms*. Oxford University Press.

Buffalo, M., & Rodgers, J. W. (1971). Behavioral norms, moral norms, and attachment: Problems of deviance and conformity. *Social Problems*, *19*(1), 101-113.

Chakroff, A., Dungan, J., Koster-Hale, J., Brown, A., Saxe, R., & Young, L. (2016). When minds matter for moral judgment: intent information is neurally encoded for harmful but not impure acts. *Social Cognitive and Affective Neuroscience*, *11*(3), 476 484.

Chakroff, A., & Young, L. (2015). Harmful situations, impure people: An attribution asymmetry across moral domains. *Cognition*, *136*, 30-37.

Cialdini, R. B., Reno, R. R., & Kallgren, C. A. (1990). A focus theory of normative conduct: Recycling the concept of norms to reduce littering in public places. *Journal of Personality and Social Psychology*, *58*(6), 1015.

Cialdini, R. B., Kallgren, C. A., & Reno, R. R. (1991). A focus theory of normative conduct: A theoretical refinement and reevaluation of the role of norms in human behavior. In *Advances in Experimental Social Psychology* (Vol. 24, pp. 201-234). Academic Press.

Cialdini, R. B., Demaine, L. J., Sagarin, B. J., Barrett, D. W., Rhoads, K., & Winter, P. L. (2006). Managing social norms for persuasive impact. *Social Influence*, *1*(1), 3-15.

Curry, O. S., Mullins, D. A., & Whitehouse, H. (2019). Is it good to cooperate? Testing the theory of morality-as-cooperation in 60 societies. *Current Anthropology*, *60*(1), 47-69.

Dang, J., King, K. M., & Inzlicht, M. (2020). Why are self-report and behavioral measures weakly correlated? *Trends in Cognitive Sciences*, *24*(4), 267-269.

Deutchman, P., Marshall, J., Lee, Y., Sansom, E., Warneken, F., & McAuliffe, K. (Preprint). Children selectively update their injunctive norm beliefs and moral evaluations after receiving descriptive norm information.

Deutchman, P., Amir, D., Jordan, M. R., & McAuliffe, K. (2022). Common knowledge promotes cooperation in the threshold public goods game by reducing uncertainty. *Evolution and Human Behavior*, *43*(2), 155-167.

Dungan, J. A., Chakroff, A., & Young, L. (2017). The relevance of moral norms in distinctrelational contexts: Purity versus harm norms regulate self-directed actions. *PloS One*, *12*(3), e0173405.

Elek, E., Miller-Day, M., & Hecht, M. L. (2006). Influences of personal, injunctive, and descriptive norms on early adolescent substance use. *Journal of Drug Issues*, *36*(1), 147-172.

Eriksson, K., Strimling, P., & Coultas, J. C. (2015). Bidirectional associations between descriptive and injunctive norms. *Organizational Behavior and Human Decision Processes*, *129*, 59-69.

Folger, R. (1998). Fairness as moral virtue. In *Managerial ethics* (pp. 23-44). Psychology Press.

Goldring, M. R., & Heiphetz, L. (2020). Sensitivity to ingroup and outgroup norms in the association between commonality and morality. *Journal of Experimental Social Psychology*, *91*, 104025.

Harrer, M., Cuijpers, P., Furukawa, T. A., & Ebert, D. D. (2021). *Doing meta-analysis with R: A hands-on guide*. Chapman and Hall/CRC.

House, B. R. (2018). How do social norms influence prosocial development?. *Current Opinion in Psychology*, *20*, 87-91.

Lindström, B., Jangard, S., Selbing, I., & Olsson, A. (2018). The role of a "common is moral" heuristic in the stability and change of moral norms. *Journal of Experimental Psychology: General*, *147*(2), 228.

Lu, H., Zou, J., Chen, H., & Long, R. (2020). Promotion or inhibition? Moral norms, anticipated emotion and employee's pro-environmental behavior. *Journal of Cleaner Production*, *258*, 120858.

Morris, M. W., Hong, Y. Y., Chiu, C. Y., & Liu, Z. (2015). Normology: Integrating insights about social norms to understand cultural dynamics. *Organizational Behavior and Human Decision Processes*, *129*, 1-13.

Parker, D., Manstead, A. S., & Stradling, S. G. (1995). Extending the theory of planned behaviour: The role of personal norm. *British Journal of Social Psychology*, *34*(2), 127-138.

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Raihani, N. J., & McAuliffe, K. (2014). Dictator game giving: The importance of descriptive versus injunctive norms. *PloS One*, *9*(12), e113826.

Reno, R. R., Cialdini, R. B., & Kallgren, C. A. (1993). The transsituational influence of social norms. *Journal of Personality and Social Psychology*, *64*(1), 104.

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.

Russell, P. S., Smith, D. M., Birtel, M. D., Hart, K. H., & Golding, S. E. (2022). The role of emotions and injunctive norms in breastfeeding: a systematic review and meta analysis. *Health Psychology Review*, *16*(2), 257-279.

Smetana, J. G. (2013). Moral development: The social domain theory view. In P. D. Zelazo (Ed.), *The Oxford Handbook of Developmental Psychology (Vol. 1): Body and mind* (pp. 832–863). Oxford University Press.

Smetana, J. G., Jambon, M., & Ball, C. (2014). The social domain approach to children's moral and social judgments. In M. Killen & J. G. Smetana (Eds.), Handbook of moral development (2nd ed., pp. 23–45). Taylor & Francis Publishers.

Smetana, J. G., Ball, C. L., Jambon, M., & Yoo, H. N. (2018). Are young children's preferences and evaluations of moral and conventional transgressors associated with domain distinctions in judgments? *Journal of Experimental Child Psychology, 173*, 284–303.

Smith, M. K., & Masser, B. M. (2012). Principles and popularity: The interplay of moral norms and descriptive norms in the context of volunteerism. *British Journal of Social Psychology*, *51*(4), 762-771.

Turiel, E., Smetana, J. G., & Killen, M. (2014). Social contexts in social cognitive development. In *Handbook of Moral Behavior and Development* (pp. 329-354). Psychology Press.

Turiel, E., & Dahl, A. (2019). The development of domains of moral and conventional norms, coordination in decision-making, and the implications of social opposition. *The normative animal: On the anthropological significance of social, moral, and linguistic norms*, 195-213.

Tworek, C. M., & Cimpian, A. (2016). Why do people tend to infer "ought" from "is"? The role of biases in explanation. *Psychological Science*, *27*(8), 1109-1122.

Webb, T. L., & Sheeran, P. (2006). Does changing behavioral intentions engender behavior change? A meta-analysis of the experimental evidence. *Psychological Bulletin*, *132*(2), 249.

White, K. M., Smith, J. R., Terry, D. J., Greenslade, J. H., & McKimmie, B. M. (2009). Social influence in the theory of planned behaviour: The role of descriptive, injunctive, and in group norms. *British Journal of Social Psychology*, *48*(1), 135-158.

Yucel, M., Drell, M. B., Jaswal, V. K., & Vaish, A. (2022). Young children do not perceive distributional fairness as a moral norm. *Developmental Psychology*, *58*(6), 1103.

Zhang, J., Cherian, J., Abbas Sandhu, Y., Abbas, J., Cismas, L. M., Negrut, C. V., & Negrut, L. (2022). Presumption of Green Electronic Appliances Purchase Intention: The Mediating Role of Personal Moral Norms. *Sustainability*, *14*(8), 4572.

7. Appendix

7.1. Vignette Type Updating Comparison

7.1.1. Introduction

Another important yet unexplored question is whether the relationship between descriptive norms, injunctive norms, and moral judgements is consistent across different categories of norms or whether it varies such that the relationship is stronger for certain kinds of behaviors. Put differently, we might update our injunctive beliefs and moral judgments about a norm to a greater or lesser extent depending on the norm in question. In support of this possibility, a growing body of work suggests there are important differences in cognition between moral domains. For example, there are a number of differences in moral cognition between harm and purity domains—researchers have found an attribution asymmetry between purity and harm domains such that people endorse more person-based attributions for impure acts compared to harmful ones (Chakroff & Young, 2015), while other work finds that people rely more on judgements of intent for harm violations than purity violations (Chakroff et al., 2015). Additionally, there is also a large body of work on social domain theory which posits that children hold a categorical distinction between moral norms and social conventions. Namely, this theory predicts that moral concepts and norms are universally applied and obligatory while social conventions and norms are perceived as more alterable and subjective (Smetna, 2013; Smetna et al., 2014). This work has found that children behave differently depending on whether a norm is conventional (e.g., wearing a school uniform) or moral (e.g., bullying and stealing someone's lunch money), such that they view conventional norm violations as less serious and deserving of punishment (Smetna et al., 2014). This tendency to view moral norms and social conventions as distinct arises early in development: by 3-4-years of age, children begin to differentiate behaviors that are conventionally normative from those that are morally normative (Smetana, 2013; Smetana et al., 2018; Turiel & Dahl, 2019).

Altogether, this previous work provides compelling evidence that we view moral and conventional norms as psychologically distinct. Given this prior work, we wanted to explore whether people would update their beliefs to different extents for conventional and moral norms. Namely, we asked whether people would be more sensitive to descriptive norm information for conventional norms compared to moral norms such as harm and fairness norms. Because conventional norms are socially agreed upon whereas moral norms are perceived as inalterable and objective (Smetna, 2006), we hypothesized that people should be less influenced by descriptive norm information for moral norms than conventional norms.

Furthermore, while there is increasing evidence that moral norms and social conventions are distinct, less work has explored the distinction between different types of moral norms, such as between fairness (e.g., cheating in a competition) and harm-related (e.g., mugging someone) norms. While past work demonstrates that there are important differences in cognition between two types of norms–harm and purity norms (Chakroff et al., 2015; Chakroff & Young, 2015)–there are several other kinds of moral norms that we frequently encounter and which we might also view as psychologically distinct. For example, recent work has identified an important distinction between two kinds of moral norms–fairness and harm norms. Namely, while fairness is often viewed as a moral behavior (Baumard et al., 2013; Curry et al., 2019; Folger, 1998), this work suggests that children view fairness norms differently than harm-based norms, such that they view harm-based norm violations as more serious than fairness-based violations (Yucel et al., 2022). Here we focus on the question of whether people view fairness and harm behaviors as differently. Specifically, we asked whether people would be less likely to update their injunctive norm beliefs and moral judgements after receiving descriptive norm information for harm behaviors than fairness behaviors.

By comparing updating between types of norms, we can shine light on whether people view them as psychologically distinct. If, for example, we find that people update their beliefs to the same extent across moral and conventional behaviors, this might suggest that people actually view these different behaviors more similarly than what is predicted by social domain theorists who posit a fundamental cognitive difference between these behaviors (Smetna, 2006; 2013). Additionally, by looking at updating across different kinds of moral behaviors (e.g., fairness vs. harm), we gain insight into whether people view different kinds of moral behavior as distinct concepts. If we find that descriptive norm information influences beliefs to a greater extent for fairness than harm behaviors, this would suggest there are important differences in how we conceptualize these kinds of moral behaviors that could have broad implications for the study of moral psychology. To that end, we focused on three kinds of norms here—conventional, fairness, and harm norms.

7.1.2. Analysis

Our modeling approach here was identical to that described previously. We ran a series of preregistered linear mixed effect models including participant and vignette identity as random effects and dependent measure difference scores as the outcome measures. All models included the interaction term between the descriptive norm condition and vignette type. The fairness vignettes were set as the reference level when making between-vignette comparisons. See our preregistrations for experiment-specific analyses.

7.1.3. Results

We find relatively mixed evidence for differences in updating between norm types. We first examined whether participants updated their prior beliefs to different extents after receiving a strong descriptive norm depending on the category of behavior. In two out of four studies, participants updated their injunctive beliefs more after receiving strong descriptive norm information for the fairness behaviors than the harm behaviors, however, this finding varied depending on the wording of the injunctive measure (see Measures). In Experiment 5, participants updated their injunctive beliefs to a greater extent for the conventional behaviors than the harm behaviors and preferences; there were no differences in updating between the fairness and conventional vignettes. In two out four studies, participants updated their behavioral intentions to engage in the behavior to a greater extent after receiving the strong descriptive norm for the fairness behaviors as compared to the harm behaviors. Experiment 5 found the same interaction effect but between conventional and harm behaviors: participants more readily updated their behavioral intentions for the conventional behaviors than the harm behaviors. In two out of four studies, participants updated their personal moral beliefs more for fairness than harm behaviors. Participants updated their personal moral beliefs more for conventional than fairness behaviors in Experiment 5, while the difference was not significant in Experiment 2. In Experiment 5, participants also updated their personal moral beliefs more for conventional behaviors than harm behaviors and preferences. In three out of four studies, participants updated their beliefs about others' moral beliefs to a greater extent for fairness than harm vignettes. In Experiment 2, participants did not update to different extents between conventional and fairness behaviors while in Experiment 5, they updated their beliefs more for conventional than fairness behaviors and updated to similar extents for fairness behaviors and preferences.

Next, turning to the weak descriptive norm updating results, we find that after receiving a weak descriptive norm, participants updated their injunctive norm beliefs more for fairness than harm behaviors in four out of four studies. Participants did not update to different extents between the fairness behaviors and preferences while in Experiments 2 and 5, participants updated more for conventional than fairness behaviors. In three out of four studies, participants updated their behavioral intentions of engaging in the behavior after receiving the weak descriptive norm to a greater extent for fairness than harm behaviors. There was no difference in updating between fairness behaviors and preferences across all studies. Participants updated their behavioral intentions to similar extents for fairness and conventional behaviors in Experiments 2 and 5. In only one out of four studies did participants update their personal moral beliefs to a larger extent for fairness than harm behaviors after receiving a weak descriptive norm. There was no difference in updating between fairness behaviors and preferences, nor between fairness and conventional behaviors. Lastly, in only one out of four studies did participants update their beliefs about others' moral beliefs more for fairness than harm behaviors after receiving the weak descriptive norm. Participants updated their beliefs about others' moral beliefs more for fairness behaviors than preferences in three out of four studies while they updated less for fairness than conventional behaviors in Experiments 2 and 5.

When comparing updating between the weak and strong descriptive norm conditions across vignette type, we find that, in three out of four studies, the updating difference in participants' injunctive beliefs between the weak and strong descriptive norm conditions was greater for the fairness vignettes than the harm vignettes. However, this finding varied depending on the wording of the injunctive measure (see Measures). In Experiment 5, participants' weak-strong descriptive norm updating difference for injunctive ratings was greater for the conventional behaviors than the harm behaviors and preferences. In four out four studies, the difference in behavioral intention updating between weak & strong descriptive norm conditions was larger for the fairness behaviors than harm behaviors. Experiment 5 found the same interaction effect but between the conventional behaviors and the harm behaviors and preferences such that the difference in updating between descriptive norm conditions was larger for conventional behaviors than harms or preferences. In three out of four studies, the weak-strong descriptive norm updating difference in personal moral beliefs was greater for fairness than harm behaviors. In Experiments 2 and 5, there was no difference in weak-strong descriptive norm updating for self-morality between conventional and fairness behaviors whereas the difference was larger for conventional behaviors than harm behaviors and preferences. The weak-strong descriptive norm updating difference for others' moral beliefs was greater for fairness than harm vignettes in two out of four studies. In Studies 2 and 5, there was no difference in the extent of weak-strong descriptive norm updating between conventional and fairness behaviors while in Experiment 5, the updating difference was larger for conventional behaviors than harm behaviors and preferences.

Interestingly, we found different results depending on the wording of the injunctive measure—participants consistently updated their injunctive beliefs more for fairness behaviors than harm behaviors when phrased in terms of others' approval whereas this effect was not significantly different when the injunctive norm was phrased in terms of what others think you should do. We also found different results when including the new harm vignettes in Experiment 5. Namely, while we found differences in injunctive belief updating between fairness behaviors and the initial set of harm behaviors, we failed to find a difference in updating between fairness behaviors and the new harm behaviors—while in the predicted direction, participants were not significantly more sensitive to descriptive norm information for fairness than these harm behaviors. However, we did replicate the difference in injunctive belief updating between the conventional behaviors and harm behaviors for both sets of harm vignettes. Importantly, this difference between the two sets of harm vignettes was specific to injunctive updating—we replicate the updating differences between norm categories for the likelihood and self- and other-morality updating with both sets of harm vignettes.

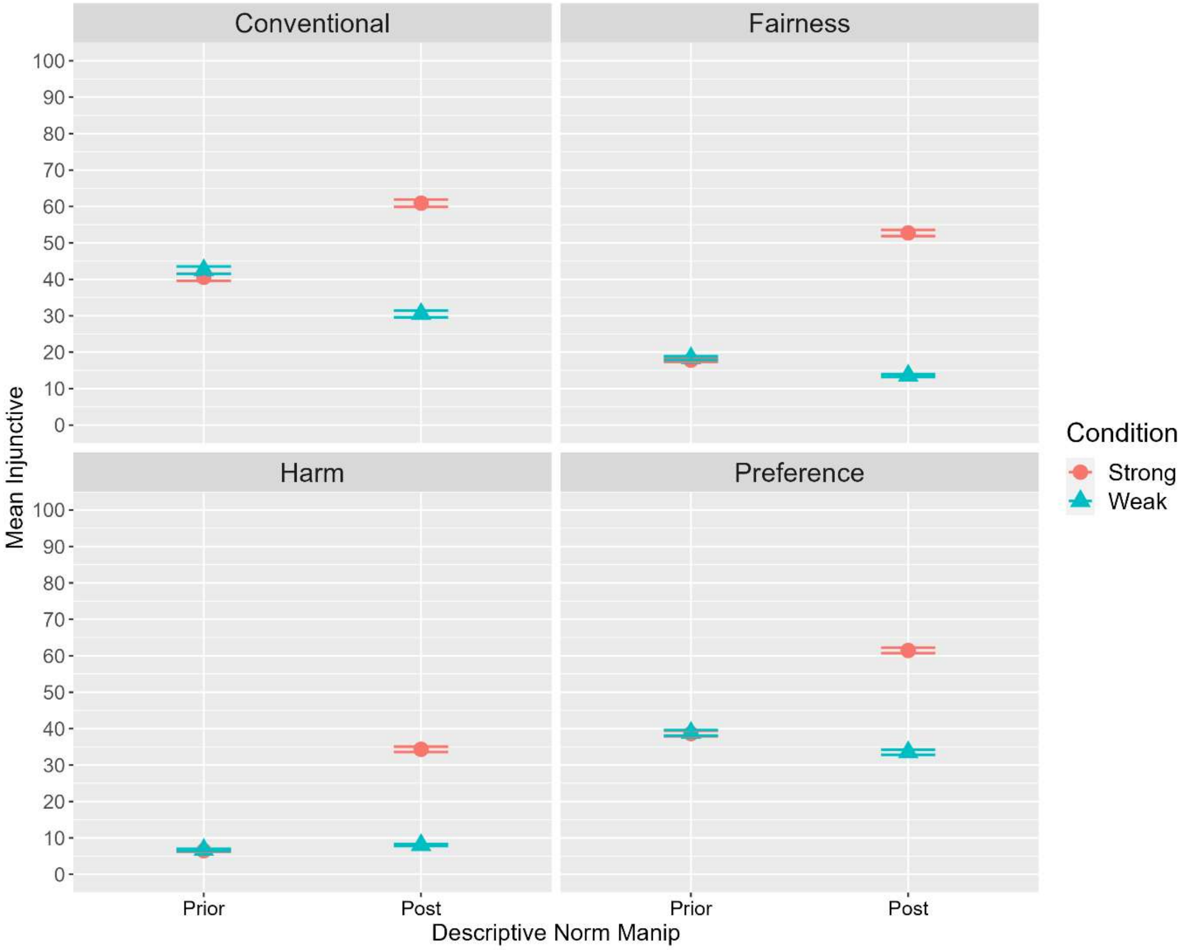


Figure 3. Plot comparing belief updating for the injunctive norm measure across all four norm types studied. Errors bars show standard error.

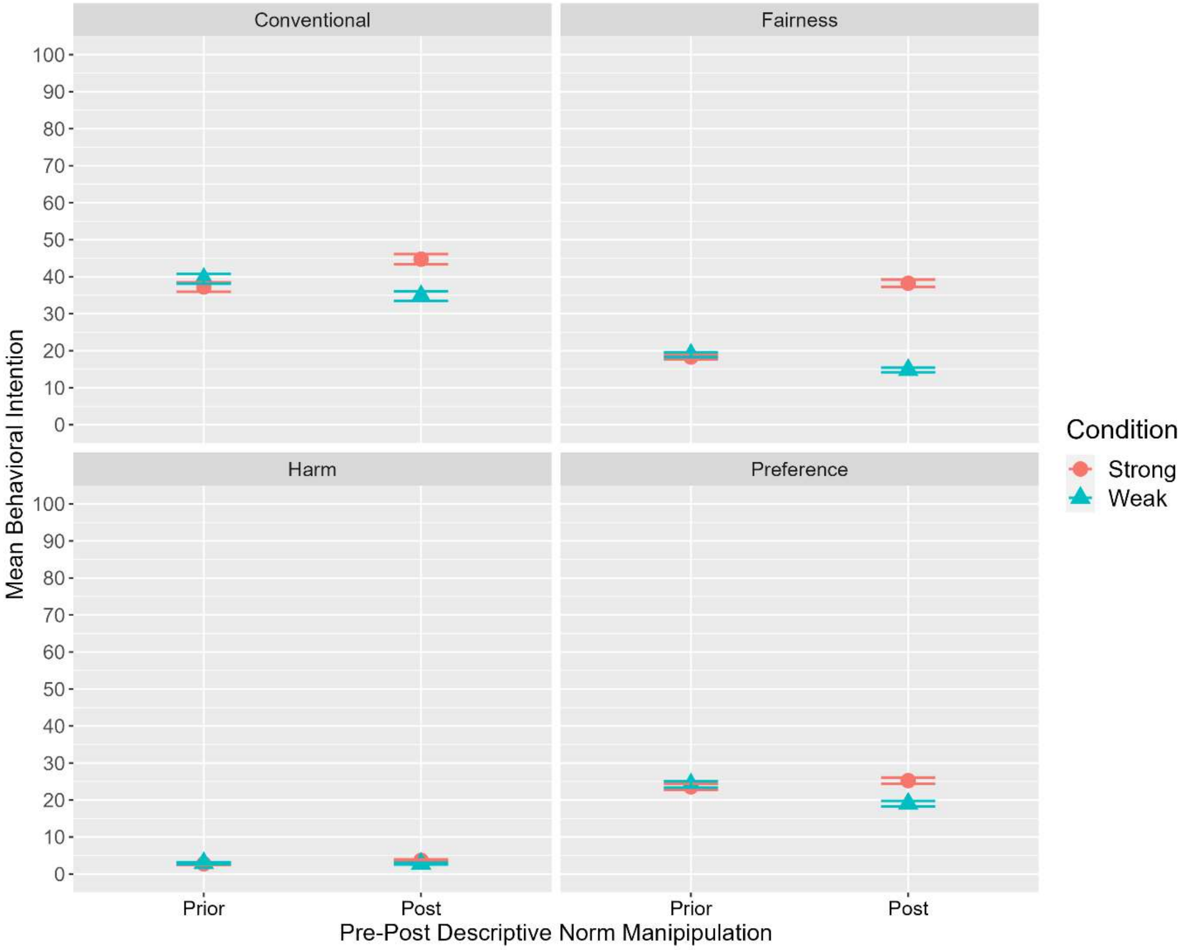


Figure 4. Plot comparing belief updating for the behavioral intention measure across all four norm types studied. Errors bars show standard error.

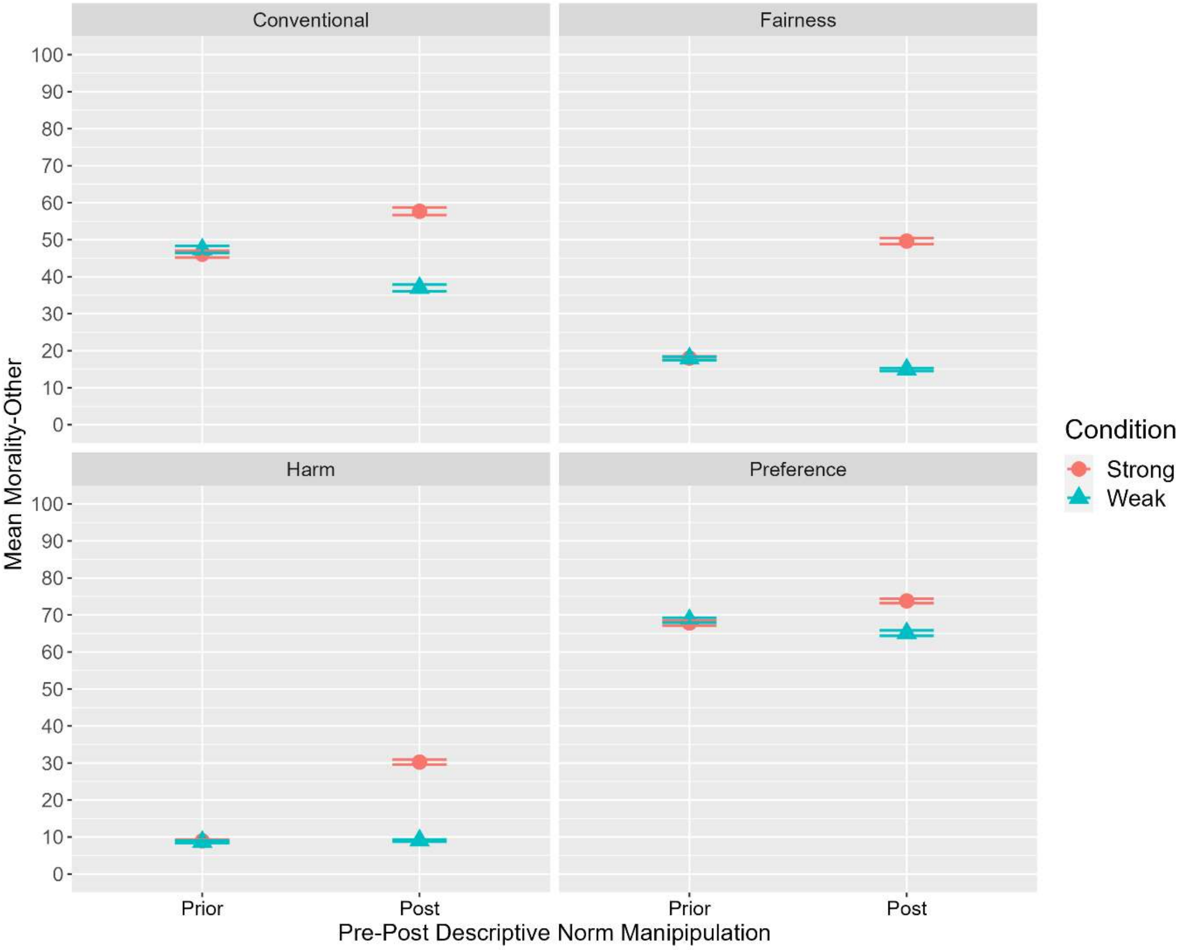


Figure 5. Plot comparing belief updating for the other-morality measure across all four norm types studied. Errors bars show standard error.

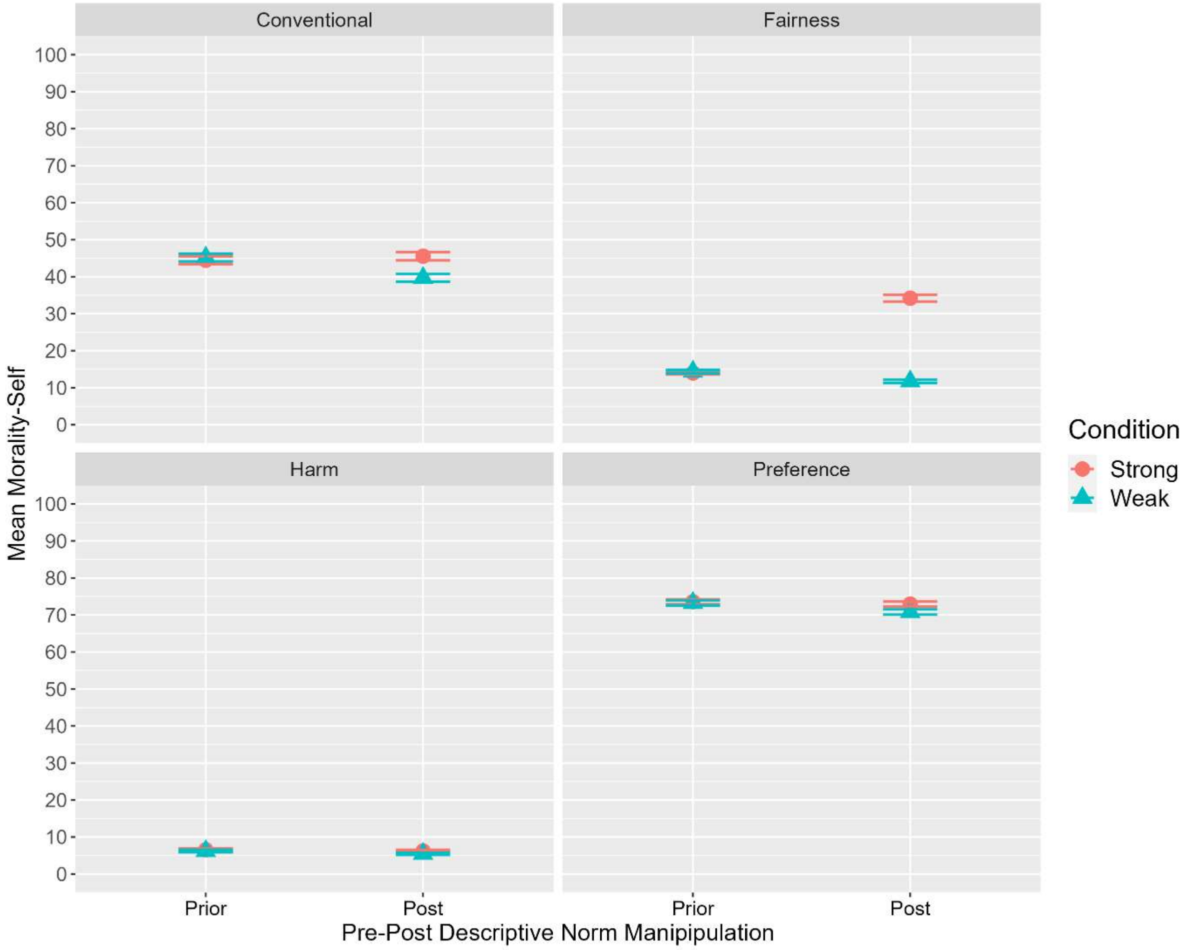


Figure 6. Plot comparing belief updating for the self-morality measure across all four norm types studied. Errors bars show standard error.

7.1.4. Discussion

We examined whether the relationship between descriptive and injunctive norm beliefs varies depending on the type of behavior in question. In other words, do we as readily update our beliefs and behavioral intentions for some normative behaviors—such as fairness & conventional norms—more than others—such as harm norms? This is an important question to answer because it can shine light on whether we view different norms as psychologically distinct, such as the distinction posited by social domain theorists between conventional and moral norms (Smetana et al., 2014; Turiel & Dahl, 2019; Yoo & Smetana, 2022). Across our studies we find evidence that updating differed between vignette type. Namely, participants generally updated their beliefs to a greater extent for conventional and fairness behaviors compared to harm-related behaviors, and in some cases, preferences. Specifically, participants updated their injunctive beliefs more conventional and fairness norms than harm norms, although this result hinged on the wording on the injunctive norm measure and the type of harm behaviors. We find a similar pattern of results for behavioral intentions and personal moral beliefs: people's reported likelihood of engaging in the behavior, and the extent to which they thought it was moral, was more influenced by descriptive norm information for fairness behaviors than harm behaviors. However, we find little evidence of updating differences in second-order moral beliefs between fairness, conventional, and harm norms.

In contrast to our predictions, we found participants updated their beliefs and behavioral intentions about preference after receiving descriptive norm information. There was also mixed evidence regarding whether people update their beliefs for preferences less than other behaviors: while participants generally updated less for preferences than fairness or conventional behaviors, there was substantial variability in updating differences between harm behaviors and preferences across the dependent measures. For example, there was little difference in injunctive and personal norm belief updating between the harm and preference vignettes whereas participants updated their behavioral intentions more for preferences than harms but updated their second-order moral beliefs more for harms than preferences. Interestingly, we generally found little difference in injunctive and moral belief updating between fairness and conventional behaviors, but did find that participants updated their likelihood ratings more for conventional than fairness behaviors in two studies.

Altogether, these results provide some initial evidence that we are more influenced by descriptive norms for certain kinds of behaviors than others. Namely, our results suggest that descriptive norm information is more strongly associated with injunctive and moral norm beliefs for fairness behaviors and social conventions than harm behaviors or preferences. These results provide some for support for work on social domain theory which finds that people perceive moral and conventional norms as psychologically distinct (Smetana et al., 2014; Turiel & Dahl, 2019; Yoo & Smetana, 2022). Specifically, that we find that people's beliefs about conventional norms are more easily swayed by descriptive norm information than their beliefs about harm behaviors suggests that there is a psychological distinction between conventional and moral (e.g., harm) norms. However, it remains unclear what specifically constitutes a moral norm—are fairness norms also moral norms or are they distinct from norms of harm which are considered prototypically moral?

Recent developmental research finds that children do not perceive distributional fairness as a moral norm—by 4-years of age, children rated harm transgressions (e.g., hitting) as more severe than fairness or conventional transgressions (Yucel et al., 2022), while other work suggests children make different moral judgements about distributional unfairness than physical and psychological harm (Smetana & Ball., 2019). That we find a difference in updating between fairness and harm norms, albeit a relatively variable one, suggests that, while researchers generally consider both to be moral norms, people perceive them differently, at least to some degree. These results, in conjunction with previous work, suggests there are meaningful psychological differences between different kinds of moral norms and highlights the need for researchers to distinguish between them in their work rather than lump fairness and harm norms together under the umbrella of moral norms. However, it is important to note that the updating differences between vignette types found here varied considerably across studies and measures and thus limits our ability to make strong claims about domain differences. For example, in Experiment 5 we included a new set of harm behaviors designed to more exclusively tap harm concerns but found that, unlike the initial harm vignettes, people updated to a similar extent for them and the fairness behaviors. Why did we see domain updating differences between these two sets of harm vignettes? One reason might be because the new harm behaviors involved psychological harm rather than physical harm—some work suggests that people view physical harm as distinct from psychological harm (Smetana & Ball, 2019). This highlights a limitation with our stimuli—because we only used several highly controlled behaviors for each norm category, it is possible that our results are an artifact of the specific behaviors we used or at the very least, might not generalize to other kinds of behaviors. Future work should continue to explore whether the effect of descriptive norms differs between types of normative behaviors using a wider array of behaviors to ensure domain differences are not an artifact of the specific behaviors used here.

While the evidence was somewhat mixed, participants generally updated their beliefs less for preferences than fairness and conventional behaviors. Because preferences are not socially conditional—we should engage in them regardless of what others are doing or expect—they should be less sensitive to descriptive norms. Our data only partially support this idea: across our studies, participants often updated their beliefs about preferences less than fairness and conventional norms but more than harm norms. Importantly, participants still updated their beliefs and behavior for preferences and harm behaviors after receiving information that the behavior was common, just to a lesser extent than the fairness and conventional behaviors. The fact that participants updated their beliefs for preferences potentially conflicts with the socially conditional account of normative behavior which holds that what distinguishes norms from preferences are our beliefs about others' expectations (Bicchieri, 2005). However, that participants updated their beliefs for preferences is likely a consequence of the specific behaviors selected as preferences here. Alternatively, it is possible that norms are socially conditional but many of the things we consider preferences (e.g., wearing socks with sandals) are in fact closer to social conventions and thus sensitive to descriptive norm information. In other words, you might have a personal preference for vanilla ice cream, but if you see everyone in the ice cream shop ordering chocolate, you might infer that the chocolate ice cream there is extraordinary (or that the vanilla is bad) and decide to order the chocolate instead. Future work should continue to explore whether preferences are influenced by social expectations.

4. Study 4: Descriptive Norms Influence Children's Injunctive and Moral Norm Beliefs

When children form opinions about what the moral thing to do is and how they should act – they often look at how most people act. That is, their injunctive norms about what they *should* do is influenced by descriptive norms of what people usually *do* do. However, it remains unclear *how* exactly these regularities in the social environment influence different kinds of normative beliefs: Are children capable of flexibly tuning those beliefs depending on the frequency and type of behavior? We examined this question in 6-9-year-olds (N = 138) from the US in a preregistered study, asking whether children's injunctive beliefs, moral evaluations, behavioral intentions, and punishment ratings are influenced by descriptive norm information that a behavior is relatively common or uncommon. Since children readily distinguish between different categories of normative behaviors, we explored whether the influence of descriptive norm information varies depending on the category of normative behavior. Because the coronavirus pandemic offered a natural case study of novel norm learning, we explored this question in relation to COVID-related behaviors which children have only more recently acquired. Participants saw eight vignettes spanning four categories of behavior—negatively-valenced conventional, positively-valenced conventional, novel preferences, and COVID-related health behaviors—in which they either received a strong descriptive norm that the behavior was common or a weak descriptive norm that the behavior was uncommon. By 6 years of age, children's injunctive beliefs, moral evaluations, and behavioral intentions were more influenced by strong descriptive norms than weak descriptive norms across behaviors with the exception of the personal preferences, in which they were largely insensitive to the descriptive norm or were more influenced by weak than strong descriptive norms. Punishment judgements were also influenced by the commonality of the behavior, although this varied across categories. Our findings suggest that children's injunctive and moral beliefs are influenced by how common or uncommon a behavior is. Importantly, this influence does not generalize to all kinds of behaviors, pointing to a special role of social influence on beliefs for behaviors of social consequence.

This paper is co-authored with Emma Sansom, Julia Marshall, Young-Eun Lee, Felix Warneken, and Katherine McAuliffe

1. Introduction

Social norms are a foundational part of human societies and pervade nearly every aspect of social life–from how we eat and dress to how we share resources. A large body of work has found that children and adults acquire, conform, and enforce social norms across a diverse range of contexts and behaviors, spanning social conventions (Song et al., 1987; Yoo & Smetana, 2022), moral behaviors (Vaish et al., 2011; Yucel & Vaish, 2018), sharing resources (House, 2018; McAuliffe et al., 2017; McQuire et al., 2018), and playing conventional rule games (Diesendruck & Markson, 2011; Kanngiesser et al., 2022; Langenhoff et al., 2022; Rakoczy et al., 2008) to name only a few. How are we able to acquire these different kinds of norms? Researchers have investigated different means of norm acquisition, with much of the work focusing on direct instruction and pedagogy (Butler et al., 2015; Rakoczy et al., 2008; Rakoczy et al., 2010), as well as imitation and social learning (Hardecker & Tomasello, 2017; Muthukrishna et al., 2016). When it comes to theoretical approaches to explaining these behaviors, the Minimal Account of norms posits that we possess an innate norm psychology that is designed to expect cues and regularities in one's social environment in order to acquire information about local norms (Kelly & Davis, 2018). This paper draws on this account, asking how we acquire norms from observations of behavioral regularities in our social environment. We frequently receive information in our social world about what is common or uncommon in our group and use that information to shape our beliefs and guide our behavior. Consequently, studying this route for acquiring social norms is important for informing our understanding of the social cognition of norms.

Much of the recent work on social norms has examined two distinct kinds of normative information—injunctive norms, what we think others approve of or expect of us (norms of *ought*), and descriptive norms, what we think other people commonly do (norms of *is*; Cialdini et al., 1990). While generally thought of as conceptually distinct, recent work suggests that there is a close relationship between descriptive and injunctive norm beliefs. This work finds that people have an implicit association between descriptive and injunctive norms, make explicit bi-directional inferences between them, and infer what ought to be from what is typically done (Eriksson et al., 2015; Tworek & Cimpian, 2016). There also appears to be a close relationship between descriptive and moral norms with other work finding that people infer moral beliefs from descriptive norms such that they find behaviors that are common as more moral than those that are uncommon (Lindstrom et al., 2018).

In the developmental literature, previous work suggests that children make normative judgements about how one should behave from what is common in a group (Roberts et al., 2017; Roberts & Horii, 2019), a finding that replicates cross-culturally (Roberts et al., 2018). Specifically, when presented with information about common, morally neutral behaviors (e.g., the kind of music they listen to) in a novel social group, children were more likely to disapprove of agents that did not conform to the behavior than those who did. Other work suggests that 4-7-year-old children make these injunctive (*ought*) inferences from observations of typical behaviors in part because they make value-based judgements that something is right from its inherent features (e.g., people should give others roses on Valentine's day because roses are beautiful; Tworek & Cimpian, 2016). The tendency to infer 'ought' from 'is' emerges early in development: 3-year-old children infer social norms, and are willing to enforce them, after watching a single action by an adult (Schmidt et al., 2016). Moreover, additional work finds that children make prescriptive judgments about what ought to be from what is typical for a category (e.g., Zebra's *should* have stripes because the typical Zebra does; Foster-Hanson et al., 2021; Foster-Hanson & Lombrozzo, 2022). While this evidence suggests that descriptive norm information about group regularities generally influences injunctive norm beliefs, other work finds that the types of norms are sometimes dissociated such that children's beliefs about how others should behave can diverge from their expectations of how they themselves will behave (DeJesus et al., 2014; Smith et al., 2013). Thus, while there is generally a close association between descriptive and injunctive norms, the fact that they are sometimes dissociated suggests that they are distinct types of normative information.

While previous work suggests that children infer injunctive norm information from what is common (Roberts et al., 2017; Roberts & Horii, 2019), little work has explicitly manipulated the frequency of behavioral regularities to measure how it changes injunctive norm beliefs in addition to behavioral intentions, moral evaluations, and punishment judgements, all within one experimental design. Doing so will allow us to not only better understand the relationship between descriptive and injunctive norm beliefs but also how regularities in the environment shape a suite of other important normative beliefs about morality, punishment, and behavior. Furthermore, it remains unclear whether the influence of behavioral regularities on behavior is consistent across types of normative behaviors, particularly those that differ in their social consequences (e.g., the extent to which they affect other agents). That is, do children make the same kind of injunctive inferences from descriptive norms for all kinds of behavior or do they do so selectively for some behaviors but not others? If children selectively change their beliefs in response to social regularities depending on social consequence of the behavior, this might suggest that there are important psychological differences between how children view these different behaviors. Moreover, answering this question would inform our understanding of the contexts in which children are the most likely to be influenced by the behavior of their peers.

Here we investigate whether children flexibly tune their injunctive and moral beliefs in response to different frequencies of behavioral regularity across different kinds of social contexts. Using a developmental approach is critical to addressing this question because middle childhood is a period in development where children begin to become rapidly more sensitive to social norm information and thus represents an ideal age to study norm learning and cognition. While previous work has found that children start to infer injunctive norm information from descriptive norm information by 3-4-years-old, other research suggests that children become increasingly influenced by social norms throughout middle childhood and are not sensitive to certain normative information, such as advantageous inequity aversion or sharing norms, until 6-8-years of age (Blake et al., 2015; House & Tomasello, 2018; House et al., 2020). Consequently, we identified 6-9-years of age as a critical point in middle childhood where children might be most likely to show a sensitivity to descriptive norms across different kinds of normative behaviors. More generally, taking a developmental approach can also offer insight into how foundational these inferences are in social cognition–if children are already highly sensitive to descriptive norms by age 6, that would suggest that the ability to infer normative beliefs from behavioral regularities is a relatively foundational aspect of social cognition (Olson & Dweck, 2008; Olson & Dunham, 2010).

How do children learn what is right and what is wrong? While pedagogy and explicit teaching clearly play a large role (Butler et al., 2015; Rakoczy et al., 2008; Rakoczy et al., 2009), observations about how other people behave is likely to also be an important contributor to norm learning. However, it remains unclear whether children selectively incorporate descriptive norm information in their beliefs for different categories–updating their prior normative beliefs more for some behaviors than others in response to the same information–or whether they inflexibly incorporate descriptive norm information to similar extents for any category of behavior. If the effect of descriptive norm information on normative beliefs varies by behavior, this would lend credence to the hypothesis that there are meaningful psychological differences between different categories of behaviors, such as predicted by social domain theory (Killen et al, 2006; Smetana, 2013). If, on the other hand, children respond similarly to descriptive norm information for different kinds of behaviors, this might suggest that we view different categories of behavior similarly, at least in the context of norm learning.

Additionally, the coronavirus pandemic has provided a naturalistic case study of how children learn novel social norms in real time. For norms of fairness or honesty, children have already received substantial information from their social environment about the behaviors in question, making it difficult to separate social learning from the underlying norm cognition. To strip away prior social learning and experience to better understand how readily adapted humans are for norm cognition, researchers have used entirely novel norms and groups, such as aliens on an unknown planet (Roberts et al., 2017; Roberts et al., 2019). While this work has made an important contribution to our understanding of norm cognition, relatively less work has examined how children learn novel norms that they actually experience in their lives. As a result of the coronavirus pandemic, children have had to rapidly learn new norms that they previously had no prior information about. While in the intervening two years since the World Health Organization declared coronavirus a pandemic children have received substantial social information about COVID-related norms (Cucinotta & Vanelli, 2020), this still constitutes significantly less social information than they have received for other normative behaviors (e.g., their entire lives). Moreover, these novel norms (e.g. mask wearing and social distancing) are not as ingrained in the lives of children as common moral norms such as hurting others or stealing that have been an integral part of stories, fables, films, songs, informal and formal moral education at home, in schools or religious institutions for generations. Consequently, investigating how descriptive norms influence beliefs about COVID behaviors offers an informative, naturalistic case study of how children form normative beliefs.

1.1. Differences between norm categories

While past work suggests there is a close relation between descriptive and injunctive norms, the strength of that relation might vary depending on the kind of behavior in question. There is now a sizeable body of work on social domain theory which finds an important distinction between moral norms, which are universally applied and obligatory, and social conventional norms, which are perceived as subjective and alterable, finding that by age 4 children start to consistently differentiate between these behaviors (Killen et al., 2006; Smetana, 2013; Yoo & Smetana, 2022). Indeed, recent work with adults suggests that the effect of descriptive norms on injunctive norm beliefs varies depending on the kind of norm, such that people are generally less sensitive to descriptive norms for harm behaviors than conventional or fairness behaviors (Deutchman et al., preprint). Another important distinction is between social norms and personal preferences. Preferences are largely independent and unconditional, meaning that they are things we are internally motivated to do regardless of what others do or expect of us (e.g. if it's raining outside, I'm going to use my umbrella regardless of what other people are doing or expect (Bicchieri, 2017). Furthermore, personal preferences typically do not impact others in the same way that social conventional norms do (my decision to use an umbrella does not affect others–whereas my decision to help clean up or share a resource has direct consequences for other people). This distinction suggests that children's injunctive beliefs about preferences should be less influenced by descriptive norm information, if at all, as compared to conventional behaviors that are socially conditional and consequential, meaning that our decision to comply with them depends on others' beliefs and behavior and affects other agents. Indeed, young children tend to conform more to behaviors framed as conventional norms than personal preferences (Li et al., 2021), while adults update their beliefs less for preferences than conventional norms, suggesting they view preferences differently (Deutchman et al., preprint).

While much of the work in this area has examined the effect of group regularities on norm learning in the context of third-party stories (Roberts et al., 2017; Roberts et al., 2019), comparatively less work has examined how children conceptualize new norms that they encounter in their actual social environments. As described above, the coronavirus pandemic presented a rare opportunity to study how children spontaneously learn new social norms such as wearing masks and social distancing. Therefore, the fourth category of behaviors we included were COVID-related health behaviors. As described above, including COVID-related norms allows us to study how children learn relatively novel norms and thus can inform our understanding of norm learning. Additionally, while these behaviors have become an important aspect of children's lives, it remains unclear how children think about these new norms. Therefore, including COVID norms offers the practical benefit of teaching us more about how children conceptualize norms that carry important health consequences. One possibility is that they treat these types of norms more like socially-consequential conventional norms, such as talking in the library or cleaning up your lunch tray, and thus might be more sensitive to the influence of descriptive norm information. Alternatively, children might view COVID-related norms as distinct from conventional norms. For example, it is possible children might view these behaviors more like personal preferences that lack social consequences–in which case, children might be less sensitive to descriptive norm information than they would be for conventional norms. To help address these questions, we explored whether children's beliefs about COVID-related health behaviors were as readily influenced by descriptive norm information as other kinds of behaviors. If children's beliefs about COVID behaviors are resilient to descriptive norms, this might suggest that children view them distinctly from conventional norms.

1.2. Present study

In the present preregistered study, we examined whether children's injunctive norm beliefs, moral evaluations, behavioral intentions, and punishment judgements are influenced by descriptive norm information that a behavior is relatively common or uncommon. Because past work suggests there are cognitive differences between different kinds of normative behaviors, we assessed whether the influence of descriptive norms on beliefs varies depending on the type of normative behavior. To investigate this, we presented 6-9-year-old children with a series of animated vignettes depicting different behaviors in which there was either a strong descriptive norm that the behavior was common or a weak descriptive norm that the behavior was uncommon, such that almost everyone, or almost no one, was engaging in the behavior, respectively. To explore whether the relationship between descriptive and injunctive norms varies across behaviors, as well as to gain insight into whether children view social conventions and personal preferences as psychologically distinct, we studied four different types of behaviors. We included negatively valenced conventional behaviors (e.g., talking in the library) and positively valenced conventional behaviors (cleaning up your lunch tray) that are socially consequential (e.g., complying with or violating the norm affects other agents), personal preferences (writing with a red pen rather than a blue pen), in which we did not expect descriptive norm information to influence injunctive beliefs, and COVID-related health behaviors for the two reasons mentioned above.

If children find the behaviors as more injunctive in the strong than the weak descriptive norm condition, then this would suggest that they infer injunctive norm beliefs from descriptive norm information and are sensitive to the strength of that information. If we find that children rate the behavior to be more moral and report being more likely to engage in it, when there’s a strong descriptive norm than a weak descriptive norm, then this would suggest that descriptive norms directly influence moral evaluations and behavioral intentions. If children rate violating the positive conventional norms as more deserving of punishment when there is a strong descriptive norm than a weak descriptive norm, this would indicate that children rely on descriptive norm information to infer the wrongness of a behavior. In other words, if most people are cleaning up their lunch trays, children will view failing to do so as deserving of more punishment than if only a few are cleaning up their trays. Similarly, if children rate complying with the negative conventional norms as more deserving of punishment when there is a weak descriptive norm than a strong descriptive norm, that would provide additional evidence that children infer the wrongness of a behavior from how common it is. If most people are talking in the library, children will view doing so as less deserving of punishment compared to when only a few are talking in the library.

Lastly, if children perceive the different normative behaviors as psychologically distinct, then the association between descriptive norms and their beliefs and behavioral intentions should vary across categories of behaviors. Namely, children’s beliefs should be more influenced by descriptive norms for conventional behaviors than personal preferences, where there should be little to no effect of descriptive norms on beliefs. If children view COVID-19 health behaviors differently from conventional norms and more like personal preferences which do not affect others, they should be less sensitive to descriptive norms for the COVID behaviors than the conventional behaviors. If children view COVID norms similarly to socially-consequential conventional norms, then they should be similarly sensitive to descriptive norms for COVID-related behaviors as the conventional behaviors.

Answering these questions will inform our understanding of how regularities in our social environment contribute to shaping injunctive and moral norms as well as shine light on whether there are meaningful psychological differences between conventional norms with social consequences and personal preferences. Additionally, it will also inform our understanding of how children conceptualize COVID norms, teaching us about how children form norm beliefs in relatively novel contexts as well as whether children perceive COVID norms more similarly to social conventions or personal preferences.

2. Method

2.1. Participants

We tested N = 138 6-9-year-old children recruited into two age groups: 6-7-year-olds (N = 65, M = 7.05, SD = 0.57, range = 6.01-7.97, 52.3% females) and 8-9-year-olds (N = 73, M = 8.96, SD = 0.55, range = 8.00-9.96, 46.7% females). Participants were majority white (White = 68.9%, Asian = 11.6%, Hispanic = 4.3%, Black = 3.6%, Biracial = 7.9%, Other = 2.9%). Participants were recruited via a lab data base and Facebook ads across the United States. The minimum sample size of N = 138 was determined using a power simulation with simr which found that we would have 80% power to detect a medium interaction effect between descriptive norm condition and vignette type (Green & MacLeod, 2016). All participants were from the United States and were tested online via Zoom conferencing technology in moderated sessions between June 11th 2021 and

September 16th 2022. While we initially recruited N = 144, 6 children were excluded from data analysis for meeting our preregistered exclusion criteria, including parental interference (1), equipment failure (1), failing the comprehension checks (2), failing to complete the study in its entirety (1), and severe inattention (1). This study was approved by the IRB (#16.242.04-33) and preregistered prior to data collection: https://aspredicted.org/TH6_64X.

2.2. Materials

Participants saw a series of animated gifs of scenes depicting four different categories of behavior and descriptive norm information (see Figure 1). Animations were created using Vyond animation software. The animations and questions were embedded in a Qualtrics survey that an experimenter controlled and was shown to participants via Zoom's screen sharing function (see Sheskin et al., 2020 for a discussion of the validity of remote data collection).

2.3. Design

Children saw eight animated scenarios depicting four different categories of normative behaviors which were presented in a randomized order: negatively valenced conventional behaviors, positively valenced conventional behaviors, COVID health behaviors, and personal preferences. Participants were assigned between-subjects to either the weak or strong descriptive norm condition. Depending on condition, children saw that either almost no one (one out of five characters; *weak descriptive norm condition*) or almost everyone (four out of five characters; *strong descriptive norm* *condition*) was engaging in the behavior depicted in the vignette. The presentation order of the dependent variables was counterbalanced.

2.4. Procedure

Children were first introduced to the experiment and provided their consent to participate. They then had the opportunity to select one of two avatars (one male, one female) to represent them in the stories which we included in order to promote children's engagement with the vignettes. The avatar that participants selected was used in the stimuli for the rest of the experiment.

Participants then saw eight different animated scenarios that corresponded to four different categories of behavior: negatively valenced conventional behaviors (hereafter *negative conventional*), positively valenced conventional behaviors (hereafter *positive conventional*), COVID health-related behaviors, and personal preferences. See Figure 1 for a diagram of the task. We chose both negatively and positively valenced behaviors in order to ensure the relationship between descriptive and injunctive norms does not vary depending on the emotional valence of the conventional norm. The conventional vignettes detailed scenarios with socially conventional norms such as talking in the library and walking on someone's yard (negative conventional) or leashing your dog and cleaning up your lunch tray (positive conventional). The COVID health behavior vignettes depicted two scenarios that took place during the coronavirus pandemic, double-masking inside and social distancing.[9] All the characters in the COVID vignettes

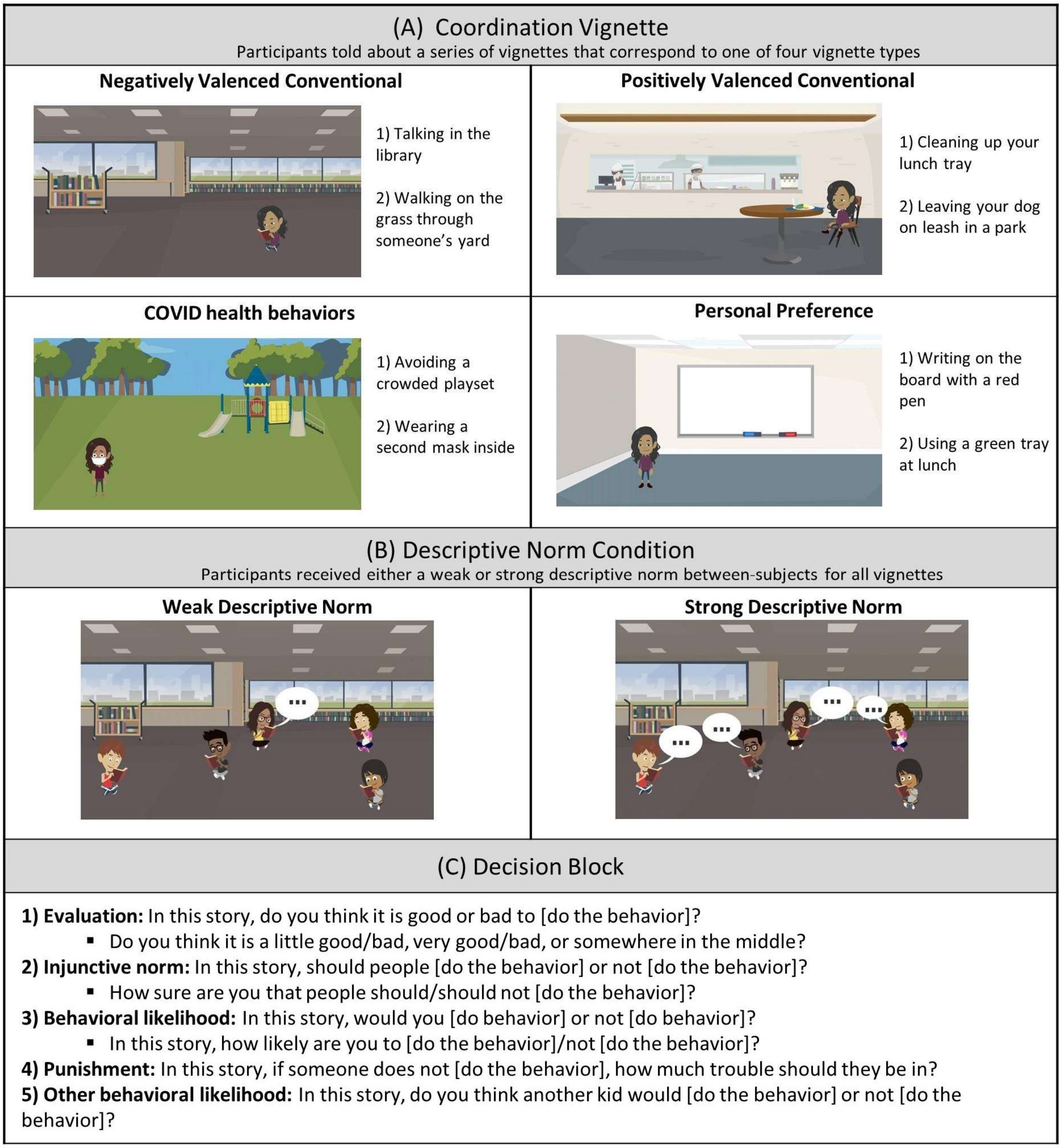
[9] When we designed these vignettes there was still uncertainty regarding the value of double masking and whether coronavirus spread via surfaces, topics on which attitudes rapidly changed throughout the pandemic.

were wearing masks to visually denote that they were taking place during the coronavirus pandemic. Lastly, the personal preference vignettes depicted scenarios with arbitrary behaviors that did not impact others and so solely hinged on one's personal preference: writing on the board with a red pen (as opposed to a blue pen) and using a green lunch tray (as opposed to an orange tray). At the start of each vignette type block, participants answered a comprehension question assessing whether the story took place during or before the coronavirus pandemic (e.g. "are these stories happening during COVID times or before COVID times?"; see supplement for comprehension question text). They were given three attempts to answer the question correctly before moving on (see exclusion criteria in our preregistration).

In each vignette, participants were asked to imagine themselves in the scenario and were shown an animated scene depicting the normative behavior. In the weak descriptive norm condition, participants were told almost no one was doing the behavior and saw an animation in which only one out of five characters were engaging in the behavior. In the strong descriptive norm condition, participants were told almost everyone was doing the behavior and saw an animation of four out of five characters engaging in the behavior. After receiving the weak or strong descriptive norm for each vignette, participants answered a comprehension check to ensure they understood the descriptive norm manipulation (e.g., "in this story, is almost everyone or almost no one keeping their dog on the leash?"). Participants had three attempts to answer the question correctly. If participants answered incorrectly, they were reminded of the answer ("you look around and see almost everyone is keeping their dog on the leash") and asked again.

If they answered incorrectly on their third attempt, they were reminded of the correct answer and moved on to the next screen but were excluded from data analysis.

After receiving the descriptive norm information and answering the comprehension checks, participants answered the five dependent variables. First, participants answered a moral evaluation asking whether the behavior in the vignette was good or bad and then depending on their choice, received a 5-point scale assessing how good or bad it is. Next, participants answered an injunctive measure asking whether people should engage in their behavior and then rated their certainty in that choice (continuous: 1-5). Participants then answered a behavioral intention measure of whether they would engage in the behavior or not, and then depending on their decision, the likelihood of doing that behavior or not doing the behavior (continuous: 1-5). To measure punishment evaluations, participants rated how much trouble someone should be in if they violate the norm; however, for the conventional negative vignettes, participants were asked how much trouble some should be in for engaging in the norm. Lastly, to measure others' expected behavior in the scenario, participants were asked to predict whether they thought another kid would do the behavior or not do the behavior. Participants answered these questions eight times, once for each vignette. The dependent variables were presented in two orders which were counterbalanced across participants. After completing all the vignettes, participants answered six questions assessing whether or not they thought the various behaviors in the task are injunctive in general (e.g., "In general, should people talk in a library?"; see supplement). All of our code, materials, and data are publicly available online at the Open Science Framework (https://osf.io/qpzwv/?view_only=0ed5812f3beb4384a878b9924951b109).


(A) Coordination Vignette
Participants told about a series of vignettes that correspond to one of four vignette types
Negatively Valenced Conventional
1) Talking in the library
2) Walking on the grass through someone's yard
Positively Valenced Conventional
1) Cleaning up your lunch tray
2) Leaving your dog on leash in a park
COVID health behaviors
1) Avoiding a crowded playset
2) Wearing a second mask inside
Personal Preference
1) Writing on the board with a red pen
2) Using a green tray at lunch
(B) Descriptive Norm Condition
Participants received either a weak or strong descriptive norm between-subjects for all vignettes
Weak Descriptive Norm
Strong Descriptive Norm
(C) Decision Block

1) Evaluation: In this story, do you think it is good or bad to [do the behavior]?
- Do you think it is a little good/bad, very good/bad, or somewhere in the middle?

2) Injunctive norm: In this story, should people [do the behavior] or not [do the behavior]?
- How sure are you that people should/should not [do the behavior]?

3) Behavioral likelihood: In this story, would you [do behavior] or not [do behavior]?
- In this story, how likely are you to [do the behavior]/not [do the behavior]?

4) Punishment: In this story, if someone does not [do the behavior], how much trouble should they be in?

5) Other behavioral likelihood: In this story, do you think another kid would [do the behavior] or not [do the behavior]?

Figure 1. Diagram of the norm task with images from the stimuli. In the negatively valenced conventional vignettes, the punishment measure asked how much trouble someone should be in if *they did* the behavior. © 2022 GoAnimate, Inc.[10]

[10] Images are copyrighted by and used by permission of VYOND™. VYOND is a trademark of GoAnimate, Inc., registered in Australia, Brazil, Canada, Chile, Egypt, the European Union, Hong Kong,

2.5. Analyses

We randomly selected 20% of participant videos to be coded by an independent coder who found that there was 99.96% agreement (Cohen's Kappa = 1) between the videos and our Qualtrics data on all eight of our dependent variables. All analyses were conducted in R version 4.1.2 (R Core Team, 2020). In order to capture the bidirectional range of variance in one measure, we combined the 5-point scales for each binary choice (e.g., good, bad) to create 10-point scales for the evaluation and likelihood measures as done with previous work (Marshall et al., 2022). If participants evaluated a behavior as good, their response was re-coded from 6-10 (e.g., a five indicating 'very good' would be a 10), while if they evaluated a behavior as bad, their response was reverse coded (e.g., a five indicating 'very bad' would be a 1; see supplement for details). We ran eight preregistered generalized linear mixed models predicting our dependent measures by descriptive norm condition (models 1-6, 8) and by the interaction between vignette type and norm condition (model 7).

First, to examine whether descriptive norm information influences moral evaluations, model 1 predicted evaluation ratings (continuous: 1-10) by descriptive norm condition (categorical: weak, strong) while model 2 predicted binary evaluations (1 = good, 0 = bad) by norm condition. Second, to explore whether descriptive norms influence injunctive norm beliefs, model 3 predicted binary injunctive beliefs (1 = injunctive, 0 = not injunctive) by norm condition while model 4 predicted certainty that the behavior was injunctive (model 4; continuous: 1-5). To explore whether descriptive

India, Indonesia, Israel, Japan, Malaysia, Mexico, New Zealand, Norway, OAPI, the Philippines, Russia, Singapore, Switzerland, the United Kingdom and Vietnam.

norm condition influenced behavioral intention ratings, model 5 predicted binary behavior (0 = don't do behavior, 1 = do behavior) by norm condition. Model 6 predicted continuous likelihood (1-10) by condition; while we preregistered that this model would include the 5-point likelihood scale ("in this story, how like are you to do X"), we combined them into a 10-point scale to capture the full range of variance (we report the results of the 5-point scale analysis in the supplement). To examine whether descriptive norm condition differentially influences punishment judgements depending on the type of norm, model 7 predicted punishment ratings by norm condition, vignette type (conventional negative, conventional positive, COVID health behaviors, & personal preferences; baseline = conventional negative), and their interaction. Lastly, to test whether descriptive norm condition predicts beliefs about others' future behavior, model 8 asked whether norm condition predicted others' anticipated behavior (binary: 1 = do behavior, 0 = not do behavior).

While we did not expect to find any age effects, to test whether our results were robust to participant age, we ran a series of exploratory models including age as a covariate in our main models (models 1-6, 8) and compared them to models without the age term. We note here that these comparisons find that age added no explanatory value (see Supplement for details). We also ran a series of exploratory models predicting injunctive beliefs, evaluations, and behavioral intention by descriptive norm condition, vignette type, and their interaction. Next, we created an exploratory model predicting certainty ratings that the behavior was not injunctive by descriptive norm condition (see the supplement for these exploratory analyses). Lastly, as an exploratory analysis, we tested for mediation with a multilevel structural equation model. This model included behavioral intentions (continuous: 1-10) as the endogenous variable, descriptive norm condition (weak, strong) as the exogenous variable, and evaluations (continuous: 1-10) as the mediator. To decompose the interactions between descriptive norm condition and vignette type, we made pairwise comparisons using estimated marginal means adjusted using the multivariate t method (MVT) to correct for multiple comparisons. We report the results of these comparisons in the supplement unless otherwise noted.

3. Results

3.1. Do descriptive norms influence children's normative beliefs?

Figure 2 displays the results for the evaluation (panel A), injunctive (panel B), likelihood (panel C), predicted behavior (panel D), and punishment judgements measures (panel E) between the different behaviors (see the supplement for results separated by individual vignette). Participants were more likely to evaluate the behavior as good when there was a strong descriptive norm than a weak descriptive norm, both when treating evaluations as a binary outcome measure (B = 1.25, SE = 0.26, *p* < .001, OR: 3.48, 95% CI: 2.10, 5.78) and as a continuous measure (B = 0.52, SE = 0.19, *p* = .006, 95% CI: 0.16, 0.89). Participants were also more likely to view the behavior as an injunctive norm when there was a strong descriptive norm that the behavior is common than a weak descriptive norm that it is uncommon (B = 0.94, SE = 0.23, *p* < .001, OR: 2.56, 95% CI: 1.62, 4.03). When examining certainty in injunctive beliefs, we find that participants were not more certain in the strong than the weak descriptive norm condition (B = -0.16, SE = 0.12, *p* = .21, 95% CI: -0.40, 0.09). Participants were, however more likely to report that they would engage in the behavior when there was a strong descriptive norm than a weak descriptive norm for the binary measure (B = 0.48, SE = 0.20, *p* = .02, OR: 1.62, 95% CI: 1.09, 2.40), however this difference was trending on significance for the continuous measures (B = 0.49 , SE = 0.26, *p* = .057, 95% CI: -0.01, 1.00). Participants were significantly more likely to expect another kid to engage in the behavior when there was a strong descriptive norm than a weak descriptive norm (B = 2.29, SE = 0.26, *p* < .001, OR: 9.89, 95% CI: 6.00, 16.32).

3.1. Does the influence of descriptive norms vary by category of behavior?

We found a significant interaction between norm condition and the negative conventional and positive conventional behaviors on punishment ratings such that punishment ratings were higher in the weak than the strong descriptive norm condition for the negative behaviors but higher in the strong than the weak descriptive norm condition for the positive behaviors (B = 0.45, SE = 0.17, *p* = .009, 95% CI: 0.11, 0.79). Similarly, there was a significant interaction between norm condition and the negative conventional and COVID behaviors such that punishment ratings were higher in the weak than the strong descriptive norm condition for the negative behaviors but higher in the strong than the weak descriptive norm condition for the COVID behaviors (B = 0.35, SE = 0.17, *p* = .045, 95% CI: 0.01, 0.69). The interaction between the conventional negative behaviors and the personal preference was not significant (B = 0.28, SE = 0.17, *p* = .11, 95% CI: -0.06, 0.62). This suggests that the effect of descriptive norm information on punishment judgements depends on the category of behavior such that people punish more when there’s a violation of a common positive conventional (e.g., cleaning up your lunch tray) or COVID norm (e.g., double masking) but punish complying with a negative-conventional norm (e.g., talking in the library) more when uncommon than common.

We found no interaction between descriptive norm condition and vignette type on binary evaluations: the interactions between descriptive norm condition and the negative and positive conventional behaviors (B = -0.39, SE = 0.55, *p* = .47, OR: 0.67, 95% CI: 0.23, 1.98), negative conventional and COVID behaviors (B = -0.26, SE = 0.56, *p* = .64, OR: 0.77, 95% CI: 0.26, 2.32), and negative conventional and preference behaviors (B = -0.85, SE = 0.53, *p* = .11, OR: 0.43, 95% CI: 0.15, 1.21) were not significant. This suggests that the influence of descriptive norms on evaluations of the behavior as good or bad was consistent across all vignette categories. For binary injunctive ratings, we found a significant interaction between descriptive norm condition and the conventional negative and COVID behaviors (B = -0.95, SE = 0.48, *p* = .047, OR: 0.39, 95% CI: 0.15, 0.99) and negative conventional and personal preferences (B = -2.02, SE = 0.45, *p* < .001, OR: 0.13, 95% CI: 0.05, 0.32) such that participants were more likely to report the behavior as an injunctive norm when it was common than uncommon for the negative conventional behaviors (e.g., talking in the library) than the COVID behaviors (e.g., double masking) or preferences (e.g., using an orange tray). The interaction between the negative and positive conventional behaviors was not significant (B = 0.03, SE = 0.64, *p* = .96, OR: 1.03, 95% CI: 0.29, 3.59), indicating that the effect of descriptive norm information was consistent across positively- and negatively-valenced conventional norms.

Lastly, looking at behavioral intentions, we found a significant interaction between descriptive norm condition and the negative conventional and personal preferences such that participants were more likely to say they would do the behavior in the strong than the weak descriptive norm condition for the negative behaviors but more likely in the weak than the strong descriptive norm condition for the personal preferences (B = -1.45, SE = 0.41, *p* < .001, OR: 0.23, 95% CI: 0.10, 0.52). In other words, participants were *more* likely to report they would talk in the library (negative conventional) when it was common than uncommon but that they would be *less* likely to use an orange lunch tray (personal preference) when it was common than uncommon. The interactions between descriptive norm condition and conventional negative and conventional positive (B = 0.69, SE = 0.53, *p* = .19, OR: 2.01, 95% CI: 0.72, 5.66) and conventional negative and COVID behaviors were not significant (B = 0.07, SE = 0.42, *p* = .86, OR: 1.07, 95% CI: 0.48, 2.43), suggesting that the descriptive norm influenced behavioral intentions similarly between the conventional and COVID behaviors.

Evaluations of the behavior partially mediated the effect of descriptive norm condition on behavioral intentions. The direct effect of descriptive norm condition on behavioral intentions was significant (*b* = 0.52, SE = 0.26, *p* = .044). The path from descriptive norm condition to evaluations was significant (*b* = 0.46, SE = 0.19, *p* = .015), with norm condition explaining 37.8% of the variance in evaluations. The path from evaluations to behavioral intentions was also significant (*b* = 0.73, SE = 0.04, *p* <.001), explaining 55.2% of the variance in behavioral intentions. Critically, the indirect effect was significant (*b* = 0.334, SE = 0.138, *p* = .015), explaining 20.9% of the total variance, suggesting that descriptive norm information influenced behavioral intentions in part because it increased beliefs that the behavior was moral.

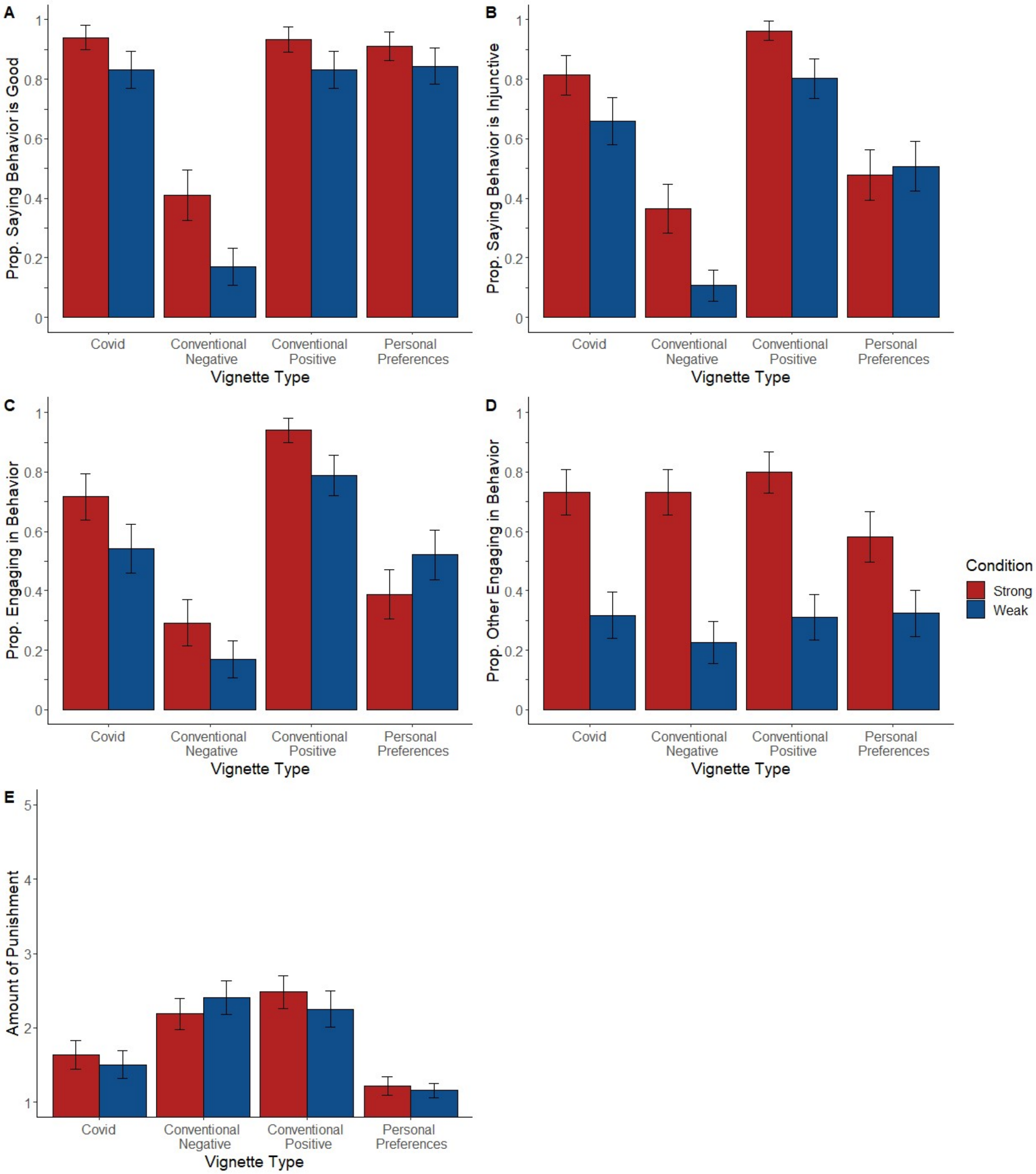


Figure 2. Proportion of participants who responded saying the behavior was good (A), injunctive (B), likelihood to engage in it (C), likelihood another kid would engage in it

(D), and amount of punishment deserved for violating the norm (or complying with it in the negative-conventional vignettes; E). Error bars indicate 95% confidence intervals.

4. Discussion

We examined how learning that a certain behavior is common or uncommon influences children's perception of a behavior as normative. We find that differences in the frequency of a behavior changes children's perceptions in several key indicators of behaviors as being normative: injunctive beliefs about how one should behave, moral evaluations, behavioral intentions, and punishment judgements. There were three key findings from this study. First, we found that by 6-years of age, children's injunctive norm beliefs about how they should behave were influenced by descriptive norm information about how common the behavior was. Namely, children were more likely to think the behavior was injunctive in the strong descriptive norm condition where the behavior was common than the weak descriptive norm condition where it was uncommon. Additionally, we found that observing that individuals commonly engage in a behavior resulted in children not only predicting that other individuals would be more likely to engage in the behavior, and reporting that they would be more likely to do so as well, but also that children attached moral value to these behaviors by judging norm followers as more moral than norm violators. Second, we found that children flexibly adjusted their normative beliefs depending on the category of behavior: children were influenced more by strong descriptive norm information for conventional behaviors of social consequence while they were not influenced at all, or influenced more by weak descriptive norm information, for personal preferences that do not affect other agents.

Furthermore, we found that the effect of descriptive norm information on children's punishment judgements partly depended on the type of behavior: Children rated violations of positive conventional norms (e.g., cleaning up your tray) as more deserving of punishment when most people were following the norm than when few were. Conversely, children thought complying with the negative conventional norm (e.g., talking in the library) was more deserving of punishment when it was uncommon to do so than when it was common. Third, we found that children were similarly influenced by descriptive norms for conventional and COVID-related behaviors but differed compared to personal preferences, suggesting that children perceive the COVID norms as more similar to socially impactful conventional norms than preferences.

Our findings suggest that children are highly attuned to regularities in their social environment, readily adjusting their normative beliefs in response to descriptive norm information that a behavior is common. Additionally, that descriptive norm information influenced moral evaluations in addition to injunctive beliefs suggests that there is likely an important relationship between injunctive norms and moral judgements. More generally, our finding that the relationship between descriptive and injunctive norms is already present by 6 years of age indicates that the is-to-ought association is likely a fundamental aspect of human norm cognition. Furthermore, the results of the exploratory meditation analysis indicate that the effect of descriptive norms on behavioral intentions was partially mediated by moral evaluations. These results support previous work which has found a close association between injunctive and descriptive norm information in adults (Eriksson et al., 2015) as well as other developmental work which finds that children infer injunctive norm information from regularities in the social environment

(Roberts et al., 2017; Tworek & Cimpian, 2016). Our findings extend this past work, showing that children adjust several of their normative beliefs in response to the frequency of behavioral regularities and do so for behaviors they regularly encounter beyond the context of novel norms and groups (Roberts et al., 2017; Roberts et al., 2019). Our results are also congruent with adult work which finds that people associate commonality with morality, finding common behaviors to be more moral than uncommon behaviors (Lindstrom et al., 2018). Additionally, our mediation results support recent work with adults that descriptive norm information influences behavioral intentions in part because it changes beliefs about the morality of the behavior (Deutchman et al., preprint).

We found that the effect of descriptive norm information on beliefs varied depending on the category of behavior. One measure for which we found meaningful between-behavior differences is punishment. Our results indicate that children rely on descriptive norm information about how frequent a behavior is to infer the wrongness of a behavior. In support of our prediction, we found a significant interaction for punishment ratings: children judged not following the positive conventional norms as more deserving of punishment when common than uncommon but rated following the negative conventional norms as more deserving of punishment when uncommon than common. In other words, when most people were cleaning up their lunch trays (i.e., positive conventional), children viewed failing to do so as deserving of more punishment than if only a few were cleaning up their trays. In contrast, when most people were talking in the library (i.e., negative conventional), children viewed doing so as less deserving of punishment compared to when only a few were talking in the library. This finding supports recent work with adults which finds that both second- and third-party punishment are influenced by descriptive norm information about how common it is for others to cooperate and punish free-riders (Li et al., 2021; Lois et al., 2019). However, this result conflicts with recent adult work which found that descriptive norm information had little influence on punishment judgements (Deutchman et al., preprint). Altogether, this evidence suggests that punitive behavior is influenced by descriptive norm information and that this association is present relatively early in development. Future work should further explore the contexts in which descriptive norms do and do not influence punishment judgements and behavior.

Our results also suggest that children were similarly influenced by descriptive norm information for socially consequential conventional norms and COVID norms: We found little difference in the effect of descriptive norms on beliefs and judgements between the negative conventional, positive conventional behaviors, and the COVID health behaviors. That said, the effect of descriptive norm condition on injunctive beliefs, and to a lesser extent, evaluations, was larger for the negative conventional behaviors than the positive conventional or COVID behaviors which tended to pattern similarly. This might suggest that children are more sensitive to social information for selfish behaviors than prosocial behaviors, which has been found in the context of norms of resource allocations in 4-5-year-olds but not 6-9-year-olds (McAuliffe et al., 2017). Conversely, the larger influence of descriptive norms on beliefs for the negative conventional behaviors might simply reflect that there was greater range for movement since ratings for these behaviors were closer to the midpoint of the scale while the positive conventional and COVID behaviors were closer to ceiling.

Our finding that children were already sensitive to descriptive norm information about the frequency of COVID-related behaviors by 6 years of age indicates that children can rapidly acquire relatively novel norms from their environment. More generally, this provides evidence that children make descriptive-to-injunctive normative inferences for novel norms they have experience with in their everyday lives, rather than solely in the context of entirely novel, fictional norms in which this question has frequently been studied (Roberts et al., 2017; Roberts et al., 2019). Our results also shine light on how children conceptualize these novel COVID norms. Namely, that children were influenced by strong descriptive norms for the COVID behaviors but not the preferences, suggests that they do not perceive them like personal preferences which do not impact others. Whereas our finding that children were as influenced by descriptive norms for the COVID behaviors as the conventional behaviors suggests that children perceive COVID-related norms much like conventional norms that affect other agents. This is an important finding given the increasing intensity and frequency of climate-related natural disasters and epidemics (Marani et al., 2021; Ritchie and Roser, 2019), it is likely that children will encounter other social crises in the future in which their individual behavior will have a strong impact on the health and wellbeing of others. Therefore, it is important for future work to study how children learn and think about similar kinds of novel norms that carry important social consequences.

While our results indicate children viewed the COVID norms like conventional norms, there are a number of alternative possibilities. For one, because we did not include harm-related behaviors, it is possible that participants would have been equally influenced by the descriptive norms for harm behaviors as the COVID and conventional behaviors. We did not include harm behaviors because we were primarily interested in social conventions since they are a key driver of human social behavior and cross-cultural variability, meaning that they constitute an especially interesting area for studying how we acquire novel norms. In contrast, previous work suggests there is relatively less cultural variation in harm-related norms and that our beliefs about them are largely inflexible to normative information (Barrett et al., 2016; Deutchman et al., preprint; Song et al., 1987; Yao & Smetana, 2003). Additionally, from a practical perspective, we were constrained by the number of stimuli we could include in a testing session and including harm behaviors would have presented other challenges (e.g., it would have likely strained credulity to inform children that a majority were doing something truly harmful). Thus, we cannot say with certainty whether children viewed the COVID behaviors more as social conventions or moral norms (e.g., harm-related norms). It is also possible that our results are a consequence of the specific COVID behaviors we selected and that children might have responded differently to other COVID-related norms. The relevance of these chosen norms—double masking and social distancing—likely varied across data collection as vaccines were rolled out nationally and we learned more about disease transmission which shifted public perception of behaviors like double masking. This might suggest that, while these behaviors were initially perceived as moral, they increasingly became viewed as socially conventional over time. Additionally, other factors such as parental political affiliation likely influenced the extent to which participants viewed them as moral or conventional (Gollwitzer et al., preprint).

Our findings also suggest children did not generalize social information to the same extent for all categories of behavior but rather did so selectively depending on the social context, indicating that there are likely meaningful psychological differences in how children view these different norms. Specifically, we found a difference between the conventional and COVID behaviors and the personal preferences, such that participants' injunctive beliefs were largely not influenced by descriptive norm information for independent personal preferences while they were for socially consequential behaviors that could affect other agents. This result aligns with recent work which found that 3.5-year-old children conform more to conventional norms than personal preferences (Li et al., 2021). However, in contrast to our predictions and past research, we observed an effect in the opposite direction for the behavioral intention measure: participants were actually more likely to say they would engage in the preference behavior when it was uncommon than common. Interestingly, this differed from participants' beliefs about what other children would do such that they expected others to be more likely to follow the norm for the preference behaviors when it was common than uncommon. Why were children more likely to report they would engage in the preference behavior when it was uncommon? One possibility is that, because the preference behaviors involved picking a red vs blue pen or green vs orange tray, children simply preferred the less commonly used item (e.g., the blue pen in the strong descriptive norm condition), either because it was more unique or because they inferred there would be more of that resource available. This finding comports with work which finds that children prefer a more abundant resource over one that was scarce despite believing that the scarce resource was more preferred by other children (Smith-Flores et al., 2021).

While this work shines important light on the influence of descriptive norms on injunctive norm beliefs, it also raises some important questions for future work. First, because of constraints imposed with working with young children, we were only able to include two behaviors per vignette type to keep the study to a manageable length. Thus, it is possible that the results found here are an artifact of the specific behaviors chosen and might not generalize more broadly. However, while we acknowledge that we must be careful when generalizing these findings, given the similarities between each behavior within vignette type (see supplement), we think this explanation is unlikely to explain our results. Future work should explore the effect of descriptive norms on injunctive beliefs using a larger set of behavioral stimuli. Second, a limitation of our design is that behavioral intentions were self-reported and thus did not measure actual behavior. There's a large body of work in developmental psychology on the knowledge-behavior gap which suggests that children's knowledge about how they should behave often diverges from how they actually behave (Blake, 2018; Blake et al., 2014). Therefore, children's actual behavior may not be as influenced by descriptive norms as their behavioral intention ratings suggest. That said, previous work has found that descriptive norms influence children's resource allocations which suggests that descriptive norms do influence children's actual behavior to some extent (Liu et al., 2022; McAuliffe, et al., 2017). Future work should further explore how descriptive norms influence children's behavior and the potential role of injunctive beliefs may in mediating that relationship.

An important question in the study of social cognition is how we are able to acquire and enforce social norms, and more specifically, how behavioral regularities in our social environment shape our normative beliefs about how we should behave in different social contexts. We find that by 6 years of age, children are sensitive to social information such that their injunctive norm beliefs, moral evaluations, punishment judgements, and behavioral intentions and expectations were influenced by descriptive norm information about how common a behavior is. Children flexibly tuned their beliefs in response to behavioral regularities depending on the social consequence of the behavior such that the influence of descriptive norms differed for personal preferences compared to conventional behaviors and novel COVID-related health norms in which complying with the norms affects other agents. That children perceive COVID-related norms similarly to conventional norms has implications for future social crises in which acquiring and complying with novel norms can impact the health and wellbeing of others. Overall, these findings showcase the early emerging ability to make normative inferences from descriptive norm information.

6. References

Butler, L. P., Schmidt, M. F., Bürgel, J., & Tomasello, M. (2015). Young children use pedagogical cues to modulate the strength of normative inferences. *British Journal of Developmental Psychology, 33*(4), 476-488.

Bicchieri, C. (2017). Norms in the wild: How to diagnose, measure, and change social norms. Oxford University Press.

Blake, P. R., McAuliffe, K., Corbit, J., Callaghan, T. C., Barry, O., Bowie, A., Kleutsch, L., Kramer, K., Ross, E., Vongsachang, H., Wrangham, R., & Warneken, F. (2015). The ontogeny of fairness in seven societies. *Nature, 528*(7581), 258-261.

Blake, P. R. (2018). Giving what one should: Explanations for the knowledge-behavior gap for altruistic giving. *Current Opinion in Psychology, 20*, 1-5.

Blake, P. R., McAuliffe, K., & Warneken, F. (2014). The developmental origins of fairness: The knowledge–behavior gap. *Trends in Cognitive Sciences, 18*(11), 559-561.

Cialdini, R. B., Reno, R. R., & Kallgren, C. A. (1990). A focus theory of normative conduct: Recycling the concept of norms to reduce littering in public places. *Journal of Personality and Social Psychology, 58*(6), 1015.

DeJesus, J. M., Rhodes, M., & Kinzler, K. D. (2014). Evaluations versus expectations: Children's divergent beliefs about resource distribution. *Cognitive Science, 38*(1), 178-193.

Deutchman, P., Kraft-Todd, G., Young, L., & McAuliffe, K. (Preprint). People update their injunctive norm beliefs and moral judgements after receiving descriptive norm information.

Diesendruck, G., & Markson, L. (2011). Children's assumption of the conventionality of culture. *Child Development Perspectives, 5*(3), 189-195.

Eriksson, K., Strimling, P., & Coultas, J. C. (2015). Bidirectional associations between descriptive and injunctive norms. *Organizational Behavior and Human Decision Processes, 129*, 59-69.

Foster-Hanson, E., Roberts, S. O., Gelman, S. A., & Rhodes, M. (2021). Categories conveyprescriptive information across domains and development. *Journal of Experimental Child Psychology, 212*, 105231.

Foster-Hanson, E., & Lombrozo, T. (2022). How "is" shapes "ought" for folk-biological concepts. *Cognitive Psychology, 139*, 101507.

Gollwitzer, A., Marshall, J., Deutchman, P., Warneken, F., McAuliffe, K. (Preprint). Parent and Community Partisanship Predicts Children's Health Behaviors.

Green P, MacLeod CJ (2016). "simr: an R package for power analysis of generalised linear mixed models by simulation." *Methods in Ecology and Evolution*, *7*(4), 493-498.

Hardecker, S., & Tomasello, M. (2017). From imitation to implementation: How two-and three year-old children learn to enforce social norms. *British Journal of Developmental Psychology, 35*(2), 237-248.

House, B. R. (2018). How do social norms influence prosocial development?. *Current Opinion in Psychology, 20*, 87-91.

House, B. R., & Tomasello, M. (2018). Modeling social norms increasingly influences costly sharing in middle childhood. *Journal of Experimental Child Psychology, 171*, 84-98.

House, B. R., Kanngiesser, P., Barrett, H. C., Broesch, T., Cebioglu, S., Crittenden, A. N., Erut, A., Lew-Levy, Sebastian-Enesco, Smith, A.M., Yilmaz, S. & Silk, J. B. (2020). Universal norm psychology leads to societal diversity in prosocial behaviour and development. *Nature Human Behaviour, 4*(1), 36-44.

Kanngiesser, P., Schäfer, M., Herrmann, E., Zeidler, H., Haun, D., & Tomasello, M. (2022). Children across societies enforce conventional norms but in culturally variable ways. *Proceedings of the National Academy of Sciences, 119*(1), e2112521118.

Kelly, D., & Davis, T. (2018). Social norms and human normative psychology. *Social Philosophy and Policy, 35*(1), 54-76.

Killen, M., Smetana, J. G., & Smetana, J. (2006). Social–cognitive domain theory: Consistencies and variations in children's moral and social judgments. In Handbook of moral development (pp. 137-172). Psychology Press.

Langenhoff, A. F., Dahl, A., & Srinivasan, M. (2022). Preschoolers learn new moral and conventional norms from direct experiences. *Journal of Experimental Child Psychology, 215*, 105322.

Li, L., Britvan, B., & Tomasello, M. (2021). Young children conform more to norms than to preferences. *PloS One*, *16*(5), e0251228.

Lindström, B., Jangard, S., Selbing, I., & Olsson, A. (2018). The role of a "common is moral" heuristic in the stability and change of moral norms. *Journal of Experimental Psychology: General, 147*(2), 228.

Lois, G., & Wessa, M. (2019). Creating sanctioning norms in the lab: The influence of descriptive norms in third-party punishment. *Social Influence, 14*(2), 50-63.

McAuliffe, K., Raihani, N. J., & Dunham, Y. (2017). Children are sensitive to norms of giving. *Cognition, 167*, 151-159.

McGuire, L., Rizzo, M. T., Killen, M., & Rutland, A. (2018). The development of intergroup resource allocation: The role of cooperative and competitive in-group norms. *Developmental Psychology, 54*(8), 1499.

Muthukrishna, M., Morgan, T. J., & Henrich, J. (2016). The when and who of social learning and conformist transmission. *Evolution and Human Behavior, 37*(1), 10-20.

Olson, K. R., & Dweck, C. S. (2008). A blueprint for social cognitive development. *Perspectives on Psychological Science, 3*(3), 193-202.

Olson, K. R., & Dunham, Y. (2010). The development of implicit social cognition. In B. Gawronski & B. K. Payne (Eds.), Handbook of implicit social cognition: Measurement, theory, and applications (pp. 241–254). The Guilford Press.

R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R project.org/.

Rakoczy, H., Warneken, F., & Tomasello, M. (2008). The sources of normativity: young children's awareness of the normative structure of games. *Developmental Psychology, 44*(3), 875.

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*(2), 445-456.

Rakoczy, H., Hamann, K., Warneken, F., & Tomasello, M. (2010). Bigger knows better: Young children selectively learn rule games from adults rather than from peers. *British Journal of Developmental Psychology, 28*(4), 785-798.

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., & 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.

Schmidt, M. F., Butler, L. P., Heinz, J., & Tomasello, M. (2016). Young children see a single action and infer a social norm: Promiscuous normativity in 3-year-olds. *Psychological Science, 27*(10), 1360-1370.

Smetana, J. G. (2013). Moral development: The social domain theory view. In P. D. Zelazo (Ed.), The Oxford handbook of developmental psychology (Vol. 1): Body and mind (pp. 832–863). Oxford University Press.

Smith, C. E., Blake, P. R., & Harris, P. L. (2013). I should but I won't: Why young children endorse norms of fair sharing but do not follow them. *PloS One, 8*(3), e59510.

Smith-Flores, A. S., Applin, J. B., Blake, P. R., & Kibbe, M. M. (2021). Children's understanding of economic demand: A dissociation between inference and choice. *Cognition, 214*, 104747.

Song, M. J., Smetana, J. G., & Kim, S. Y. (1987). Korean children's conceptions of moral and conventional transgressions. *Developmental Psychology, 23*(4), 577-582.

Tooby, J., Cosmides, L., & Barkow, J. (1992). The adapted mind. Evolutionary psychology and the generation of culture. New York: Oxford University Press. Tworek, C. M., & Cimpian, A. (2016). Why do people tend to infer "ought" from "is"? The role of biases in explanation. *Psychological Science, 27*(8), 1109-1122.

Vaish, A., Missana, M., & Tomasello, M. (2011). Three-year-old children intervene in third party moral transgressions. *British Journal of Developmental Psychology, 29*(1), 124-130.

Cucinotta, D., & Vanelli, M. (2020). WHO declares COVID-19 a pandemic. *Acta Biomedica: Atenei Parmensis, 91*(1), 157.

Yau, J., & Smetana, J. G. (2003). Conceptions of moral, social-conventional, and personal events among Chinese preschoolers in Hong Kong. *Child Development, 74*(3), 647–658.

Yoo, H. N., & Smetana, J. G. (2022). Distinctions between moral and conventional judgments from early to middle childhood: A meta-analysis of social domain theory research. *Developmental Psychology*, *58*(5), 874-889.

Yucel, M., & Vaish, A. (2018). Young children tattle to enforce moral norms. *Social Development, 27*(4), 924-936.

5. General Discussion

This dissertation presents four studies that investigate how mentalizing allows people to cooperate and successfully coordinate their behavior. Study 1 presents evidence that people are more likely to cooperate in the TPGG when they have common knowledge because it reduces uncertainty about what whether their group members will also contribute and thus whether they will reach the threshold to receive the public good. Study 2 provides evidence that the ability to use common knowledge to coordinate emerges by 6 years of age, suggesting that common knowledge is likely a relatively early-emerging social-cognitive ability underlying children's cooperative behavior. Study 3 finds that people update their injunctive norm beliefs about what other people approve of based on what is common, and that these beliefs partly underlie the effect of descriptive norms on behavioral intentions. This finding suggests that beliefs about other agents' beliefs play an important role in social norm cognition, a critical ability for coordinating human behavior on a large scale. Lastly, Study 4 finds that children's beliefs about how they should behave are influenced by descriptive norms about what others are doing. This provides some initial evidence that the ability to make inferences about other agents' belief, and use those beliefs to guide behavior, based on what is commonly done emerges by 6 years of age. This finding suggests that the ability to use higher-order beliefs in social norm cognition is early emerging and an important feature of social cognition. Together, these results demonstrate the importance of mentalizing in cooperative behavior and coordination specifically, highlighting two forms of mentalizing—common knowledge and higher-order beliefs—that allow agents to successfully coordinate their behavior by predicting what others think, believe, and do.

There are three main themes that connect these four studies together. The first is that a major driver of cooperative behavior is the belief that others will also cooperate and that we're highly sensitive to cues of this information. The second is the need for an integrative approach to studying cooperative behavior using a mix of vignette-based surveys and economic games across different points in development to better understand the ontogeny of mentalizing in cooperation. The third is that there are large contextual differences in cooperative behavior depending on the kind of behavior in question.

The first theme that emerges across the studies presented here is that an important driver of cooperative behavior is the expectation that others will also cooperate, even across a range of cooperative contexts. In Studies 1 and 2, participants' own decision to cooperate was influenced by their certainty in other agents' cooperative behavior in that when they were more certain that others would cooperate—such as cases in which they had common knowledge—they were more likely to cooperate themselves. While in Studies 3 and 4, participants were more likely to report they would engage in the norm when there was a strong descriptive norm that most others were also engaging in the behavior. These findings align with previous work on the importance of expectations in cooperative behavior (Köll & Quercia, 2021; Pletzer et al., 2018) and more generally, suggest that in most cooperative contexts, promoting the belief that others will pitch in is a motivator of actual cooperative behavior. This has important implications for promoting cooperative behavior such as designing behavioral interventions. For example, communicating that a funding goal is common knowledge (e.g., everyone knows that everyone knows we need $10,000) might reduce uncertainty about others' cooperative behavior and the likelihood of successfully funding the initiative, which in turn might motivate greater contributions.

The second theme is the need to take an integrative approach to studying social cognition. The studies presented here used a complimentary mix of vignette-based surveys and behavioral experiments to study the role of mentalizing in cooperative behavior. Perhaps most importantly, this work highlights the importance of collecting developmental data: investigating the role of mentalizing in cooperative behavior across ontogeny offers insight into how foundational these inferences are in social cognition (Olson & Dweck, 2008; Olson & Dunham, 2010). Cognitive abilities that emerge early in development—such as a sensitivity to normative information (House et al., 2020)—are more likely to be universal and less likely to be shaped by cultural information. By studying when in ontogeny children are able to represent others' beliefs and use that information to guide their cooperative behavior, we can gain insight into how important the ability is for social cognition. Namely, that we find that the ability to use common knowledge to coordinate, or infer others' beliefs from their behavior, is already present by middle childhood suggests that these mentalizing abilities emerge relatively early in ontogeny and are a core feature of social cognition. Additionally, studying these questions using a mix of vignette-based studies and behavioral tasks provides complimentary evidence that is stronger than an individual approach.

The third theme emerging from this work—particularly Studies 3 and 4—is that there are substantial contextual differences in the effect of mentalizing on cooperative behavior which reveal important psychological distinctions between different behaviors. Namely, in Study 3, we found that the effect of descriptive norms on injunctive and moral beliefs varies across behaviors such that information about what others are doing—and thus what they believe one should do—is less influential for harm behaviors and personal preferences than conventional or fairness behaviors. Whereas in Study 4, children were influenced by descriptive norms for conventional behaviors but were largely not influenced for personal preferences, suggesting that children selectively incorporate social information depending on the context. That people made weaker injunctive inferences from others' behavior for personal preferences and harm-related norms than conventional or fairness norms suggests that the role of mentalizing might vary across contexts, such that we rely on it more in certain contexts than others. Specifically, mentalizing might play a more important role in conventional and fairness norms because they are cases in which it is more important to coordinate our behavior—it's more important for our own behavior to know whether other people are likely to cheat on the assignment (because that can affect our own grade) than if they like to wear socks with sandals. This can also inform our understanding of how people conceptualize harms: that people are less sensitive to others' beliefs and behavior for harm behaviors indicates that harms might be more internalized than other behaviors, meaning that we comply with them regardless of what others do or believe.

What do these studies tell us about the social cognition underlying social norms? My findings highlight the central role of mentalizing in norm cognition—we rely on mentalizing to make inferences about others' beliefs and moral judgements from their behavior. Thus, our beliefs about what others know or do represent a major factor influencing norm conformity. Additionally, the findings reported here support the prominent accounts of social norms put forward by Cristina Bicchieri which suggests that a defining feature of social norms is that there are social expectations, meaning we comply with a norm because we believe that others think we should and expect that we do so (2005; 2016). Our results highlight the need for future research to investigate the role of mentalizing when studying social norms and cooperative behavior.

Ongoing and future work

One aim of ongoing work is to further explore the development of social norm cognition—while we now have data to suggest that descriptive norms influence children's injunctive and moral beliefs, it remains unclear whether and how they update their prior beliefs as found in adults. To test this, I'm recruiting 6-9-year-old children in a version of the task used in Study 4 in which children see the vignettes before and after receiving either weak or strong descriptive norm information. We expect to find that, like adults, children update their injunctive norm beliefs and moral evaluations after receiving descriptive norm information and update to a larger extent after receiving the strong than the weak descriptive norm. We also expect that there will be between-vignette differences in updating, such that children should update their beliefs more for positive and negative conventional behaviors than for personal preferences. These findings would provide stronger evidence that by middle-childhood, children can make rich normative inferences from behavioral regularities and that this varies across social contexts.

In another ongoing study, I'm exploring how the source (e.g., peer or adult) and content of a norm (e.g., prosocial or antisocial) interact to influence social norm cognition and behavior. In this study, I'm presenting 6-11-year old children with a series of vignettes with either injunctive norm information from a peer or adult—the baseline condition—or injunctive norm information followed by conflicting descriptive norm information with a different norm type and norm source—the updating condition. For example, in this condition, if children received a prosocial injunctive norm from an adult, they would then immediately receive an antisocial descriptive norm from peers. By manipulating the source and type of norm content, we can better understand which source of normative information is more influential in shaping children's normative beliefs and moral judgements and how that changes across development. We expect to find that children will update their beliefs after receiving a conflicting descriptive norm and that they will be differentially influenced depending on whether the norm is coming from an adult or peer, or whether it's an antisocial or prosocial norm. Specifically, we expect that children will be more influenced by prosocial norms—and less influenced by antisocial norms—and become increasingly sensitive to normative information from their peers as they get older. These findings would shine important light on the forces that influence children's normative beliefs and inform our understanding of age-related changes in norm cognition.

The present work also raises a number of exciting avenues for future research. One important future direction will be to explore the role of common knowledge in social norm cognition. Common knowledge is likely an important feature of social norms—in general, norms are common knowledge within a group (e.g., everyone knows that everyone knows to drive on the right side of the road). Findings from Studies 1 and 2 offer preliminary evidence that this might be the case. Namely, in these studies, I find that, like social norms, common knowledge plays an important role in promoting cooperation by allowing individuals to align their expectations about each other's cooperative behavior. Indeed, one potential hypothesis for the high levels of cooperation found in Study 1 is that participants contributed because they possessed preexisting fairness norms—even in the absence of explicit norm information—that they should pitch in to a group effort or cooperate with fellow Mechanical Turk workers (Almaatouq et al., 2020). If implicit fairness norms did influence contributions in the TPGG, this may suggest that common knowledge plays a similar function for other kinds of normative information. Specifically, common knowledge might play a critical role in social norms by allowing individuals to align their social expectations about how to behave—if you are unsure whether other people know the norm (and that they know that you know, etc.) you cannot be confident they will comply, which in turn might influence your own behavior.

While no work to my knowledge has explicitly investigated common knowledge in norm cognition, other work has studied a similar concept in this context—shared attention. Shared attention is the social cognitive state in which we collectively co-attend to a given stimuli such that we experience it from "our" perspective (sometimes called 'collective attention'; Shteynberg, 2018; Shteynberg et al., 2020; Tomasello et al., 2005). Because shared attention necessarily establishes the co-attended stimulus as common knowledge, this would suggest that common knowledge also plays a role in the acquisition of social norms. Indeed, some researchers even suggest that shared intentionality and collective beliefs are in fact necessary to create and maintain social norms and other cultural knowledge (Tomasello & Carpenter, 2007). Future research should explicitly investigate the role of common knowledge representation in underlying social norm cognition. For example, by manipulating whether normative information is learned under shared attention or not—and thus whether it is or is not common knowledge—we can explore the influence of common knowledge on norm acquisition, compliance, and enforcement. If common knowledge is critical for the effect of social norms, we would expect that people would be more likely to acquire norms, comply with them, and enforce violations of them when that normative information is common knowledge than when it is not.

The present work also highlights the importance of studying other factors that influence conformity. Namely, in Study 3 we found that when examining the role of injunctive beliefs in mediating the effect of descriptive norm information on behavioral intentions, injunctive norm beliefs only explained 26% of the variance. This suggests that while the injunctive inferences we're making from others' behavior influence our own behavioral intentions, there are other factors at play that contribute to our decision to comply with a social norm beyond the social expectations of our group members. For example, imagine that you're waiting at the gate for your flight, and even though the plane is ten minutes from boarding, other passengers have begun to crowd around the entrance to the boarding bridge. As more passengers gather around the entrance, you might feel the need to do so as well, even though you're in group Z and are last to board. Here, you're mindlessly copying the behavior of others without making any inferences about what they approve of or expect of you; in other words, you're conforming to the descriptive norm even in the absence of social expectations.

One reason why descriptive norms might influence behavior outside of their effect on injunctive beliefs is due to a herd (or mob) mentality consisting of unreflective conformity (Raafat et al., 2009). That is, we possess a heuristic to simply copy the behavior and actions of others in a group, particularly in new or uncertain social situations, without necessarily making inferences about what they approve of or expect of others. This is a useful heuristic in many situations in which it is beneficial to make a split-second decision—it's better to start running if you see a group of people running in terror than to wait around to think about why they are running or whether they expect you to join them. Therefore, while our social expectations and injunctive norm beliefs play a central role in the influence of descriptive norms on compliance, they do not solely explain the effect of descriptive norms; other psychological phenomenon, such as herd mentality, likely contribute to the effect of descriptive norms on behavior. While this phenomenon would fall outside the conception of social norms as defined by Bicchieri's account (as it is not a socially conditional preference), it still constitutes a major influence on social behavior that many would colloquially think of as a norm. It will be important for future work to investigate the relationship between herd mentality and social norms moving forward.

Another valuable direction for future work will be to study social norm cognition in a population with mentalizing deficits such as Autism Spectrum Disorder (ASD). Individuals with ASD have been found to have impairments in their theory of mind ability and thus show a reduced ability to reason about others' intentions and emotions (Andreou & Skrimpa, 2020; Boucher, 2012). This is an interesting population to study because, if mentalizing plays an important role in norm cognition, then individuals with ASD should also show deficits in their normative reasoning. Namely, they might struggle to make inferences about others beliefs from what is commonly done or are unable to use common knowledge to guide their cooperative behavior. Indeed, while minimal work has examined this question, the work that has finds circumstantial evidence that ASD contributes to impairments in norm acquisition and compliance (Heerey et al., 2003; Heerey et al., 2005). Studying this population can also shine light on the relationship between common knowledge and ToM, informing our understanding of whether these are distinct abilities for mentalizing as has been proposed (De Freitas et al., 2019) or whether ToM underlies the ability to represent something as common knowledge. Namely, if people with ASD can understand and use common knowledge to guide their behavior but show an impairment in ToM, this would suggest that common knowledge is a distinct cognitive ability from ToM. Whereas if people with ASD cannot understand and use common knowledge to guide their behavior, that might suggest that common knowledge is at least in part supported by ToM networks. Furthermore, if individuals with ASD do show evidence of a deficit in common knowledge representation, then studying social norms with this population might more generally elucidate the role of common knowledge in normative cognition. Namely, if common knowledge is important for social norms, then individuals with ASD will be less able to readily acquire social norms as neuropsychologically typical individuals.

Conclusion

Findings from these studies contribute to a fuller picture of the role of mentalizing in cooperative cognition. Namely, this work provides evidence that common knowledge allows our species to solve coordination problems by reducing uncertainty about the likelihood of others' cooperative behavior. Additionally, this work demonstrates that we make complex inferences about others' beliefs from their behavior and that these higher-order beliefs influence our own norm compliance and enforcement. Altogether, the studies in this dissertation highlight the importance of studying mentalizing in cooperative behavior and social norm cognition.

6. References

Andreou, M., & Skrimpa, V. (2020). Theory of mind deficits and neurophysiological operations in autism spectrum disorders: a review. *Brain Sciences, 10*(6), 393.

Apperly, I. A. (2012). What is "ToM"? Concepts, cognitive processes and individual differences. *Quarterly Journal of Experimental Psychology*, *65*(5), 825-839. https://doi.org/10.1080/17470218.2012.676055

Baltag, A., Moss, L. S., & Solecki, S. (2016). The logic of public announcements, common knowledge, and private suspicions (pp. 773-812). Springer International Publishing.

Boucher, J. (2012). Putting theory of mind in its place: psychological explanations of the socio emotional-communicative impairments in autistic spectrum disorder. *Autism, 16*(3), 226-246.

Bicchieri, C. (2005). *The grammar of society: The nature and dynamics of social norms*. Cambridge University Press.

Bicchieri, C. (2016). *Norms in the wild: How to diagnose, measure, and change social norms*. Oxford University Press.

Böckler, A., & Zwickel, J. (2013). Influences of spontaneous perspective taking on spatial and identity processing of faces. *Social Cognitive and Affective Neuroscience, 8(*7), 735-740.

Bohn, M., & Köymen, B. (2018). Common ground and development. *Child Development Perspectives, 12*(2), 104–108.

De Freitas, J., Thomas, K., DeScioli, P., & Pinker, S. (2019). Common knowledge, coordination, and strategic mentalizing in human social life. Proceedings of the

National Academy of Sciences of the United States of America, 116(28), 13751 13758.

Fang, C., Kimbrough, S. O., Pace, S., Valluri, A., & Zheng, Z. (2002). On adaptive emergence of trust behavior in the game of stag hunt. *Group Decision and Negotiation, 11*, 449-467.

Frith, C. D., & Frith, U. (2006). The neural basis of mentalizing. *Neuron, 50*(4), 531-534.

Gallagher, H. L., & Frith, C. D. (2003). Functional imaging of 'theory of mind'. *Trends in Cognitive Sciences*, *7*(2), 77-83.

Heerey, E. A., Keltner, D., & Capps, L. M. (2003). Making sense of self-conscious emotion: linking theory of mind and emotion in children with autism. Emotion, 3(4), 394.

Heerey, E. A., Capps, L. M., Keltner, D., & Kring, A. M. (2005). Understanding teasing: Lessons from children with autism. *Journal of Abnormal Child Psychology, 33*, 55-68.

Higgs, S. (2015). Social norms and their influence on eating behaviours. *Appetite, 86*, 38-44.

House, B. R., Kanngiesser, P., Barrett, H. C., Broesch, T., Cebioglu, S., Crittenden, A. N., Erut, A., Lew-Levy, Sebastian-Enesco, Smith, A.M., Yilmaz, S. & Silk, J. B. (2020). Universal norm psychology leads to societal diversity in prosocial behaviour and development. *Nature Human Behaviour, 4*(1), 36-44.

Olson, K. R., & Dweck, C. S. (2008). A blueprint for social cognitive development. *Perspectives on Psychological Science, 3*(3), 193-202.

Olson, K. R., & Dunham, Y. (2010). The development of implicit social cognition. In B. Gawronski & B. K. Payne (Eds.), Handbook of implicit social cognition: Measurement, theory, and applications (pp. 241–254). The Guilford Press.

Raafat, R. M., Chater, N., & Frith, C. (2009). Herding in humans. *Trends in Cognitive Sciences*, *13*(10), 420-428.

Rubinstein, A. (1989). The electronic mail game: Strategic behavior under "almost common knowledge." *The American Economic Review, 79*, 385–391.

Schelling, T. C. (1960). The strategy of conflict. Cambridge, MA: Harvard University Press.

Shteynberg, G. (2018). A collective perspective: Shared attention and the mind. *Current Opinion in Psychology*, *23*, 93-97.

Shteynberg, G., Hirsh, J. B., Bentley, R. A., & Garthoff, J. (2020). Shared worlds and shared minds: A theory of collective learning and a psychology of common knowledge *Psychological Review*, *127*(5), 918.

Skyrms, B. (2004). The stag hunt and the evolution of social structure. New York, NY: Cambridge University Press.

Smith, J. R., Louis, W. R., Terry, D. J., Greenaway, K. H., Clarke, M. R., & Cheng, X. (2012). Congruent or conflicted? The impact of injunctive and descriptive norms on environmental intentions. *Journal of Environmental Psychology, 32*(4), 353 361.

Snidal, D. (1985). Coordination versus prisoners' dilemma: Implications for international cooperation and regimes. *American Political Science Review, 79*(4), 923-942.

Thomas, K. A., DeScioli, P., Haque, O. S., & Pinker, S. (2014). The psychology of coordination and common knowledge. *Journal of Personality and Social Psychology*, *107*(4), 657.

Thomas, K. A., DeScioli, P., & Pinker, S. (2018). Common knowledge, coordination, and the logic of self-conscious emotions. *Evolution and Human Behavior*, *39*, 179 190.

Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharingintentions: The origins of cultural cognition. *Behavioral and Brain Sciences*, *28*(5), 675-691.

Tomasello, M., & Carpenter, M. (2007). Shared intentionality. *Developmental Science*, *10*(1),121-125.

Tsoi, L., Hamlin, J. K., Waytz, A., Baron, A. S., & Young, L. L. (2021). A cooperation advantage for theory of mind in children and adults. *Social Cognition, 39*(1), 19-40.

Tsoi, L., & McAuliffe, K. (2020). Individual differences in theory of mind predict inequity aversion in children. *Personality and Social Psychology Bulletin, 46*(4), 559-571.

Young, H. (2007). Social Norms. Department of Economics (University of Oxford).

Young, H. P. (2015). The evolution of social norms. *Economics, 7*(1), 359-387.