

The Development of Language and Morality as Forms of Social Action

by

Leon Li

Department of Psychology and Neuroscience
Duke University

Date: _____

Approved:

Michael Tomasello, Supervisor

Rick Hoyle

Steven Asher

Bahar Köymen

Dissertation submitted in partial fulfillment of
the requirements for the degree of Doctor
of Philosophy in the Department of
Psychology and Neuroscience in the Graduate School
of Duke University

2022

ABSTRACT

The Development of Language and Morality as Forms of Social Action

by

Leon Li

Department of Psychology and Neuroscience
Duke University

Date: _____

Approved:

Michael Tomasello, Supervisor

Rick Hoyle

Steven Asher

Bahar Köymen

An abstract of a dissertation submitted in partial
fulfillment of the requirements for the degree
of Doctor of Philosophy in the Department of
Psychology and Neuroscience in the Graduate School of
Duke University

2022

Copyright by
Leon Li
2022

Abstract

Language and morality are two of the most striking manifestations of human social cognition. Each has been investigated in depth individually, but relatively little research has examined how they are related. To help address this gap, the present dissertation outlines key ways in which language and morality co-evolved during human evolution and co-develop during human ontogeny.

To begin, Chapter 2 provides a theoretical framework for viewing language and morality as interrelated forms of cooperative social action. Both evolved as adaptations for contexts in which collaboration was necessary for survival, and both stem from the more general social cognitive capacity to engage in shared intentionality (i.e., to align, exchange, and interact with others' mental states). Furthermore, language is used for many moral functions (e.g., to initiate, preserve, revise, and act on aspects of morality), which are operative even in young children. Building on the theoretical foundations established in Chapter 2, the next two chapters describe novel empirical studies into specific moral functions of language.

Highlighting the function of language as a means of signaling normativity, Chapter 3 reports that young children conform more to the choices of another person when those choices are framed as socially normative. In this study, 3.5-year-old children helped set up items for a tea party. A confederate, who was either an adult female or a 6-

year-old girl, endorsed various items in terms of either conventional norms (e.g., “For tea parties at Duke, we always use this kind of plate”) or personal preferences (e.g., “For my tea party today, I feel like using this cup”). Children conformed more to the model’s choices when the choices were framed as norms, as opposed to preferences.

Highlighting the influence of linguistically mediated social interactions on children’s moral development, Chapter 4 identifies features of social experiences that are conducive to development. In this study, children from 4 to 5.5 years of age discussed simple moral decisions (how to allocate things between different recipients) with a puppet interlocutor. The puppet (i) either agreed or disagreed with the child’s ideas and (ii) either asked the child to justify themselves or not. Experiences of being disagreed with and experiences of being asked for justification both encouraged children to make fair decisions.

Overall, the chapters illustrate the interconnectedness of language and morality in human development. This work may serve as a helpful basis for further research into how language and morality shape each other.

Dedication

I dedicate this to Durham and its lovely people for being my home these past few years.

Contents

Abstract.....	iv
Dedication.....	vi
List of Tables.....	xi
List of Figures	xii
Acknowledgements	xiii
Chapter 1. Introduction	1
Chapter 2. On the Moral Functions of Language.....	6
2.1 The Evolution of Human Cooperation.....	8
2.2 The Moral Functions of Language	14
2.2.1 Language is Used to Initiate Morality	17
2.2.2 Language is Used to Preserve Morality.....	22
2.2.3 Language is Used to Revise Morality	26
2.2.4 Language is Used to Act on Morality	28
2.3 Conclusion	32
Chapter 3. Young Children Conform More to Norms Than to Preferences.....	37
3.1 Method.....	43
3.1.1 Participants	43
3.1.2 Procedure.....	44
3.1.2.1 Conformity Trials.....	45
3.2 Results.....	48

3.3 Discussion.....	52
Chapter 4. How Social Interactions Contribute to Moral Development.....	60
4.1 Moral Reasoning in Young Children.....	61
4.2 Present Study.....	64
4.2.1 Operationalizing Moral Development.....	65
4.2.2 Hypotheses.....	67
4.3 Method.....	72
4.3.1 Participants.....	72
4.3.2 Procedure.....	73
4.3.2.1 Pre-Training Phase: Warm-Up.....	73
4.3.2.2 Pre-Training Phase: False Belief.....	74
4.3.2.3 Training Phase.....	76
4.3.2.4 Test Phase.....	81
4.3.3 Coding and Reliability.....	84
4.4 Results.....	87
4.4.1 Analyses of Allocations.....	89
4.4.1.1 Training Trials.....	89
4.4.1.2 Distributive Fairness Test Trials.....	91
4.4.1.3 Retributive Fairness Test Trials.....	92
4.4.2 Analyses of Reasoning in Conjunction with Allocations.....	94
4.4.2.1 Correlations Between Allocations and Reasoning.....	95
4.4.2.2 Training Trials.....	95

4.4.2.3 Distributive Fairness Test Trials.....	96
4.4.2.4 Retributive Fairness Test Trials.....	97
4.4.3 Analyses of False Belief.....	97
4.4.3.1 Allocations	98
4.4.3.2 Reasoning in Conjunction With Allocations	98
4.4.4 Ancillary Analyses	100
4.4.4.1 Correlations Between Age and Reasoning.....	101
4.4.4.2 Responses on the First and Second Stories.....	101
4.4.4.3 Examining Children With an Equality Bias	102
4.5 Discussion.....	102
4.5.1 Limitations	106
4.6 Conclusion	107
Chapter 5. Conclusion.....	109
Appendix A. Children’s Initial Preferences and Subsequent Choices	112
Appendix B. Protest Measure.....	114
Method.....	114
Results.....	115
Discussion.....	115
Appendix C. Additional Analyses.....	118
Main Effect of Order	118
Crossing Order and Informant	119
Crossing Order and Endorsement.....	119

Crossing Order, Informant, and Endorsement.....	122
The Adult and Child Informant Conditions	123
References	126
Biography.....	137

List of Tables

Table 1: Propositional language facilitates all aspects of morality and is even necessary for certain aspects of morality.	33
Table 2: Summary of the linear mixed effects model of conformity as predicted by Informant (Child, Adult) and Endorsement (Preference, Norm).	49
Table 3: The scripts of the six stories in the training phase.	77
Table 4: The scripts of the six stories in the test phase.	82
Table 5: The coding scheme for children’s justifications of their allocations.	85
Table 6: Children’s initial preferences and subsequent choices.	112
Table 7: Effects of Order and Endorsement.	121
Table 8: Effect of Endorsement.	124

List of Figures

Figure 1: Ecological changes in foraging contexts led early humans to engage in obligate collaborative foraging.	13
Figure 2: Children conformed more to norms than to preferences.	52
Figure 3: The scripts for how Hedgy responded to the child’s decisions in the four experimental conditions.	80
Figure 4: Children’s allocations to the more deserving recipient on the training trials....	91
Figure 5: Children’s allocations to the more deserving recipient on the distributive fairness test trials.....	92
Figure 6: Children’s allocations of punishment to the more serious transgressor on the final retributive fairness test trial (Story 12).....	94
Figure 7: Children’s allocations to the more deserving recipient in conjunction with providing valid reasoning on the distributive fairness test trials.....	97

Acknowledgements

I am forever grateful to my family, friends, mentors, and colleagues for everything.

Chapter 1. Introduction

Human beings are distinctively cooperative creatures. They inhabit not only the physical world, as other animals do, but also a social reality of their own making. By putting their heads together, humans have been able to co-construct, cooperatively, a rich social world comprised of norms, conventions, and other cultural practices (Tomasello, 2016a, 2019). Two of the most striking manifestations of this cooperative social reality are language and morality. Questions about how they emerged in humans have long been of interest in psychology. Evolutionarily, one important question is how humans came to differ so drastically from other great ape species in their linguistic and moral capabilities. Developmentally, a central question is how human children, who begin life unenculturated, are able to develop into fully linguistic and moral beings. Thirdly, another question of interest is how language and morality interact—how they co-evolved over evolutionary time and how they co-develop during ontogeny.

The present dissertation aims to contribute to these lines of inquiry. To begin, Chapter 2 provides a theoretical framework for viewing language and morality as interrelated forms of social action. Next, the subsequent chapters each describe novel experimental demonstrations of the powerful effects of language—linguistic cues in Chapter 3, linguistically mediated social interactions in Chapter 4—on children’s socially normative inferences, behaviors, judgments, and justifications. The contents of these three research chapters are described in detail below.

Chapter 2 provides a theoretical account of the interrelations between language and morality over evolutionary and ontogenetic timescales. This text was originally published by Li and Tomasello (2021) in the journal *Social Cognition*. Leon Li wrote the paper as an extension of his master's thesis, and Michael Tomasello provided guidance and feedback. The argument of Chapter 2 is that language and morality are both forms of cooperative social action that evolved as adaptations for obligate collaboration. Some 400,000 years ago, humans' hominid predecessors faced contexts in which survival was dependent on cooperation (Tomasello, Melis, Tennie, Wyman, & Herrmann, 2012). Such contexts selected for humans who had the social cognitive capacities to collaborate well, such as the capacity to engage in shared intentionality (i.e., to align, exchange, and interact with others' mental states) and the capacity to sustain common ground (i.e., mutual knowledge).

These social cognitive capacities enabled early humans to construct basic forms of communication (i.e., means of influencing one another's mental states) and social normativity (i.e., expectations about how individuals should treat one another). Later, as human groups scaled up in size and complexity, human communication and social normativity scaled up into modern forms of propositional language and culturally elaborated morality (Li & Tomasello, 2021). In addition to outlining this evolutionary story, Chapter 2 also pinpoints specific moral functions of language—how humans use language to initiate, preserve, revise, and act on various aspects of morality. The

theoretical treatment of language and morality as interrelated forms of social action serves as a foundation for the later chapters of the dissertation.

Chapter 3 describes an experimental study on one moral function of language: the use of language to signal what is socially normative behavior. This text was originally published by Li, Britvan, and Tomasello (2021) in the journal *PLOS ONE*. Leon Li contributed to the study's conceptualization, formal analysis, investigation, manuscript drafting, and manuscript editing. Bari Britvan contributed to the study's conceptualization, investigation, and manuscript editing. Michael Tomasello contributed to the study's conceptualization, supervision, and manuscript editing. This study examined whether subtle linguistic framing could influence young children's inferences about what actions are socially normative. Young children at 3.5 years of age were recruited, as children at this age are still relatively unenculturated but nonetheless linguistically competent enough to comprehend linguistic cues.

In this study, children were invited to help set up items for a tea party along with a confederate, who was either an adult female or a 6-year-old girl. The experimental manipulation was the way in which the confederate described the items that children could choose for the tea party. The confederate endorsed some items in terms of conventional norms (e.g., "For tea parties at Duke, we always use this kind of plate") but endorsed other items in terms of personal preferences (e.g., "For my tea party today, I feel like using this cup"). Children's relative rates of conformity to norms versus

preferences were observed. Overall, this study was advantageous for helping to assess whether subtle linguistic cues could affect the socially normative inferences and behaviors of children even at a considerably young age.

Chapter 4 describes an experimental study on another moral function of language: articulating reasons and justifications about moral issues in the context of joint decision-making. In this study, children from 4 to 5.5 years of age discussed what to do in simple moral scenarios (namely, how to allocate various things between different recipients) with a puppet interlocutor. The aim of this study was to assess which features of social interactions are the most effective at promoting children's moral development. Although there is a broad consensus that moral development is strongly influenced by social interactions, there has been surprisingly little experimental research examining which features of social interactions are the most conducive to development.

To address this gap, an experimental design was employed in the present study. The experimental manipulation was the way in which the puppet interlocutor responded to the child's ideas about what to do. In a factorial design, the puppet (i) either agreed or disagreed with the child's ideas and (ii) either asked the child to justify themselves or not. The impacts of the different features of social interactions (e.g., being disagreed with, being asked for justification) on children's moral development were assessed. Moral development was operationalized in terms of children's abilities to make appropriate allocation decisions and justify their decisions with reference to

common ground values of fairness. Overall, this study was advantageous for helping to clarify which specific features of social interactions effectively promote moral development.

In sum, these three research chapters serve to illustrate the powerful influence of language on moral development. Although language and morality have each been investigated to great extents as individual phenomena, relatively little research has examined their confluence and co-development. This dissertation aims to help address this gap.

Chapter 2. On the Moral Functions of Language

This chapter provides a theoretical account of the interrelations between language and morality over evolutionary and ontogenetic timescales. This text was originally published by Li and Tomasello (2021) in the journal *Social Cognition*. Leon Li wrote the paper as an extension of his master's thesis, and Michael Tomasello provided guidance and feedback.

Among the most distinctive characteristics of humans, as opposed to other animal species, are the human capacities for language and morality. These are clearly separate capacities, but at the same time we may ask if they share some underlying psychological processes—or at least some phylogenetic and/or ontogenetic sources—and whether the two capacities interact with one another in ways that are important for human cognitive and social development. The most well-known theory on this topic argues that language and morality are comparable in that each relies on an innate, computational grammar. Mikhail's (2007) theory of Universal Moral Grammar seeks to describe the computational and representational structures underlying the linguistic and moral judgments that humans make. Of course, linguistic and moral judgments are different, as linguistic judgments concern issues such as comprehensibility, grammaticality, and pragmatic appropriateness, whereas moral judgments concern issues such as fairness, obligation, and fidelity to norms. Mikhail's (2007) theory attempts to formalize these differences in terms of different Chomskyan grammars.

While not denying the importance of the representational aspects of human cognition, we propose that there is another way of relating language and morality that may reveal some interesting psychological commonalities. The key is to recognize that Mikhail's (2007) theory is a cognitively *internalist* account, aiming to describe how humans make private mental judgments about linguistic and moral contents. An alternative perspective focuses not on content but on process and function. More specifically, it focuses on the fact that both language and morality are functionally directed "outwardly" towards influencing other people's mental states: in language, to communicate; in morality, to regulate; in both, to cooperate. This perspective is thus a socially *externalist* account of the nature of the interpersonal interactions—what happens between, not just within, minds—when humans engage with one another linguistically and/or morally.

This functionalist perspective, as we may call it, reflects a different view of human linguistic competence than the standard Chomskyan view. The Chomskyan view focuses on *language as representation*, that is, language as it pertains to cognitive representations, syntax, and internal thought (e.g., Chomsky, 1967; Yang, Crain, Berwick, Chomsky, & Bolhuis, 2017). Another body of theory and research instead focuses on *language as social action*, that is, language as it pertains to speech acts, the exchange and alignment of mental states, and the co-construction of reality (e.g., Bybee & Beckner, 2010; Holtgraves, 2002; Langacker, 2013; Tomasello, 2008). Interestingly, a

similar theoretical division may be found within moral psychology. Beginning with the foundational research of Kohlberg, most of moral psychology has focused on *morality as representation*, so to speak, asking about the nature of moral judgments of right and wrong (e.g., Killen & Dahl, 2018; Kohlberg, 1973; Rizzo, Li, Burkholder, & Killen, 2019; Turiel, 2018). But one can also investigate *morality as social action* by focusing on questions such as how and why individuals help one another, cooperate with one another, co-construct value systems, and regulate one another's behaviors and relationships (DeScioli & Kurzban, 2018; Haidt & Graham, 2007; Rai & Fiske, 2011; Tomasello, 2016a).

To date, there has been little theory or research relating language and morality to one another from the perspective of social action. Our aims here are to specify the ways in which language and morality as social actions relate to one another and, in addition, to specify how language makes possible crucial aspects of the ways that human individuals and social groups initiate, preserve, revise, and act on morality.

2.1 The Evolution of Human Cooperation

The animating idea of our account is that both language and morality are forms of cooperative social interaction. This is fairly uncontroversial in the case of morality, since moral actions are fundamentally cooperative in that they consider the desires and well-being of others. But based on the analyses of Grice (1989) and other theorists (e.g., Clark & Wilkes-Gibbs, 1986), human linguistic communication is fundamentally

cooperative as well. Except in the aberrant cases of lying and deception, the communicator's motive is to inform the recipient of something of interest or relevance to her, and the recipient relies on this expectation of cooperativeness in her comprehension (Wilson & Sperber, 2012). Thus, evolutionarily, the original phenomenon was cooperative interaction. Uniquely human communication—first in natural gestures and then in linguistic conventions—emerged to coordinate and facilitate cooperative interactions (Tomasello, 2008; Zlatev, 2014). And for cooperative interactions to become evolutionarily stable strategies, the participants had to recognize their interdependence with one another and become responsible to one another morally (Tomasello, 2016a).

As for why humans evolved to be so cooperative, the evolutionary story began when individuals started having to rely on one another in new and especially urgent ways. Some 400,000 years ago, our hominid predecessors began foraging for food sources that could only be secured via cooperation (e.g., hunting large animals that individuals could not take down alone). These contexts of obligate collaborative foraging selected for individuals with the skills and motivations to cooperate well with others (Tomasello, Melis, Tennie, Wyman, & Herrmann, 2012). In these contexts, our ancestors evolved capacities for shared intentionality—that is, capacities for aligning, exchanging, and interacting with one another's mental states (Tomasello, 2019). Shared intentionality enabled individuals to construct *common ground*, which is a state of mutual

knowledge between individuals wherein “we both know that we both know” about some shared referent to which we are both attending.

Common ground, in turn, enabled and motivated the development of cooperative communicative systems to facilitate the exchange of information and the coordination of social actions with respect to expectations in common ground (Tomasello, 2019). To that end, early humans engaged in basic forms of cooperative communicative actions, such as informing others of useful knowledge, sharing attitudes with others, and requesting things from others (Tomasello, 2008). We use the term *natural communication* to refer to these early forms of communication. At first, natural communication likely relied on more gestural methods, such as pointing, rather than on the more vocal, propositional kinds of language that we use today (Tomasello, 2008). But common to both natural communication and modern forms of language is the use of cooperative social actions—whether gestures or linguistic utterances—in order to influence and coordinate with others’ mental states (Tomasello, 2003).

In the context of cooperative activities, humans also developed notions of *social normativity*, by which we mean, broadly, our mutual expectations about how individuals ought to treat one another. Early forms of social normativity likely included prosocial concerns about others’ well-being (e.g., sympathy), interdependence concerns (e.g., expectations about how “you” and “I” ought to treat each other in the context of a joint commitment), and basic notions of fairness in resource allocations (Tomasello, 2019;

Vaish & Tomasello, 2014). We characterize these early forms of social normativity as a kind of *natural morality* because individuals are naturally disposed towards these prosocial attitudes independent of what they learn from culturally elaborated norms (Tomasello, 2016a). Natural morality and natural communication were intertwined and mutually supportive. Our prosocial concerns motivated us to engage in helpful acts of communication, such as informing others of knowledge that would be useful to them. Furthermore, we used communication to achieve prosocial goals, such as to coordinate the fulfillment of complementary roles in the context of joint commitments (Tomasello, 2008).

The human cooperative mode of living was adaptive and successful. Over time, human populations scaled up in size, and human modes of cooperation scaled up in complexity as group practices became more elaborated. As cooperation scaled up, the common ground required for sustaining such cooperation scaled up as well. The greater demands of common ground motivated the development, in turn, of more complex communicative systems, leading eventually to the emergence of formal, abstract languages with arbitrary word-to-meaning lexicons as well as syntactic capacities for forming propositional statements. The development of these forms of *culturally elaborated communication*—of which *propositional language* is the prototype—enabled, in turn, the development of more complex forms of social normativity, including what we would call *culturally elaborated morality*. We define culturally elaborated morality as a group's

system of norms, values, and practices that the group prescribes regarding how members of the group should act (Tomasello et al., 2012).

Culturally elaborated morality, in this sense, includes both norms that are moral in the traditional, philosophical sense (e.g., moral norms pertaining to fairness, harm, and rights) as well as norms that are more conventional in nature (e.g., conventional norms pertaining to group identity or group customs, which are not inherently moral concerns). To be clear, we do not mean to advocate a relativistic view that moral norms are equivalent to conventional norms. A vast body of philosophical and psychological literature has indicated that moral norms and conventional norms differ along several dimensions, such as authority dependence or relevance to harm (Turiel, 2014). Here, we mean to use the term “morality” in a broad and functional sense—rather than in the philosophical sense of “morality” that distinguishes morals from conventions—because we mean to describe how communication relates to normative modes of social action generally.

Overall, communication (e.g., “language”) and social normativity (e.g., “morality”) are not monolithic entities; they can each be divided into their respectively more “natural” and more “culturally elaborated” versions (Figure 1). As such, the relation between communication and social normativity is not monolithic either. Thoroughly characterizing the relation between communication and social normativity will require paying attention to the internal complexities of each phenomenon.

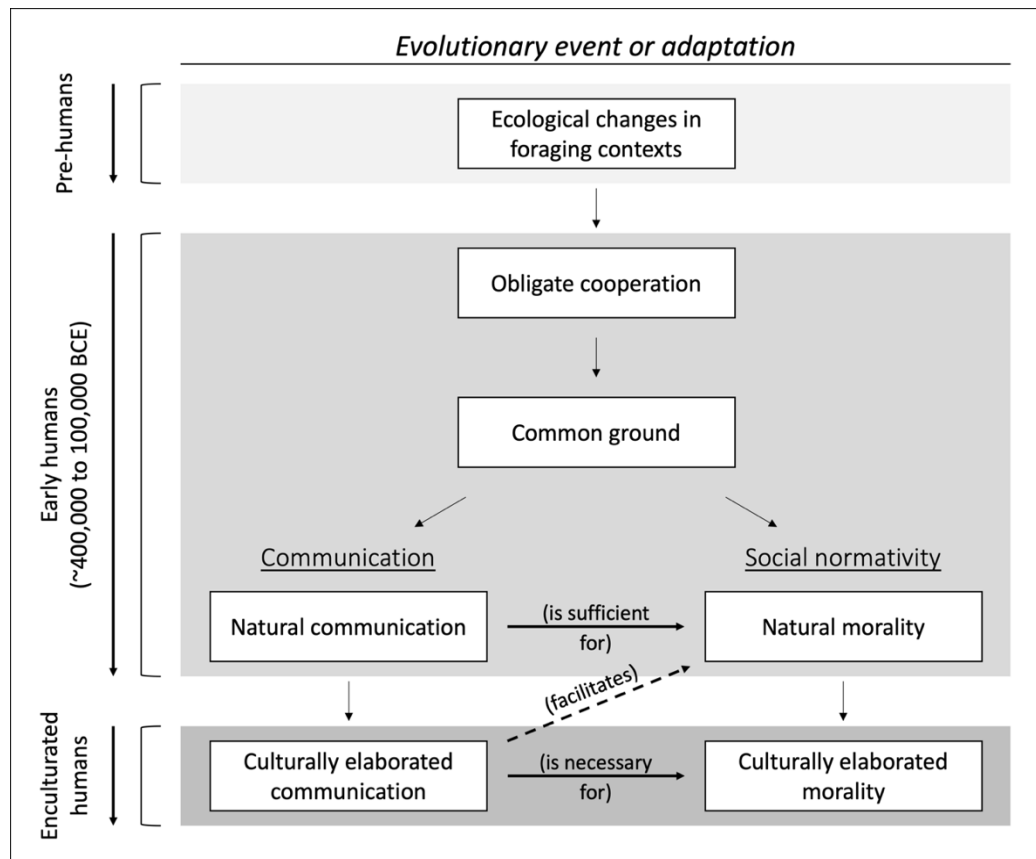


Figure 1: Ecological changes in foraging contexts led early humans to engage in obligate collaborative foraging. In these contexts, humans evolved skills and motivations for cooperation, including the ability to share mutual knowledge in common ground. The ability to share common ground enabled and motivated the development of two key social cognitive adaptations: communication and social normativity. As human cooperation increased in complexity over time, communication and social normativity scaled up from more “natural” versions to more “culturally elaborated” versions.

Interestingly, the progression from the more natural to the more culturally elaborated forms of communication and social normativity manifests not only on the scale of evolution but also on the scale of individual ontogeny (Tomasello & Vaish, 2013; Zlatev, 2014). Taking an ontogenetic view, our account may lend an explanation for the robust empirical finding that children undergo a “normative turn” at age 3 (Killen &

Dahl, 2018; Tomasello, 2018). Children younger than age 3 seem to possess, at most, a second-personal form of prosocial normativity; they do things such as comfort distressed others (Paulus, 2014), provide instrumental help (Warneken & Tomasello, 2006), and form basic joint commitments (Warneken, Chen, & Tomasello, 2006). It is only after age 3, however, that children understand normativity from a group-level, or third-personal, perspective, as evidenced by the enforcement of social norms from a third-party stance (Tomasello & Vaish, 2013). It may be no coincidence that this normative turn at age 3 coincides with the developing linguistic ability to engage in not only personal but also cultural common ground (Tomasello, 2019). If culturally elaborated morality requires cultural common ground, and the communicative skills for sustaining cultural common ground do not emerge until age 3, then it stands to reason that culturally elaborated morality would not emerge until this age either.

2.2 The Moral Functions of Language

Now that we have described the more natural versus the more culturally elaborated versions of communication and social normativity, we are in a position to more precisely specify the relations between their various elements. To begin, a key point is that natural communication is not as effective at conveying complex or abstract information compared to culturally elaborated communication. Imagine, for instance, how hard it would be for a team of builders to construct a large house if they could only use gestures (natural communication) but were unable to communicate using blueprints,

writing, or speech (forms of culturally elaborated communication). There are many proximal reasons for why culturally elaborated communicative systems outperform natural communication (e.g., they have larger vocabularies, are more capable of putting propositions together hierarchically and recursively, and are better at signifying entities that are not present in the immediate situation), but the ultimate, overarching advantage of culturally elaborated communication is, in our view, its greater ability to construct common ground from the bottom-up. Whereas natural communication and culturally elaborated communication may both be effective for referencing mental contents that are already shared in common ground, culturally elaborated communication is far superior at creating and referencing common ground in the absence of previously established common ground. Thus, culturally elaborated communication is much better than natural communication for conveying informationally rich propositions.

Returning to the example of building a large house (and this example is interchangeable with any other case of large-scale coordination on a sophisticated project), it may be in principle possible for the builders to use only natural communication—provided that they already knew, in common ground, about their roles and the required steps in the process. But in cases where common ground is absent or incomplete, natural communication may be insufficient. In contrast, one may easily create common ground using propositional language even if common ground did not previously exist. With propositional language, one may reference complex, abstract

concepts far removed from daily life—and far removed from the ability to be gesturally referenced—simply by saying certain words. Relevant to morality, many moral concepts are indeed the kinds of complex, abstract entities that are hard or impossible to signify gesturally. For instance, how would one signify, in the absence of existing common ground, concepts such as justice, fairness, or rights?

Thus, we argue that natural communication can only support some—but not all—aspects of social normativity. Namely, natural communication can support certain aspects of natural morality, given that common ground expectations for the types of issues covered by natural morality (e.g., issues relating to avoidance of harm) may be robust enough that one may reference these expectations easily. However, natural communication is not sufficient to support the more complex, abstract, and culturally elaborated aspects of morality. In contrast, culturally elaborated communicative systems, such as propositional language, can support and facilitate all aspects of social normativity, spanning both natural morality and culturally elaborated morality. What is more, propositional language is also necessary, not just facilitative, of certain aspects of culturally elaborated morality—for the reason that natural communication is simply inadequate for establishing or navigating common ground when it comes to the complex, abstract contents of culturally elaborated morality (see Figure 1).

In the following sections, we present a (non-exhaustive) catalogue of some key ways in which propositional language facilitates or is necessary for aspects of social

normativity. We support these arguments by referencing selected developmental studies on how even young children, who have just begun to master language, use their communicative skills for socially normative ends. Given that children are less enculturated than adults, children may serve as better, more prototypical examples of how language and morality are related in essence. For ease of reading, we will use the term “language” as a stand-in for culturally elaborated communication in general, as propositional language is the most ubiquitous and widely used form of culturally elaborated communication for moral purposes.

2.2.1 Language is Used to Initiate Morality

To begin, humans use communication to initiate (that is, to create, establish, or otherwise bring into existence) certain aspects of social normativity. A basic form of social normativity is the joint commitment, wherein two individuals agree to do something together (Gilbert, 1990). For simple joint commitments, propositional language is not necessary, as one may imagine initiating a joint commitment using solely gestural means. For instance, one study by Siposova, Tomasello, and Carpenter (2018) showed that even eye contact may be sufficient for generating a sense of commitment. In their study, 5- to 7-year-old children played a “stag hunt” game with an adult partner. Each player could pursue either a small reward individually (a “hare”) or a large reward that could only be obtained if both players decided to pursue it (a “stag”). When, prior to the decision phase, the adult gave the child a “communicative” look with raised

eyebrows (as opposed to a more ambiguous look without raised eyebrows), children chose the stag option more often and also protested more when the adult defected (Siposova et al., 2018).

However, more sophisticated joint commitments may require language. For example, language may be necessary for establishing joint commitments when there is a high risk of defection. To that end, language enables the “making public” that one has committed to an agreement (Pettit, 2018). Without public commitment, one could always potentially feign that one did not understand an agreement in the precise sense intended by the other party. Thus, one could always defect with some semblance of excusability. With language, however, and with the public, explicit declaration of one’s understanding of and commitment to an agreement, it no longer becomes plausibly excusable to defect from the agreement (Pettit, 2018). Children as young as 3 may understand this. One study found that 3-year-olds were more likely to “take leave” from a boring activity (e.g., notify the partner or apologize) if they had made an explicit commitment than if they had not (Gräfenhain, Behne, Carpenter, & Tomasello, 2009). Similarly, another study showed that 3-year-olds were more likely to remain committed to an activity and resist bribes to defect if they had made an explicit commitment than if they had not (Kachel & Tomasello, 2019). Along these lines, as argued by Chwe (2001), rituals and ceremonies serve the function of establishing common ground and commitment within a community. By bringing everyone’s attention together during a

salient event, it is possible to achieve common ground about important issues (e.g., transitions in power, marriages, changes to existing rules), such that pleas of ignorance become inadmissible as excuses for defection.

Propositional language is also necessary for initiating complex types of status functions. As described by Searle (2001), status functions take the form of “X counts as Y in C.” For instance, “green slips printed by the Department of the Treasury” count as “dollar bills” in “the United States.” Status functions are socially normative. Insofar as one wants to be part of group C, one “should” treat X as if it’s Y, just as the other members of C treat X. Language enables groups to assign status functions to objects (e.g., tools, currencies, sacred artifacts) and other kinds of referents, such as people (e.g., roles, responsibilities, group membership) or events (e.g., holidays, ceremonies). Relevant to this, research has shown that generic language (e.g., “pencils are for writing”) aids children’s understanding of the conventionality of tools (Diesendruck & Markson, 2011) and also influences children’s categorization of social groups (Rhodes, Leslie, & Tworek, 2012). The use of language to assign status functions enables groups to construct increasingly complex social realities that build on previously created social realities (e.g., currencies scale up to banks, which scale up to financial systems). Without language, it would be impossible to construct such elaborate cultural institutions.

One early form of normative status assignment to objects is pretense (e.g., pretending that a block of wood “counts” as a bar of soap during a play session).

Children take a normative stance towards pretense status functions. For example, in one study by Rakoczy (2008), a child and an adult pretended that certain objects actually “counted” as other kinds of objects. Then, a puppet arrived and used the objects in a manner that was “incorrect” by the prevailing pretense standards. Children at both 2 and 3 years of age protested against the puppet’s incorrect usage. That is, the children enforced the socially normative expectation to treat the objects as “we agreed” to treat them (Rakoczy, 2008). Plausibly, the pretense abilities that emerge in childhood may be the foundations for the more sophisticated, group-level forms of object normativity that prevail in adult life, such as currencies or sacred artifacts (Wyman, Rakoczy, & Tomasello, 2009).

Language is also necessary, not just facilitative, for specifying the scope of norms. Some norms, in their abstract formulations, are almost certainly universal across cultures. Consider, for instance, the general moral principle to avoid intentionally harming protected parties for no reason (Dijker, 2018; Gert, 2004; Nichols, 2018; Scanlon, 2008; Schein & Gray, 2018). Everyone may agree with this principle, but where people disagree is in the specification of the boundaries relevant to the principle (e.g., specifying who counts as part of the category of “protected party” or what kinds of reasons count as valid justifications for inflicting harm). Language is necessary for achieving common ground about the scope and specification of such principles, even if the prosocial dispositions underlying such principles are preverbally innate to some

extent (e.g., Hamlin & Van de Vondervoort, 2018). Language can also aid discussions about the boundaries between moral and conventional issues as well as the boundaries of the scope of authority, such as when parents and adolescents discuss the scope of parental authority over prudential, conventional, and moral issues (Smetana & Asquith, 1994).

Another way that language gives rise to an aspect of morality, albeit indirectly, is that language gives rise to new categories of moral transgressions, such as lying, promise-breaking, and cheating. To be clear, we are not saying that propositional language is required for deceiving others or forming agreements with others; it is possible to deceive others or form agreements with others using gestural communication alone. However, insofar as one defines lying as, specifically, using propositional language to deceive, and insofar as one defines promising as, specifically, using propositional language to make an explicit agreement, then propositional language is indeed necessary for lying and promise-breaking to occur. As shown by Kanngiesser, Köymen, and Tomasello (2017), children as young as 3 understand that promising something creates an obligation to fulfill what one has promised. In their study, 3-year-olds protested more against a puppet who did not share resources if the puppet had promised to share than if the puppet had not promised to share; tellingly, some of the children's expressions of protest made reference to the fact that the puppet had made a promise (Kanngiesser et al., 2017). Additionally, cheating is a moral transgression

insofar as one is behaving in a way that violates a previously specified set of rules; this set of rules would presumably have required language to have been established.

2.2.2 Language is Used to Preserve Morality

Norms and normative practices are not always enduring. They face three kinds of threats to their existence, so to speak, all of which may be counteracted with communication (indeed, the need to counteract such threats was likely an important evolutionary pressure that contributed to the development of communication in the first place; Pettit, 2018).

The first threat is the threat of *ignorance*. A norm may cease to exist—or may not even begin to exist in the first place—if people do not know about the norm. The threat of ignorance is counteracted by teaching norms (i.e., bringing novices into the fold of common ground expectations) and codifying norms (i.e., preserving norms in common ground). Prototypically, teaching is thought of in terms of adult instructors teaching child learners, such as when adults talk to children about moral and conventional issues (Killen, Breton, Ferguson, & Handler, 1994). However, recent research has shown that children can also be teachers of norms. For instance, 3-year-olds in a study by Kachel, Svetlova, and Tomasello (2018) taught a confederate how to play a game when they perceived that the confederate lacked knowledge about the game. Notably, the children engaged in more teaching when the confederate appeared ignorant about how to play the game as opposed to unwilling to play the game, which speaks to the children's

ability to assess their common ground with their partners (Kachel et al., 2018). Other research has shown that 5-year-olds can not only invent a new game but can also teach a newly created game to newcomers (Hardecker, Schmidt, & Tomasello, 2017).

Propositional language is not necessary for teaching all aspects of morality—one may plausibly teach some basic norms using gestural means—but it is difficult to imagine how inventing new norms with peers could occur without language. For sure, propositional language is necessary for teaching and codifying more complex norms (e.g., legal codes or religious rules).

A second threat to norms is the threat of *defection*. Some norms, such as those pertaining to situations in which one's selfish interests conflict with group interests, may be unstable due to a high risk of defection. Namely, a norm may be unstable if one's compliance to the norm is conditional upon others' compliance, and the extent to which others can be trusted to comply with the norm is itself uncertain (Bicchieri, 2006; Sripada, 2005). Norm enforcement (e.g., the public punishment of defectors) counteracts this threat by making known in common ground that the norm is still in effect and still being enforced—thereby encouraging conditional compliance. Language may not be necessary for public norm enforcement in every case, but it certainly helps. In addition, language can be very helpful for making threats against defectors, which is another method in addition to punishment for deterring defection (Schelling, 1980).

A third threat to norms is the threat of *unjustifiability*. People do not wish to follow norms that they consider to be unfair, unreasonable, or otherwise unjustified. Language counteracts this threat by enabling individuals and institutions to bolster norms via justification. Research has investigated how children produce, consider, and discuss reasons and justifications interpersonally in both moral and non-moral domains. For instance, in one study by Köymen, Rosenbaum, & Tomasello (2014), dyads of children who were 3 or 5 years old were asked to place toy items in a toy zoo. The items were either conventional for a zoo setting (e.g., a polar bear toy, which would conventionally go in the ice rink) or unconventional (e.g., a piano toy, which had no conventional location). Even 3-year-olds made the warrant explicit (i.e., justified their decisions with respect to premises) more for unconventional items than for conventional items, reflecting their understanding of the need to explain actions with respect to shared expectations in common ground.

A follow-up study by Mammen, Köymen, and Tomasello (2018) similarly elicited reasoning from dyads of 3-year-olds and 5-year-olds but this time about conventional and moral issues. In their study, children justified the punishment of transgressors who violated either social or moral rules. When justifying the punishment of a transgressor who violated a social rule, children tended to reference the rule. But when justifying the punishment of a transgressor who violated a moral rule, the children tended to refer simply to the transgressor's action (without referring to the rule prohibiting such

actions). Mammen et al. (2018) interpreted these findings to mean that young children recognize that others share moral values with them in common ground; the logic is that one does not need to explicitly reference common ground values when giving reasons and justifications because one already knows that the other person is viewing the situation with the same values in mind. Relatedly, classic research by social domain theorists has shown that young children justify normative judgments, such as by saying that moral violations are wrong because they cause harm (Smetana, Jambon, & Ball, 2014). However, this line of social domain research has mostly focused on children's reasoning in the sense of children's responses to experimenter-generated questions about hypothetical scenarios, as opposed to children's reason-giving during naturalistic interactions with other people.

Moreover, on the historical scale of societal change, language sustains the reasoning and justification processes that select, in a manner akin to natural selection, the norms that are the most "adaptive" in the sense of being the most justifiable and acceptable (DeScioli & Kurzban, 2018). Of course, we do not intend to make the unrealistic claim that fair norms have always triumphed over unfair norms; history is full of examples of powerful people committing moral transgressions and imposing unfair norms on others. However, it may still be worth mentioning that the general historical trend does seem to point towards an expanding adoption of fairer and more inclusive norms in societies worldwide as well as a corresponding reduction in the less

inclusive norms that people consider to be less justified (Acemoglu & Robinson, 2012; Pinker, 2012). This historical process of iterative norm selection facilitated by reasoning and justification—what one might call “moral progress”—requires propositional language to occur, given that discourse about societal laws occurs at a level of abstraction that exceeds what could be referenced with non-propositional communication.

2.2.3 Language is Used to Revise Morality

Normative beliefs and practices are not fixed; they change over time (Prinz, 2018). We often use language to enact these changes. For instance, we use propositional language to revise joint commitments, as in the examples of the children in the study by Gräfenhain et al. (2009) who took leave when exiting joint activities. Propositional language may not always be needed for revising joint commitments, as one may imagine that gestural communication may be sufficient for revising joint activities in some cases (e.g., when taking a walk with a friend, one may change the direction of the walk simply by pointing to a new direction).

However, for revising aspects of social normativity that are more sophisticated than simple joint commitments, propositional language is indeed necessary. For instance, it is difficult to envision how complex status functions or rules (that were themselves brought into being only through language) could be revised without using propositional language. One example of how verbal language facilitates norm revision

was provided by Grocke, Rossano, and Tomasello (2015), who taught 5-year-old children how to allocate resources among themselves using a wheel of fortune toy. Groups of children who were assigned a fair wheel accepted the procedure, but groups of children who were assigned an unfair wheel often attempted to rectify the unfairness by changing the rules of the game (Grocke et al., 2015). It is hard to imagine how the children could have discussed and proposed changes with reference to the propositional rules of the game without using propositional language themselves.

In addition, specifying adjustments to the scope of norms (e.g., deciding whether to include or exclude certain outgroups from the category of protected parties that members of one's cultural group must not intentionally harm) requires language. For instance, it is difficult to envision how governments could specify laws about naturalization (e.g., who counts as a citizen?) without propositional language, given the complexities of such situations. It is also difficult to envision how individuals could engage in meta-ethical discourse about morality (e.g., discussing whether two people who disagree about a norm could "both be right") without using propositional language. With propositional language, however, even children as young as 4 years of age can respond to questions about meta-ethics, with children around 9 years of age displaying an understanding that two parties who disagree about a norm could "both be right" in some cases (Schmidt, Gonzalez-Cabrera, & Tomasello, 2017). Finally, language is required to modify complex social realities and norms, such as cultural institutions

and laws. Often, cultures even have codified language games for revising complex norms (e.g., the legislative processes for overturning laws or adding constitutional amendments).

2.2.4 Language is Used to Act on Morality

Humans do not only use language to initiate, preserve, and revise aspects of morality—humans also use language to “act on” morality. We mean that humans “act on” morality in the sense that humans act with reference to common ground expectations about how individuals ought to treat one another. For instance, we reference common ground expectations when we make moral judgments. Whereas cognitively internalist views have focused on how moral judgments reflect innate, unconscious, and private computational processes (e.g., Mikhail, 2007), our externalist emphasis is on how people make moral judgments with reference to what they have co-constructed as right and wrong in their common ground expectations. This is not to deny that we do, as individuals, engage in cognitively internal processes such as computing private moral judgments about events (Mikhail, 2007). Nor do we deny that people do not always follow norms but sometimes actively oppose them (indeed, even opposing existing norms requires making reference to what one is opposing). Rather, our intention here is simply to outline how humans make reference to moral contents in meaningful ways.

Humans make reference to norms when they express moral judgments, convey reactive attitudes (e.g., gratitude, resentment), and protest norm violations, to name a few examples. A wealth of developmental research has found that young children express moral judgments interpersonally to observers (Rakoczy & Schmidt, 2013; Tomasello & Vaish, 2013). For instance, children protest—from a third-party stance—against violations of moral as well as conventional norms (Schmidt, Rakoczy, & Tomasello, 2012). The fact that children protest from a third-party stance against norm violations that do not personally involve them is significant, as it means that their concern for upholding norms goes beyond a purely selfish motive to protect one’s own self-interest. Evolutionarily, the capacities to express reactive attitudes and moral judgments were likely adaptive for regulating relationships and group functioning. In fact, some theorists, such as Rai and Fiske (2011), have argued that the primary function of morality is indeed to help people regulate social relationships.

Here, in focusing on the expression of moral judgments, we emphasize that moral judgments are more than just private mental events. As noted by Sinnott-Armstrong (2018), an ecologically valid conception of moral judgment is not restricted to one’s private consideration of the moral status of an act but also encompasses the public expression (i.e., the *verdict*) of one’s judgment. Moral judgments and reactive attitudes would not be very useful, after all, if we only kept them to ourselves, judging privately

in our own heads whether certain actions are right or wrong. They are useful precisely because we express them and use them to influence how others behave.

As for the expression of moral judgments, propositional language is not necessary in all cases. One may express moral disapproval with a gesture, a non-propositional vocalization, or even a well-placed facial expression. Language is necessary, however, for acting on norms in more sophisticated ways than expressing one-off moral judgments. Gossip, for instance, requires propositional language. Here, we are envisioning gossip not merely as an adolescent pastime but rather as a morally relevant practice. Gossip involves making reference to norms for justifying one's attitude about whomever one is discussing, and gossip can even be an important form of third-party punishment that facilitates partner selection (Tomasello, 2019). Indeed, one study found that children around 5 to 7 years of age consider gossip information when choosing collaborative partners (Haux, Engelmann, Herrmann, & Tomasello 2017). A developmentally early form of gossip—tattling—was investigated by Vaish, Missana, and Tomasello (2011), who found that children as young as 3 tattle on moral transgressors to others.

Furthermore, we use propositional language for the interpersonal practice of reason-giving, wherein we request, exchange, and co-construct reasons and justifications for the beliefs and actions of ourselves and others. Here, we follow Mercier and Sperber (2011) in conceptualizing reasoning from an externalist perspective. In their landmark

paper, Mercier and Sperber (2011) argued that reasoning is functionally oriented not just for introspecting about one's own beliefs but also for interacting with other people's beliefs. For one, we use reasoning to evaluate the reasons and justifications that others employ when they are trying to convince us of things; that is, we practice what is known as epistemic vigilance. In turn, we also use reasoning to refine the reasons and justifications that we employ when we are trying to convince others of things. It is our hope, for example, as the authors of this paper, that our reasoning has convinced you, the reader, of at least some things.

As for how reasoning relates to morality, researchers have described how we engage in morally relevant reasoning behaviors, such as persuading others to take our side in disputes (DeScioli & Kurzban, 2018), advocating for norms that benefit us (Gibbard, 1990; DeScioli & Kurzban, 2018), and responding to others' moral judgments of our behaviors with excuses and justifications (Tomasello, 2019). Relatedly, recent developmental research has found that young children justify the punishment of transgressors by referencing rules, which speaks to their understanding that seemingly antisocial actions (e.g., punishing) require justification (Mammen et al., 2018). Another recent study found that 4- and 6-year-olds were more likely to justify their views about moral dilemmas when discussing them with their peers than when discussing them with their mothers (Mammen, Köymen, & Tomasello, 2019). This may reflect a recognition that peers, unlike mothers, are not epistemically advantaged over one's self;

as such, the peers have a greater need to back up their views with reasoning and may also be more amenable to being persuaded by one's own reasoning (Mammen et al., 2019). Altogether, moral reason-giving in the externalist sense remains an underexplored topic in developmental moral psychology, and much more work remains to be done. The overall impression that we hope to have provided is that morality is far from limited to the formation of private moral judgments. We act on morality interpersonally in the many ways that we try to persuade, influence, and regulate others with reference to the morality that we share with them in common ground.

2.3 Conclusion

We have offered here a broad, externalist view of the moral functions of language. Humans use language to practice morality in several respects. In particular, humans use language to (i) initiate, (ii) preserve, (iii) revise, and (iv) act on various aspects of morality (Table 1). For some of these aspects of morality, language plays a facilitative but not strictly necessary role. Other aspects of morality, however, require propositional language to function. Given the many links shown in Table 1 between language—itsself a significant domain of human cognition and social action —and morality, it is clear that morality is a major hub of social cognition, as described by the present special issue of *Social Cognition* that is aptly titled *Morality as a hub: Connections within and beyond social cognition*.

Table 1: Propositional language facilitates all aspects of morality and is even necessary for certain aspects of morality.

	Language facilitates these aspects of morality.	Language is necessary for these aspects of morality.
Initiate morality.	<ul style="list-style-type: none"> • Form joint commitments. 	<ul style="list-style-type: none"> • Make one’s commitment public. • Assign status functions. • Establish social realities. • Specify the scope of norms.
Preserve morality.	<ul style="list-style-type: none"> • Teach norms. • Deter defection (e.g., via threats or public norm enforcement). 	<ul style="list-style-type: none"> • Codify complex norms. • Bolster norms via justification.
Revise morality.	<ul style="list-style-type: none"> • Change joint commitments. 	<ul style="list-style-type: none"> • Reassign status functions. • Adjust the scope of norms. • Modify social realities.
Act on morality.	<ul style="list-style-type: none"> • Express moral judgments. • Convey reactive attitudes. • Protest against norm violations. 	<ul style="list-style-type: none"> • Gossip. • Engage in moral justification and reasoning.

Morality is a major hub of human social life in the socially externalist sense.

Many of the things we do that involve other people, such as forming joint commitments, teaching, or exchanging reasons and justifications, we do with reference to socially co-constructed value systems—with reference, that is, to morality. Thus, viewing morality as a hub of human social cognition is a generative perspective that can motivate much future research. Language is one phenomenon that permeates the hub of human

morality, but several other key phenomena are also intimately interconnected with morality, such as emotion, identity, group dynamics, and culture (Nadelhoffer, Nahmias, & Nichols, 2010). It will also be interesting for future research to examine how language permeates morality's many interconnections with these other phenomena.

Our framework may motivate future research, both experimental and observational, on the specific ways in which children use language to achieve moral ends. Experimentally, a straightforward direction for research is to present children with morally relevant challenges and then examine how they approach these problems with or without access to language. If, as we have argued, language is necessary for certain aspects of morality but not necessary for other aspects, then children would fail at certain kinds of moral tasks when they cannot use language but still be able to succeed on other kinds of moral tasks even without language. Observationally, it would also be interesting to examine the kinds of language that children use in their naturalistic peer interactions involving moral concerns; one such observation was reported by Killen and Cords (2002), who found that children navigate peer interactions with sophisticated discourse strategies, such as threatening, bargaining, compromising, giving moral justifications, and providing collaborative suggestions.

Overall, our aim has not been to deny the importance of the cognitively internalist paradigm in moral psychology. It is undeniably true and fascinating that we form private moral judgments via conscious and unconscious pathways. Furthermore,

an exciting direction for future research will be to integrate the internalist and externalist accounts of moral cognition. Researchers have already made promising progress in this direction. For instance, researchers have begun to ground elements of interpersonal discourse that were traditionally believed to arise from pragmatic inferences rather than from internal, grammar-based computations (e.g., irony, implicature, metaphor) in plausible cognitive mechanisms of probabilistic inference (Goodman & Frank, 2016). This approach, which is known as the rational speech act model, may fruitfully inform research examining how people understand moral types of speech acts, such as declarations of joint commitments or expressions of moral judgments. Researchers have also referenced probabilistic inference as a potential cognitive mechanism underlying how individuals learn the scope of normative rules (Nichols, Kumar, Lopez, Ayars, & Chan, 2016) as well as how individuals learn from pedagogical instruction in general (Shafto, Goodman, & Griffiths, 2014). These theoretical developments hold much promise for the overarching endeavor to provide a complete, integrated account of human moral cognition.

To conclude, our aim has been to shed light on the underexplored topic of how humans construct morality between—and not just within—minds with the help of language. In a telling analogy, Mikhail (2007) likened individuals to “intuitive lawyers who possess a natural readiness to compute mental representations of human acts in legally cognizable terms” (p. 145). But far from working solely within the realm of

private moral intuitions, lawyers make extensive references to legal precedents and the nuanced facts of situations when deliberating about laws and verdicts. Indeed, lawyers often engage in reasoning processes precisely for the purpose of persuading others to share their views. Lawyers, then, are actually among the best examples of how humans use language to interpersonally construct, preserve, revise, and act on moral norms in common ground. To build on Mikhail's (2007) analogy, we humans are indeed like intuitive lawyers, for we are intuitively disposed, given our uniquely human skills and motivations for shared intentionality, to construct morality with others and to guide our lives with reference to what we have constructed.

Chapter 3. Young Children Conform More to Norms Than to Preferences

This chapter describes an experimental study on one moral function of language: the use of language to signal what is socially normative behavior. This text was originally published by Li, Britvan, and Tomasello (2021) in the journal *PLOS ONE*. Leon Li contributed to the study's conceptualization, formal analysis, investigation, manuscript drafting, and manuscript editing. Bari Britvan contributed to the study's conceptualization, investigation, and manuscript editing. Michael Tomasello contributed to the study's conceptualization, supervision, and manuscript editing.

To become functioning members of a culture, young children must learn about not only physical reality but also social reality. This poses a considerable learning challenge, as many aspects of social reality, such as norms, conventions, and rituals, are causally opaque with no obvious instrumental functions (Clegg & Legare, 2016a, 2016b; Kenward, 2012; Legare & Nielsen, 2015). Still, despite their causal opacity, young children need to learn and perform them for reasons such as affiliating and identifying with members of their culture (Keupp, Behne, & Rakoczy, 2013; Legare & Nielsen, 2015). Along these lines, several researchers have advanced an intriguing hypothesis: that young children are motivated to act conventionally (Legare & Nielsen, 2015; McGuigan & Robertson, 2015; Schmidt, Rakoczy, & Tomasello, 2011; Tomasello, 2016b). The hypothesis, to put it more concretely, is that when young children perceive that a certain

action is conventional within their cultural group, they will be motivated to perform that action simply out of a desire to act conventionally— independent of other possible motives for performing that action.

This hypothesis is significant because it helps explain how human groups are able to preserve and transmit cultural practices over generations. Cultural practices would not persist, after all, if younger generations were not motivated to acquire them. To date, the hypothesis that young children are motivated to act conventionally has received support from several empirical studies. In this paper, we aim to add to this literature by taking a closer look at whether young children’s motivation to perform conventional actions is indeed independent of other possible motives for performing such actions. This inquiry is warranted because, as we describe below, previous studies on this issue did not completely rule out other possible explanations for why children may be motivated to perform conventional actions.

On the basis of previous studies, researchers have argued that children interpret certain social cues to be indications that actions are conventional and thus important to adopt (Legare & Nielsen, 2015). Intentionality has been posited to be one such cue. Even before two years of age, children imitate intentional actions more often than they imitate unintentional actions (Carpenter, Akhtar, & Tomasello, 1998). Additionally, children who observe an actor performing an action intentionally, as opposed to unintentionally, protest more when a puppet performs the action in a different way (Schmidt, Butler,

Heinz, & Tomasello, 2016). Children also protest when actors omit causally unnecessary steps of actions that the children previously saw demonstrated with the needless steps, even when the children know that the steps are unnecessary (Kenward, 2012). Both of these protest findings (Kenward, 2012; Schmidt et al., 2016) suggest that children consider others' intentional actions to be representative of the socially normative way to act. In addition, the conduct of a majority has been posited to be another cue of conventionality. When observing actors operate a device to obtain a reward, children conform more to a method used by multiple individuals than to a method used by only one individual (Haun, Rekers, & Tomasello, 2014). By "conform," it was meant that the children actually switched from their own method of operating the device to adopt the newly observed method, potentially out of a desire to be like the group (Haun et al., 2014).

However, children may be motivated to adopt actions that they see others performing intentionally or in a majority for reasons other than seeking to act conventionally per se. Regarding intentionality cues, children may potentially seek to imitate another person's intentional actions out of a desire to affiliate with the person individually. From an early age, children imitate others' actions as a means of socially bonding with others, such as in the context of preverbal protoconversations (Carpenter, Uebel, & Tomasello, 2013; Tomasello & Gonzalez-Cabrera, 2017). Social bonding is more likely to be achieved by imitating the intentional, as opposed to unintentional, actions of

others. As such, even if children adopt others' intentional actions more often than they adopt others' unintentional actions, this could be due to a desire to affiliate with others, not necessarily a desire to act conventionally.

As for majority cues, the children in Haun et al. (2014) may have inferred that the method used by multiple individuals was more instrumentally effective at obtaining the desired reward compared to the method used by only one individual. After all, perhaps the very reason the former method was more widely used than the latter method was because it was more effective at obtaining rewards. As such, the children may have adopted the method used by the majority not because they wanted to act conventionally but simply because they wanted to increase their chances of obtaining the reward.

Overall, previous studies did not rule out possible alternative interpretations for why children may follow intentionality or majority cues. Children may follow these cues not because they seek to act conventionally per se but rather because they seek to affiliate with others or obtain rewards. These issues may be addressed by modeling actions with no instrumental functions and employing linguistic cues, not intentionality or majority cues, to signal conventionality.

One straightforward way to linguistically signal conventionality is to state a rule (Zhao & Kushnir, 2018). But rules may not be an ideal operationalization of conventionality because the source of a rule's normative force may be unclear to children. It has been argued that norms have two aspects: generality and force (Rakoczy

& Schmidt, 2013). Whereas generality refers to a norm's widespread applicability to all the members of the group, force refers to group members' desire that a norm be followed and their willingness to enforce the norm on others. Young children may sometimes experience adults imposing arbitrary rules seemingly based only on their own discretion (e.g., "Does Mom want me to clean my room because cleanliness is a general expectation or only because she herself likes it clean?"). As such, rules may sometimes appear to children to have force stemming not from conventionality but rather from the authority of individual adults.

In recent studies using more subtle linguistic cues, children imitated an actor's method of making a necklace with higher fidelity when the activity was linguistically framed as conventional than when it was framed as instrumental (Clegg & Legare, 2016a, 2016b). However, these studies still used an instrumental context (necklace making) and only examined imitation, not conformity, since the children did not have a prior method that the linguistic framing overrode. In our study, therefore, we first assessed children's preferences and then examined whether they would conform to another person's different choice—and whether the conformity would be greater when the choice was linguistically framed as a conventional norm than when it was framed as a personal preference.

Children's ability to distinguish between norms and preferences has been investigated in previous research. One study examined children's relative consideration

of information about rules versus information about others' preferences when predicting others' behaviors and mental states (Kalish & Shiverick, 2004). In this study, children from 4 to 5 years of age weighed information about rules more highly, whereas children from 7 to 8 years of age weighed information about preferences more highly (Kalish & Shiverick, 2004). Another study found that children improved with age at distinguishing between group norms and their own preferences (Killen, Rutland, Abrams, Mulvey, & Hitti, 2013). But this study examined older children (9-year-olds and 13-year-olds) and also focused only on children's judgments of hypothetical stories, not their behaviors (Killen et al., 2013). We aimed to recruit the youngest children who would still be linguistically competent enough to comprehend our linguistic cues. The age of 3.5 seemed suitable, as this was approximately the earliest age at which conventional linguistic framing had been shown to have an effect in previous research (Clegg & Legare, 2016b).

We invited 3.5-year-olds to help set up a pretend tea party, a context without instrumental aims, and varied whether an informant endorsed tea party items, such as cups, in terms of either norms or preferences. In the norm condition, we avoided using prescriptive cues of normative force (e.g., "one should use this cup" or "the rule is to . . ."), instead relying on descriptive cues of generality (e.g., "we always use . . ."). This enabled more confidence that children's conformity to norms represented a respect for conventionality, not just a respect for the force stemming from the authority of the

messenger. Preferences were chosen as a control to norms because they may invoke a motivation to conform to affiliate with the informant individually but not necessarily a motivation to act conventionally. Thus, if children conform more to norms than to preferences, this would imply that children seek to act conventionally above and beyond merely seeking to affiliate with the informant individually.

One methodological concern was that if we used an adult model, as most other studies have done, then children might only conform out of a deference to adult authority, not out of a genuine respect for conventionality. To address this concern, we used models of two ages. For some children, the informant who expressed norms and preferences was an adult, whereas for other children, the informant was a 6-year-old child. By using these two models, we could test whether our hypothesized effect (greater conformity to norms than to preferences) would hold even when children did not perceive the model as having authority.

3.1 Method

3.1.1 Participants

Participants were 3.5-year-olds ($N = 104$, M age = 42 months, $SD = 2$, range: 39 to 45; 53 girls) from the Southeastern United States and were recruited via phone calls to parents in our university's database of local birth records. Because we were employing a newly invented procedure, which had never been used in previous research, we had no basis for estimating an effect size or the sample size needed to detect it. However, a

general rule of thumb for factorial designs, such as the 2 x 2 design used in this study, is to obtain at least 24 participants per cell. Our recruitment efforts enabled us to finish data collection with 26 participants per cell.

Participants' families were mostly white (75% white, 9% black, 2% Asian, 14% biracial or other) and middle-class (over 75% had family incomes exceeding \$60,000). Additional children were recruited but excluded from the final sample due to procedural error (9), parental influence (4), lack of English (2), insufficient age (1), the child being excessively distracted and not focused on the activity (1), the child misunderstanding the activity (2), or the child not engaging in the activity and thus not providing any usable data (2). Children were given a toy, book, or T-shirt for participating. This study was approved by the Institutional Review Board of Duke University on March 23, 2018. The procedure was conducted with the written and informed consent of the parents or guardians of the minors and in accordance with all applicable ethical and legal rules concerning psychological research in the United States of America.

3.1.2 Procedure

To begin, the child warmed up with two experimenters (the host, who was either a man or a woman, and the adult informant, who was always a woman) in a greeting room with toys. Once the child seemed comfortable, the two experimenters brought the child and their parent to the tea party room. Here, the host labeled the child an ingroup

member (by giving the child a blue sticker, which the host and informants also wore, and saying: “We are Duke!”), invited the child to help set up the tea party for an upcoming guest, and told the child that another tea party was occurring in another room. The host then pointed out a laptop that they could use to talk to people in the other room. Next, the adult informant left the room to presumably go to the other room, and the host then pretended to video chat with the adult informant on the laptop. This initial phase, which was meant to convince the child that the video chat was live and real, showed the adult informant with a 6-year-old girl (the child informant) in a similar playroom. In reality, all the footage shown on the laptop was prerecorded. The host then briefly played with a ball with the child before proceeding to the conformity trials.

3.1.2.1 Conformity Trials

The tea party room had 4 low shelves, each containing 4 options and 4 instances of each option for one type of tea party item (e.g., the “snack” shelf held 4 donuts, 4 cookies, 4 eggs, and 4 asparagus “veggies”). Tablecloths covered the shelves to prevent the children from handling the items prematurely. Taped on the wall above each shelf were pictures of the 4 options on the shelf.

Each child received 4 conformity trials (1 trial each for the plate, cup, tea, and snack items). On each trial, the host first asked the child which option they felt like using, which children typically indicated by pointing to a picture on the wall. Children’s indications seemed to reflect their actual preferences. When children did not conform,

they tended to pick the item they had initially indicated, as described in Appendix A. Next, the host initiated a video chat to check on what was happening in the other room. In a between-subjects design, half of the sample ($n = 52$) video chatted with the adult informant on all conformity trials, whereas the other half ($n = 52$) video chatted with the child informant on all conformity trials. However, both informants followed the same script.

During the video chat, the informant declared that they were looking for an item to use, rejected 3 options for that item, and then endorsed one of the options. To reduce the likelihood that the informant-endorsed option would be one that the children would have liked to use independent of the endorsement, the informant-endorsed option was always one of the less appealing options (e.g., the veggie snack). Appendix A includes more details about the options that were available and which ones were endorsed. In a within-subjects design, the informant gave 2 endorsements framed as norms (“For tea parties at Duke, we always use this kind of snack”) and 2 endorsements framed as preferences (“For my tea party today, I feel like using this snack”). Thus, the norms and preferences differed on several dimensions of conventionality, including reference to the ingroup (“tea parties at Duke” versus “my tea party”), subject (“we” versus “I”), temporal generality (“always” versus “today”), and generic language (“this kind of snack” versus “this snack”).

Next, the host paused the video on a blank frame, emphasized the informant's endorsement, asked the child to get an item, and removed the tablecloth from the appropriate shelf. The dependent measure was which option the child selected. Children scored 0 (non-conformity) on a trial if they chose any of the 3 non-endorsed options (e.g., the donut, cookie, or egg). In some cases, children did not have the opportunity to conform per our definition because they initially indicated a preference for the option that the informant would later endorse (e.g., the veggie). We reasoned that choosing the endorsed option in such cases did not accurately represent conformity, as the children may have been inclined to select their preferred option independent of the informant's endorsement.

Thus, the 19 children who initially preferred and subsequently chose the informant-endorsed option on at least one trial scored 0 on such trials. It was advantageous to code such individual trials as 0 rather than exclude the data of these 19 children entirely, since the data from these children's other trials could still contribute to the overall analysis. Altogether, there were 21 such trials, including 13 in the norm condition (in which the child indicated a preference for an option, heard the informant endorse that option with a norm, and then chose that option) and 8 in the preference condition (in which the child indicated a preference for an option, heard the informant endorse that option with a preference, and then chose that option). Children scored 1 (conformity) on a trial only if they chose the endorsed option after initially indicating a

preference for one of the other options. In such cases, children were truly conforming, as they were overriding their own preferences to behave like the informant.

In total, children scored from 0 to 2 in conformity to norms and 0 to 2 in conformity to preferences. For counterbalancing, we crossed order of presentation of item types (plates and cups followed by teas and snacks—or teas and snacks followed by plates and cups) by order of presentation of norms versus preferences (2 norms followed by 2 preferences—or 2 preferences followed by 2 norms). To reduce carryover effects, the host said a transitional remark (e.g., “We’re all done setting up the teas and snacks. In a minute, we’ll set up the plates and cups, okay?”) and played with a ball between the first two and final two conformity trials. Additionally, between each conformity trial, the host briefly distracted the child with a ball. Besides conformity, we also examined a second measure at the end of each session: whether children protested against a puppet who deviated from the informant’s endorsements. Due to the very low rates of protest (6% of the time), we moved discussion of this measure to Appendix B.

3.2 Results

The raw data are available in S1 Table (<https://doi.org/10.1371/journal.pone.0251228.s004>). For interrater reliability, a second coder viewed 23% of the sessions ($n = 24$) and coded which items the child initially preferred and subsequently chose. Agreement was perfect in both respects aside from one ambiguous case in which the child indicated two initial preferences. Using the lme4

package in R version 4.0.0, children’s conformity was analyzed with linear mixed effects models (Bates, Mächler, Bolker, & Walker, 2015), which included random intercepts for participants to account for individual variability rather than treating such variability as error, as in typical regression models, thereby enabling a more powerful analysis. A series of models was created. Model 1 was a null model containing only the random intercept of Participant. Model 2 added the fixed effects of Informant (Child, Adult) and Endorsement (Preference, Norm). Model 3 added the interaction of Informant by Endorsement. Model comparisons using likelihood ratio tests assessed whether each model’s inclusion of additional terms significantly improved the fit to the data. An alpha level of $p < 0.05$ was selected.

The inclusion of the main effects in Model 2 led to a significant improvement in fit compared to Model 1, $\chi^2(2) = 8.59, p = 0.01$. However, the inclusion of the interaction in Model 3 did not lead to an improvement in fit compared to Model 2, $\chi^2(1) = 0.51, p = 0.47$, and the interaction was not significant in Model 3 anyways ($b = -0.12, SE = 0.16, t = -0.72, p = 0.47$). Thus, the most parsimonious explanation of the data was Model 2, as shown in Table 2.

Table 2: Summary of the linear mixed effects model of conformity as predicted by Informant (Child, Adult) and Endorsement (Preference, Norm). The Child and Preference conditions were the reference levels. * $p < 0.05$; ** $p < 0.01$.

Formula: Conformity ~ Informant + Endorsement + (1 Participant)					
<i>Model fit:</i>	AIC	BIC	logLik	deviance	df.resid

	457.8	474.5	-223.9	447.8	203
<hr/>					
<i>Random effects:</i>	Variance	Std. Dev.			
Participant	0.2081	0.4562			
Residual	0.3373	0.5808			
<hr/>					
<i>Fixed effects:</i>	Estimate	Std. Error	df	t value	Pr(> t)
(Intercept)	0.2596	0.0942	148.3307	2.757	0.0066**
Informant					
[Adult]	0.2115	0.1204	104.0000	1.757	0.0818
Endorsement					
[Norm]	0.1923	0.0805	104.0000	2.388	0.0188*

Supporting our hypotheses, Model 2 included a significant main effect of Endorsement ($b = 0.19$, $SE = 0.08$, $t = 2.39$, $p = 0.02$), such that children conformed more to norms ($M = 0.56$, $SD = 0.80$) than to preferences ($M = 0.37$, $SD = 0.70$), as shown in Figure 2. Whereas children's rate of conformity to norms was higher than expected by chance, $\chi^2(1) = 11.39$, $p = 0.0007$, children's rate of conformity to preferences did not differ from chance, $\chi^2(1) = 0.03$, $p = 0.86$. The main effect of Informant was not significant ($b = 0.21$, $SE = 0.12$, $t = 1.76$, $p = 0.08$). That is, children's conformity to the adult informant ($M = 1.13$, $SD = 1.36$) did not differ significantly from their conformity to the child informant ($M = 0.71$, $SD = 1.11$). Appendix C describes the effects of counterbalancing order, which were all consistent with the results of Model 2 reported here, as well as how the main

effect of Endorsement held in both the adult informant and the child informant conditions separately.

Given the equivocal p value of the main effect of Informant ($p = 0.08$), we conducted further analyses in G*Power (Faul, Erdfelder, Buchner, & Lang, 2009) to assess how much power our study had to detect (or rule out) a potential effect of Informant. A sensitivity analysis showed that a study with our parameters ($\alpha = 0.05$, sample size = 104, numerator $df = 1$, number of groups = 4) could have detected an effect size of $f = 0.28$ (slightly larger than a conventionally medium effect size) with a power of 0.8. Moreover, a power analysis showed that a study with our parameters ($\alpha = 0.05$, numerator $df = 1$, number of groups = 4) would have required a sample size of 128 participants to detect a conventionally medium effect size ($f = 0.25$) with a power of 0.8. For smaller effect sizes, even larger sample sizes would have been required. Thus, our study may have lacked the power to detect (or conclusively rule out) a potential effect of Informant, so our findings should not be taken as strong evidence either for or against a potential effect of Informant. Given the ambiguities about how to interpret p values greater than 0.05—and given that a main effect of Informant would be independent of our main hypotheses in any case—we elected to refrain from drawing further conclusions about the potential effect of Informant.

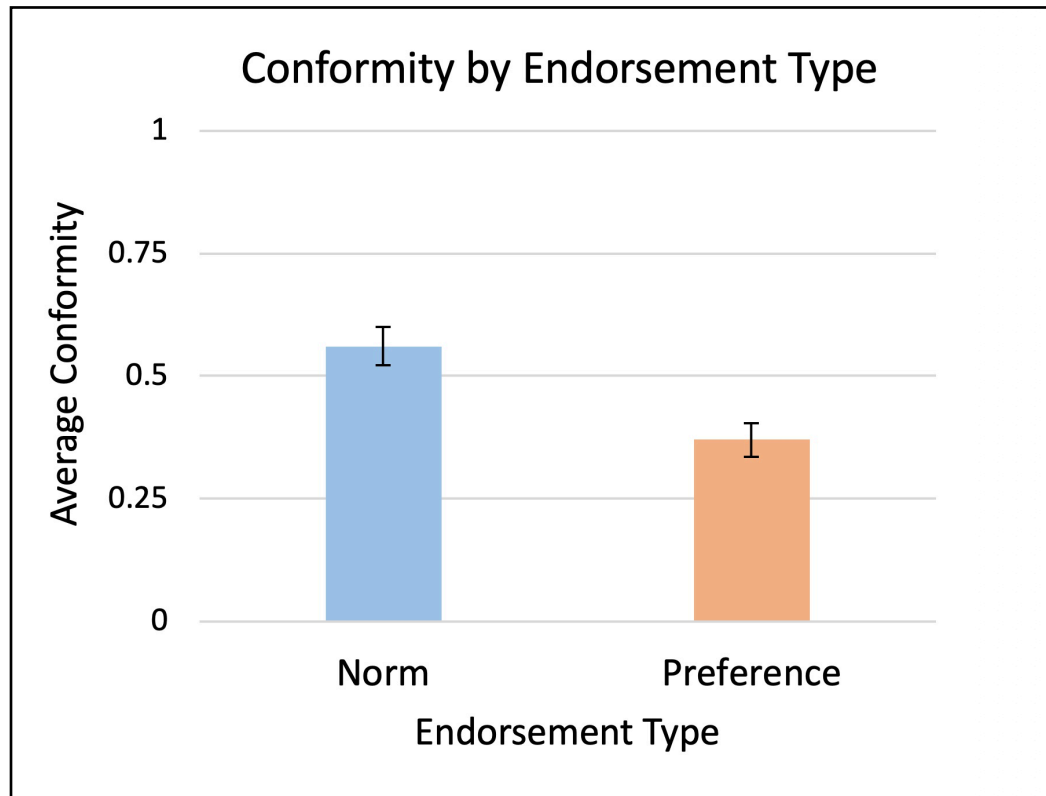


Figure 2: Children conformed more to norms than to preferences. Error bars represent standard errors. Note that the full range of conformity scores (y-axis) is from 0 to 2, although only the range from 0 to 1 is depicted here.

3.3 Discussion

In a pretend tea party, 3.5-year-old children first indicated which items they preferred to use but then heard another person, who was either an adult or another child, give endorsements of other items. Children overrode their own preferences and conformed to the other person more when the endorsements were framed as conventional norms than when they were framed as personal preferences. Moreover, children conformed more to norms than to preferences whether the informant was an adult or a child, suggesting that it is conventionality, not just adult authority, that

matters. Although the children in our norm condition conformed relatively infrequently (28%), this conformity rate was consistent with the conformity rate (also 28%) in a previous study of children's conformity (Haun et al., 2014). The relatively low conformity rate was unsurprising, given that choosing to conform meant going against one's own preferences. Considering the subtle linguistic framing that we used, which specified norms in terms of what group members descriptively "always" do instead of what they prescriptively "should" do, it is noteworthy that children still overrode their own preferences to conform.

Our findings complement previous evidence and arguments suggesting that children are motivated to act conventionally (Haun et al., 2014; Legare & Nielsen, 2015; Li & Tomasello, 2021; McGuigan & Robertson, 2015; Schmidt et al., 2016; Schmidt et al., 2011; Tomasello, 2016b). Such a motivation to act conventionally would have been adaptive throughout human evolution. It would have helped children navigate important social challenges, such as affiliating with one's cultural group, learning causally opaque but meaningful social practices, and selectively learning which actions performed by others are necessary to adopt (e.g., social norms) and which are not (e.g., others' preferences). Given its many functions, such a motivation to act conventionally would have contributed substantially to the development of human culture and human uniqueness.

Humans inherit from their forebears not only their genes but also their cultural practices. For this process of cultural inheritance to work, younger generations must be motivated to acquire culture from older generations. Previous studies have suggested that young children are motivated to act conventionally, but these studies were limited in that they did not rule out plausible alternative explanations for why children may seek to act in conventional ways (e.g., perhaps children only perform conventional actions because they seek to affiliate with others or achieve instrumental goals). By using a task setting that accounted for these alternative explanations, we provided further evidence that young children do have a specific motivation to act conventionally.

In our study, conformity was the dependent measure of interest, but researchers have also used other kinds of dependent measures to probe children's developing understanding of norms. For instance, a large body of research has examined how children react when others violate norms. This line of research has revealed that children protest against transgressors (Schmidt & Tomasello, 2012) and tattle on transgressors to observers (Yucel & Vaish, 2018). Such acts of third-party norm enforcement, in which children intervene against transgressions that do not personally harm them, indicate that children are committed to upholding norms above and beyond simply protecting their own self-interest.

Other research has examined whether children can create new norms during peer interactions. An ability to create new norms with peers, not just follow existing

norms handed down by adults, is significant because it speaks to an understanding that norms are essentially social agreements. This line of research has shown that 5-year-olds can create novel norms with peers (Göckeritz, Schmidt, & Tomasello, 2014; Hardecker, Schmidt, & Tomasello, 2017). Moreover, 5-year-olds also teach their self-created norms to novices (Göckeritz et al., 2014; Hardecker et al., 2017). In future research, it may be worthwhile to examine whether a priority of norms over preferences would manifest not only in the dependent measure of conformity but also in other relevant dependent measures, such as protesting, tattling, norm creation, and norm teaching. Plausibly, children may protest and tattle more against deviations from norms than against deviations from preferences (although, as we note in Appendix B, we observed little protest in our study). It would also be interesting to analyze children's discourses during norm creation and transmission to see how children themselves linguistically cue conventionality.

There are some limitations of our study to consider. One limitation of our study was that the endorsements of both the adult informant and the child informant were emphasized by the adult host. Thus, in both cases, participants may have felt that the informant's endorsements were bolstered by adult authority. This methodological concern may be tempered by the fact that the host always asked the adult/child informant what they were up to and followed the informant's lead by emphasizing what

the informant said, so the source of the endorsement was known to be the informant, not the host. Nonetheless, this may merit future research with a more controlled design.

A second methodological limitation was that we did not assess the strength of the children's own preferences beyond simply asking children to indicate which option they felt like using at the beginning of each trial. Potentially, children's willingness to conform (i.e., their willingness to override their own preference and adopt someone else's choice) may have varied across trials based on how committed the children were to their own preferences on particular trials. In future research, it may be interesting to examine whether the influence of linguistic cues on children's behaviors would vary depending on the strength of the children's own preferences. A third limitation was that the child informant, a 6-year-old girl, was older than our 3.5-year-old participants. To our participants, the child informant may have actually exuded authority, so our conclusion that children prioritized norms out of respect for conventionality, not just respect for authority, invites further investigation. Future research could also examine whether children consider same-age or even younger peers to be valid messengers of norms.

A fourth limitation was that the expressions of norms and preferences differed on several dimensions, not just one. As this was the first study (to our knowledge) to compare the relative effects of norms and preferences on children's conformity, we elected to present norms and preferences as they occur naturalistically — with multiple

features differing between them – to first establish whether they differed from each other as a whole. To repeat, the norms were expressed by saying: “For tea parties at Duke, we always use this kind of (item).” The preferences were expressed by saying: “For my tea party today, I feel like using this (item).” As such, the norms and preferences differed by several features, including reference to the ingroup, subject, temporal generality, and use of generic language.

Thus, although we found a main effect of endorsement type, such that children conformed more to norms than to preferences, it remains unclear which particular features of norms were the ones responsible for promoting conformity. It is likely that multiple features, not just one, had an influence on children’s behavior, but additional research will be needed to disentangle their effects. Particularly, more research should examine linguistic cues of group membership and whether such cues are effective at influencing children’s behaviors. One relevant study found that children interpret descriptive cues about how members of a group regularly behave to be, indeed, prescriptive cues for how members of that group should behave (Roberts, Ho, & Gelman, 2017). In other words, children make an inferential leap from descriptive information (how group members regularly act) to prescriptive expectations (how group members should act) (Roberts et al., 2017). Further research into how children interpret and respond to cues of group membership could help test the influential hypothesis

(Legare & Nielsen, 2015; Li & Tomasello, 2021; Rakoczy & Schmidt, 2013; Tomasello, 2016b) that respect for the group is a powerful source of normativity for young children.

As for our study, our results were certainly consistent with the hypothesis that respect for the group is a source of normativity for young children. However, we acknowledge that we cannot conclusively assert that it was respect for the group and not some other possible feature (e.g., the temporal generality or the use of generic language in the phrasing) that led children to favor norms over preferences in our study, so further research is warranted. In future studies focusing on whether groups are perceived as sources of normativity by children, it may be advisable for researchers to assess children's reactions to being labeled as a new member of the group, such as by asking children about their attitudes towards the group.

Finally, a fifth limitation was that our sample consisted of relatively affluent children in a Western context, so our conclusions warrant further study from broader cultural contexts. Notably, previous research found that framing a necklace making activity as conventional rather than instrumental increased imitative fidelity not only for Western children from the United States but also for non-Western children from Melanesia, suggesting universal processes (Clegg & Legare, 2016a). In addition to assessing how non-Western children interpret and respond to linguistic cues of conventionality, future research may also address more targeted questions regarding, for instance, the magnitudes of such effects in different cultures or the age at which

children from different cultures begin to prioritize conventionality. Given that human uniqueness is based in large part on the human capacity for culture, further research on how young children from different cultures acquire their various cultural competencies will go a long way towards answering the timeless question of how human psychology became unique.

Chapter 4. How Social Interactions Contribute to Moral Development

The great philosopher Immanuel Kant once remarked that: “Two things fill the mind with ever new and increasing admiration and awe. . . the starry heavens above and the moral law within” (Kant, 1788/2004, p. 170). Undoubtedly, human morality encompasses a vast inner realm of private moral judgments, intuitions, and deliberations. Researchers have argued that many of these aspects of moral cognition are innately fixed from birth and independent of social influence (Mikhail, 2007). Nonetheless, this inner moral realm that Kant and others have highlighted is only half of the story of human morality. In practice, morality is not only a solitary activity of private deliberation (within-minds). Functionally, humans also participate in morality as a social activity (between-minds). Engaging with others cooperatively, humans co-construct moral values, norms, and conventions as forms of common ground (i.e., mutual knowledge), which they reference when deciding how they themselves or others should act (Li & Tomasello, 2021).

When morality is thus conceived as a social activity based on common ground, then moral development may be viewed as the process of learning how to participate effectively in this kind of social activity with others. As such, a key part of moral development is learning how to reference and articulate the moral common ground one shares with others when making joint decisions or reasoning with others. This

conceptualization of moral reasoning as a between-minds process (reasoning *with* others) stems from a broader theoretical account that considers reasoning in general to be an inherently social activity (Mercier & Sperber, 2011; Tomasello, 2014, 2019). Prototypically, humans reason with each other in contexts of cooperative joint decision-making, in which the goal is to determine the best course of action based on the knowledge available to both parties (Tomasello, 2014). Interpersonal processes of reasoning (e.g., requesting, proposing, or challenging reasons for different possible courses of action) evolved as adaptations for effective joint decision-making (Tomasello, 2014).

4.1 Moral Reasoning in Young Children

Research has shown that young children are skilled at reasoning with others based on the common ground they share with those others (Köymen, Mammen, & Tomasello, 2016; Köymen & Tomasello, 2020). When jointly deciding where to place objects in a toy zoo, for instance, children from 3 to 5 years of age provide more explicit justifications for their proposed decisions when common ground with their partner is weak (e.g., discussing where to place a piano, which has no clearly appropriate location) than when common ground is strong (e.g., discussing where to place a polar bear, which should clearly go in the ice area) (Köymen, Rosenbaum, & Tomasello, 2014). Children from 3 to 5 can also reference both direct reasons (e.g., “the footprints indicate where the animal went”) and indirect reasons (e.g., “the footprints indicate where the animal did

not go”) when discussing joint decisions (Köymen, Jurkat, & Tomasello, 2020).

Additionally, one study showed that 5-year-olds can produce valid counter-arguments to a partner’s ideas; this study also found that 3-year-olds rarely produced counter-arguments on their own, but their ability to do so could be improved with training (Köymen, O’Madagain, Domberg, & Tomasello, 2020).

Research has also examined how children reason with others about specifically moral issues. One central question, which researchers as early as Piaget (1932) have explored, is whether children reason about moral issues differently when interacting with peers versus with adults. A number of studies have suggested that children reason more actively about moral dilemmas when discussing such dilemmas with their peers as opposed to with their mothers (Kruger, 1992, 1993; Kruger & Tomasello, 1986; Mammen, Köymen, & Tomasello, 2019). This difference may be due to adult-child interactions having a larger asymmetry between interactants in terms of knowledge or power compared to peer interactions (Kruger, 1992; Mammen et al., 2019). Accordingly, peer interactions may be relatively more conducive to moral development because they afford children the opportunity to reason in earnest with a peer of equal standing, as opposed to simply deferring to the authority of an adult (Kruger, 1992; Piaget, 1932).

Relatedly, researchers have long been interested in exploring which particular features of social interactions are the most conducive to moral development. Different researchers have highlighted different features. Damon and Killen (1982) found that

children who improved in moral reasoning after a social interaction were likely to have both given and received accepting and “transforming” statements (e.g., statements that correct, extend, or compromise with the ideas of one’s partner) during the interaction. Similarly, Kruger (1992, 1993) found that children who improved in moral reasoning after a social interaction were likely to have used “transactive” reasoning (e.g., reasoning that critiques, extends, or justifies the ideas of oneself or one’s partner). Both Damon and Killen (1982) and Kruger (1993) interpreted their respective findings to mean that social experiences of disagreement followed by “co-construction” of solutions (in which partners cooperatively discuss, critique, and integrate each other’s ideas) are helpful for moral development. As stated by Kruger (1993), “what is critical. . . is the opportunity the child has to compare his or her understanding to that of another and to attempt to integrate the varying perspectives” (p. 167).

In contrast to these studies, however, Walker, Hennig, and Krettenauer (2000) found that transactive social interactions were not strongly related to moral development. In their study, the types of social interactions that seemed to be more related to moral development included “representational” interactions (in which one’s partner repeated or elicited one’s input) and supportive interactions (in which one’s partner showed positive affect or encouraged one’s input). As stated by Walker et al. (2000), “a gentle Socratic style of eliciting the other’s opinions and checking for understanding—of drawing reasoning out through the use of appropriate probes—can

be effective” (p. 1045) for promoting development. In addition, “interfering” interactions (in which one’s partner resisted or devalued the discourse) also predicted development in their study (Walker et al., 2000).

4.2 Present Study

One limitation of the aforementioned studies is that they only used correlational, not experimental, approaches to identify potential features of social interactions that may be helpful for moral development. That is, the previous studies found associations between the types of social interactions children experienced and the children’s moral reasoning, but the studies did not experimentally control the features of the social interactions that children experienced. As such, the conclusions of the previous studies (e.g., that transforming, transactive, or representational social interactions promote moral development) are not immune from alternative explanations involving potential confounds or third variables. A second limitation is that different studies had divergent findings, which means that it is still unclear which particular features of social interactions are the most conducive to moral development. In all likelihood, there are multiple features of social interactions, not just one, that promote development. But only an experimental design that isolates and disentangles the individual impacts of different features can conclusively clarify which features indeed promote development.

The present study, therefore, aimed to address these limitations by employing an experimental design. In the present study, children discussed what to do in simple

moral scenarios with a puppet interlocutor. We experimentally assigned children (between-subjects) to experience one of four types of social interaction with the puppet. The four types of social interaction followed a factorial design crossing two features that previous research has indicated may be helpful for moral development: Disagreement (the puppet either agreed or disagreed with the child's ideas) and Justification (the puppet either asked the child to justify themselves or not). Accordingly, the four conditions were termed Simple Agreement (Agree + Do Not Justify), Simple Disagreement (Disagree + Do Not Justify), Justifying After Agreement (Agree + Justify), and Resolving Disagreement (Disagree + Justify). Employing the logic of a training study, we assessed children's moral judgments and reasoning both within the ongoing social interactions (the "training phase"), which differed between the four conditions, as well as within a different context (the "test phase"), which was the same for all children.

4.2.1 Operationalizing Moral Development

As the aim of the present study was to pinpoint which features of social interactions promote moral development, a concrete operationalization of what it means to morally develop was needed. For this purpose, it was advantageous to focus on issues of fairness (e.g., how to allocate things between individuals) because children have a known starting point and developmental trajectory in their judgments and reasoning about such issues. Namely, young children initially show an inflexible equality bias—they tend to always prefer strictly equal allocations, regardless of context—but improve

with age at considering other aspects of fairness, such as need or merit, which would make it fair to enact unequal allocations in some cases (Fehr, Bernhard, & Rockenbach, 2008; Li, Rizzo, Burkholder, & Killen, 2017; Rizzo, Elenbaas, Cooley, & Killen, 2016; Rizzo, Li, Burkholder, & Killen, 2019; Shaw & Olson, 2012; Sigelman & Waitzman, 1991).

Thus, in the context of fairness judgments, moral development may be concretely operationalized in terms of a shift away from a contextually insensitive equality bias in the direction of considering other aspects of fairness, such as need or merit. This shift represents a positive development in that one is learning to make appropriate and justifiable decisions in line with common ground values. Moreover, because morality is a social activity, moral development also includes becoming better at reasoning about these common ground values when justifying one's decisions to others. In the present study, as such, we operationalized moral development in terms of being able to allocate unequally, not just equally, in cases where one recipient is more deserving of an allocation than another, as well as being able to justify oneself effectively within the social activity of reasoning with others.

The training phase included moral scenarios about two topics: how to allocate play time with fun toys between two individuals and how to allocate clean-up tasks between two individuals. To avoid testing children on the same topics that they encountered in the training phase, the test phase included moral scenarios about two other topics: how to allocate cookies between individuals (distributive fairness) and how

to allocate punishment between individuals (retributive fairness). We were interested in whether the different social interaction conditions in the training phase would have different impacts on children's reasoning about distributive versus retributive fairness, as the two topics are conceptually distinct. It would be possible, for instance, that the training would only influence children's subsequent thinking about distributive fairness but not their thinking about retributive fairness, or vice versa.

4.2.2 Hypotheses

In the Simple Agreement condition, the puppet agreed with the child's proposals about what to do and did not otherwise challenge or question the child's ideas. The Simple Agreement condition was considered the baseline condition, as it did not expose children to any feature of social interaction that would be expected to advance their moral development (beyond the mere fact that a minimal social interaction was occurring).

In the Simple Disagreement condition, the puppet disagreed with the child's ideas but did not make any further moves to discuss the issue or resolve the disagreement. The Simple Disagreement condition represented the pure effect of experiencing someone disagree with oneself, independent of engaging with that person further to uncover the rationale behind their judgment or explain the rationale behind one's own judgment. What this condition aimed to capture was the effect of *disequilibrium*. One's understanding of reality, Piaget (1954) argued, exists in a state of

equilibrium between one's prior knowledge and the novel information coming in from the world. When novel information is consistent with one's knowledge, then one may simply assimilate the information into one's existing schemas. But when novel information contradicts one's knowledge, then disequilibrium occurs, and one must then revise one's schemas to accommodate the new information. As Walker et al. (2000) put it, disequilibrium is "a state of cognitive conflict that challenges current ways of thinking and stimulates development toward more equilibrated (i.e., higher level) reasoning" (p. 1034).

Experiences of disagreement may suffice to trigger disequilibrium, as disagreement is itself a signal that there may exist information that one has not yet considered. For instance, suppose that one were to discover that someone whom one respected (e.g., a mentor or a professor) held a different political belief than oneself about some issue. Knowing the mere fact that the other person holds a different belief than oneself—even absent any discussion with that person about the rationale behind their belief—may prompt one to reconsider or revise one's own belief. After all, if one's own belief were incontestably true, then why would anyone else have a different belief? Thus, we predicted that the Simple Disagreement condition, acting as a vehicle of disequilibrium, would promote children's moral development relative to the baseline condition.

In the Justifying After Agreement condition, the puppet agreed with the child's ideas but also asked the child to justify themselves. The Justifying After Agreement condition represented the pure effect of having to justify one's decisions to someone else, independent of there being any disagreement that would typically motivate the need for justification. There were at least two features of this condition that may be conducive to moral development. First, this condition afforded children the opportunity to practice articulating the rationales underlying their own beliefs to another person. Similar to the hypothesized effect of disequilibrium, the process of "talking out loud" about one's own beliefs may be helpful for clarifying or revising one's beliefs—if only because articulating one's views out loud helps bring them into awareness.

Second, this condition confronted children with a powerful communicative pressure: the need to justify one's decisions to others. Crucially, when there is a communicative pressure to justify one's decision to others, there arises a corresponding social pressure to make decisions that others would consider justifiable. These social pressures may in turn prompt children to give more thought than they otherwise would to the justificatory criteria (i.e., the common ground norms and values) that inform moral decision-making. Overall, given these two positive features of the Justifying After Agreement condition (i.e., the opportunity to practice articulating one's reasoning out loud and the social pressure to make justifiable decisions), we predicted that this

condition would promote children's moral development relative to the baseline condition.

In the Resolving Disagreement condition, the puppet disagreed with the child's ideas and also asked the child to justify themselves. The Resolving Disagreement condition represented the effect of "co-constructing" a consensus with another person following a disagreement (e.g., Damon & Killen, 1982; Kruger, 1993). We expected that this condition would be the most stimulating to moral development. For one, this condition combined two features of social experience—disagreement and justification—that may each be independently stimulating to development. What is more, the integration of these two features within one interaction may produce an effect that is greater than the sum of the parts. Namely, practice at justification may be more effective when one is justifying one's beliefs in the face of opposition—with an aim to convince one's opponent to adopt one's beliefs—than when one is simply describing one's beliefs to a sympathetic listener.

This contrast may be framed as the difference between *explaining* and *persuading*. When simply explaining oneself, as in the Justifying After Agreement condition, one only needs to attend to one's own beliefs. But when trying to persuade another person who disagrees with oneself, one must consider not only one's own beliefs but also the other person's beliefs, as well as the arguments, counterarguments, and justificatory criteria relevant to the two positions. Thus, we predicted that the Resolving

Disagreement condition would produce the greatest advances in moral development compared to the other three conditions.

In addition to the social interaction conditions, another predictor that we included in our study was children's false belief competence. Previous studies have found links between children's false belief competence and their moral reasoning, but these studies defined reasoning in the more internal (within-minds) sense of private deliberation (Killen, Mulvey, Richardson, Jampol, & Woodward, 2011; Lane, Wellman, Olson, LaBounty, & Kerr, 2010). No study to our knowledge has examined how children's false belief competence relates to their ability to morally reason *with* others interpersonally (between-minds). Theoretically, understanding that others may have false beliefs—that is, understanding that others may be wrong—is a necessary precondition for reasoning with others effectively (Köymen & Tomasello, 2020). After all, if one assumed that others' beliefs were always true, then there would be no need to reason with others or try to change their minds. Such a need to reason with others only arises when one perceives a conflict between others' beliefs and (one's own beliefs about) reality. Thus, we hypothesized that children with better false belief competence would have more advanced moral reasoning than children with worse false belief competence. We expected that this relation between children's false belief competence and their moral reasoning would hold independent of the social interaction conditions, as the effect of false belief represents a more enduring, "trait-level" disposition

compared to the more transient, “state-level” effects that may be induced by the training manipulation.

4.3 Method

4.3.1 Participants

To determine the sample size, an a priori power analysis was conducted with the statistical software G*Power. The power analysis determined that 128 participants would be needed to reach a power of .80 for detecting a conventionally medium effect size of .25 for the interaction between Disagreement (Agree, Disagree) and Justification (Do Not Justify, Justify). As such, we aimed to recruit at least 128 children. Our recruitment efforts enabled us to complete the sample with 130 children from 4 to 5.5 years of age (M age = 57 months, SD = 5, range = 48 to 66; 74 girls). The children were recruited via emails to parents in our university’s database of local birth records in the Southern United States. An additional 21 children were recruited but later excluded due to the child not understanding the activity (10), parental influence (6), excessive Internet lag (1), extreme shyness (1), refusal to participate (1), the child’s speech being incomprehensible (1), and the researcher forgetting to record the session (1). Children’s families received a \$15 Amazon gift card for participating. This study was approved by the Institutional Review Board of Duke University on October 15, 2020. The procedure was conducted with the consent of the parents or guardians of the minors and in

accordance with all relevant ethical and legal rules concerning psychological research in the United States of America.

4.3.2 Procedure

The study took place entirely on Zoom. To begin, the experimenter (termed the *host*) connected with the child and their parent or guardian on Zoom. Then, the host shared their screen over video chat so that both the child and the host could see the same stimuli (PowerPoint slides) on their respective computer screens. The child's Self View feature was turned off so that the child could not see the video of themselves. The host also helped the child's parent or guardian adjust the Zoom interface so that both the PowerPoint slides and the video of the host were visible to the child at sufficiently large viewing sizes. The study unfolded in three phases: the pre-training phase, the training phase, and the test phase.

4.3.2.1 Pre-Training Phase: Warm-Up

The host began with a warm-up activity, which was meant to help children feel comfortable with speaking up and to familiarize children with the process of enacting allocations to virtual recipients. First, the host showed a slide with four animals (a cat, a horse, a bird, and a bunny), which were depicted in square boxes with either green or blue borders. The host asked the child whether each animal was in a green box or a blue box. Then, the host asked the child which animal was their favorite and what they liked about their chosen animal. Next, the host showed a slide with three foods (apples,

grapes, and carrots) and asked the child which food was their favorite and what they liked about their chosen food. The host then said: "We're going to give these foods to some animals." For each of the following three slides, the child was asked how to allocate foods (apples, grapes, or carrots) between two animals of the same species (horses, birds, or bunnies), one of which appeared on the left in a green box and one of which appeared on the right in a blue box (e.g., "Look, here are two horses. They both want some apples. Should we give one of these horses more apples or both horses the same amount of apples?"). Each time the child made a decision, the host moved the icons of the foods to represent enacting the decision.

The final part of the warm-up was the introduction of Hedgy, a hedgehog puppet operated by the host, who would later serve as the interaction partner for the training phase of the study. Hedgy began by saying: "Hello! I'm Hedgy, and I am so happy to talk to you today! What's your name?" After the child said their name, Hedgy then said: "Hello, (Child)! Nice to meet you!" Next, Hedgy asked the child if they like to listen to stories and followed up by saying: "Me too! I love to listen to stories! But first, I need to take a nap. I'm sleepy." Hedgy then exited.

4.3.2.2 Pre-Training Phase: False Belief

After the warm-up, children completed two false belief tasks: a change-of-location false belief task followed by an unexpected-contents false belief task. In both

tasks, visual animations accompanied the narration. Children's scores on the two tasks were summed to create a composite false belief score.

The change-of-location task was about two girls named Sally and Anne, who are in a kitchen. In the story, Anne places an apple in the fridge and then exits. While Anne is gone, Sally relocates the apple from the fridge to a backpack. At this point, the child was asked: "So, where is the apple now?" If the child said any location other than the backpack, the story was explained again until the child gave the correct answer. Next, the child was asked: "When Anne comes back into the kitchen, where will she look for the apple?" An answer of "backpack" represented failure on the task (0), whereas an answer of "fridge" represented success on the task (1). For interrater reliability, two research assistants each coded 13% ($n = 17$) of children's responses on the change-of-location test question (26% in total). They attained Cohen's κ values of 1.00 and 1.00, respectively, with reference to the complete dataset.

The unexpected-contents task featured an egg carton. The first question to the child was: "What do you think is inside the egg box?" If the child said anything other than eggs, the host talked to the child until the child was convinced that there would be eggs inside. Next, the slideshow depicted the egg box opening and containing strawberries inside. The host then remarked: "Oh, look! The box actually had strawberries, not eggs. Let's close the box." The slideshow then depicted the egg box closing. At this point, the host asked the child: "So, what is really in the box?" If the

child said anything other than strawberries, the contents of the box were explained again until the child gave the correct answer. Next, the slideshow showed a cartoon boy appearing beside the egg carton. The host then said: "Look, here is a boy named Matt. Matt has never seen inside the box before. What will Matt think is in the box?" An answer of "strawberries" represented failure on the task (0), whereas an answer of "eggs" represented success on the task (1). For interrater reliability, two research assistants each coded 13% ($n = 17$) of children's responses on the unexpected-contents test question (26% in total). They attained Cohen's κ values of 1.00 and 1.00, respectively, with reference to the complete dataset.

4.3.2.3 Training Phase

At the start of the training phase, the host told the child: "Now, we're going to listen to some more stories and decide what to do." The host asked in the direction of Hedgy: "Hedgy, are you ready to talk to us again?" Hedgy reappeared and said: "Yes." The host then told the child and Hedgy together: "(Child) and Hedgy, the two of you should talk *together* and decide *together* on what to do. You should try to make the most fair decisions." The host checked with Hedgy and the child that they were ready to begin and then said: "OK, let's begin!"

The training phase consisted of six stories in a fixed order, as described in Table 3. The first three stories were about allocating play time with toys. The next three stories were about allocating clean-up tasks. Each set of three stories included two stories in

which one recipient was arguably more deserving of the allocation than the other recipient (Stories 1, 2, 4, and 5) as well as one story in which both recipients were equally deserving (Stories 3 and 6). These stories in which both recipients were equally deserving were included to help prevent children from forming an expectation that the only “right answer” was to always allocate more to one of the recipients.

Table 3: The scripts of the six stories in the training phase.

Introduction to the stories about play time	
<p>“It’s play time at school, and the kids want to play with toys. The kids like playing with toys, like this musical instrument, this fun game, and this truck. But only one kid can play with a toy at a time. If one kid gets more play time with a toy, that means another kid gets less play time with the toy. We have to decide who gets play time with these toys. Does that make sense? OK.”</p>	
Story 1: Unequal Home Access	<p>“Here are two girls. They both want play time with the musical instrument. This girl has no musical instruments at home. This girl has many musical instruments at home.”</p>
Story 2: Unequal Prior Opportunity	<p>“Here are two boys. They both want play time with the fun game. This boy has played the fun game many times before. This boy has never played the fun game before.”</p>
Story 3: Equal Desire	<p>“Here are two girls. They both want play time with the truck. This girl wants to put blocks in the truck. This girl also wants to put blocks in the truck.”</p>
Introduction to the stories about clean-up tasks	
<p>“Look at this messy table. It’s clean-up time at school, and the kids have to help do clean-up tasks, like cleaning the tables or putting stuff away. Does that make sense? OK. The kids don’t like doing clean-up tasks. They think it’s boring, and they’d rather do something else. But there’s a lot to clean up! If one kid does fewer clean-up tasks, that means another kid has to do more clean-up tasks. We have to decide who should do the clean-up tasks. Does that make sense? OK.”</p>	

Story 4: Unequal Work	“Here are two boys. They both do not like doing clean-up tasks. This boy has cleaned 4 tables already. This boy has only cleaned 1 table already.”
Story 5: Unequal Discomfort	“Here are two girls. They both do not like doing clean-up tasks. This girl doesn’t want to do clean-up tasks because she wants to play. This girl doesn’t want to do clean-up tasks because her head really hurts.”
Story 6: Equal Work	“Here are two boys. They both do not like doing clean-up tasks. This boy has cleaned 2 tables already. This boy has also cleaned 2 tables already.”

In the visual depiction of each story, one boy/girl was shown in a green box on the left, and another boy/girl was shown in a blue box on the right. The positions of which boy/girl appeared in the green box on the left and which boy/girl appeared in the blue box on the right on every given trial were counterbalanced across two versions of the PowerPoint file. On each trial, when the host referred to the boy/girl on the left (e.g., “This girl has no musical instruments at home”), the green border surrounding that boy/girl intensified to help signify the reference. Similarly, when the host referred to the boy/girl on the right (e.g., “This girl has many musical instruments at home”), the blue border surrounding that boy/girl intensified to help signify the reference.

The host told the stories one at a time. At the end of each story, the host asked the child: “So, should we give one of these (boys/girls) more (play time/clean-up tasks) or both (boys/girls) the same amount of (play time/clean-up tasks)?” If the child said to give more to one of the boys/girls, the host followed up by asking: “Which (boy/girl)

should we give more (play time/clean-up tasks) to?" Once the child finished answering, Hedgy responded to the child's decision in one of four ways. The four kinds of responses followed a factorial design crossing Disagreement (Agree, Disagree) and Justification (Do Not Justify, Justify), as shown in Figure 3. The assignment of the four conditions to the children was between-subjects, so that each child experienced the same kind of interaction from Hedgy on all six stories.

- In the Simple Agreement (Agree + Do Not Justify) condition ($n = 32$), Hedgy agreed with the child and did not ask the child to justify their decision.
- In the Simple Disagreement (Disagree + Do Not Justify) condition ($n = 33$), Hedgy disagreed with the child and did not ask the child to justify their decision. If the child had suggested allocating equally, then Hedgy suggested allocating more to the more deserving recipient (in the cases in which one recipient was more deserving than the other) or more to the recipient on the left (in the cases in which both recipients were equally deserving). But if the child had suggested allocating more to one of the recipients, then Hedgy suggested allocating equally.
- In the Justifying After Agreement (Agree + Justify) condition ($n = 33$), Hedgy agreed with the child and then asked the child to justify their decision. After the child gave a justification, Hedgy then stated: "Oh, OK."

Simple Agreement		Simple Disagreement	
If child says give the same:	If child says give more to X:	If child says give the same:	If child says give more to X:
"I think we should give the same to both (boys/girls). You also think that."	"I think we should give more to X. You also think that."	"I think we should give more to X. But you think we should give the same to both (boys/girls)."	"I think we should give the same to both (boys/girls). But you think we should give more to X."
I guess we agree.		I guess we don't agree.	
Host: Let's hear the next story.			
Justifying After Agreement		Resolving Disagreement	
If child says give the same:	If child says give more to X:	If child says give the same:	If child says give more to X:
"I think we should give the same to both (boys/girls). You also think that."	"I think we should give more to X. You also think that."	"I think we should give more to X. But you think we should give the same to both (boys/girls)."	"I think we should give the same to both (boys/girls). But you think we should give more to X."
"So, why do <u>you</u> think we should give the same?"	"So, why do <u>you</u> think we should give more to X?"	"So, why do <u>you</u> think we should give the same?"	"So, why do <u>you</u> think we should give more to X?"
"Oh, OK."	"Oh, OK."	"Oh, OK."	"Oh, OK."
Host: Let's hear the next story.			

Figure 3: The scripts for how Hedgy responded to the child's decisions in the four experimental conditions.

- In the Resolving Disagreement (Disagree + Justify) condition ($n = 32$), Hedgy disagreed with the child and then asked the child to justify their decision. After the child gave a justification, Hedgy then stated: “Oh, OK.” If the child had suggested allocating equally, then Hedgy suggested allocating more to the more deserving recipient (in the cases in which one recipient was more deserving than the other) or more to the recipient on the left (in the cases in which both recipients were equally deserving). But if the child had suggested allocating more to one of the recipients, then Hedgy suggested allocating equally.

After each interaction with Hedgy, the host then said: “Let’s hear the next story.” The next story was then shown. Once all six stories were finished, Hedgy remarked: “I’m sleepy again. I need to take another nap. It was great talking to you!” Hedgy then exited, and the test phase began.

4.3.2.4 Test Phase

The test phase was the same for all children and consisted of six stories in a fixed order, as described in Table 4. The first three stories were about distributive fairness in the form of allocating cookies. The next three stories were about retributive fairness in the form of allocating punishment. Each set of three stories included two stories in which one recipient was arguably more deserving of the allocation than the other recipient (Stories 7, 8, 11, and 12) as well as one story in which both recipients were equally deserving (Stories 9 and 10). The visual formatting was the same as in the

training phase. On each trial, one boy/girl was shown in a green box on the left, and another boy/girl was shown in a blue box on the right. The positions of which boy/girl appeared on the left and which boy/girl appeared on the right on every given trial were counterbalanced across the two versions of the PowerPoint file. As in the training phase, when the host referred to one of the boys/girls, the border surrounding that boy/girl intensified to help signify the reference. On story 7, 10, and 11, visual animations accompanied the narration.

Table 4: The scripts of the six stories in the test phase.

Introduction to the stories about distributive fairness	
“It’s snack time at school, and the kids want to eat cookies. But we don’t have a lot of cookies. If one kid gets more cookies, that means another kid gets fewer cookies. We have to decide who should get the cookies. Does that make sense? OK.”	
Story 7: Unequal Contribution	“Here are two girls. They both like cookies. During snack time, all the kids were supposed to help make cookies to share with the school. This girl was lazy. She didn’t work hard at all, and she only made 1 cookie to share with the school. Let’s put the cookie on the plate. This girl was hard-working. She worked very hard, and she made 6 cookies to share with the school. Let’s put the cookies on the plate.”
Story 8: Unequal Hunger	“Here are two boys. They both like cookies. This boy is very hungry because his mom forgot to give him breakfast. This boy is not hungry because his mom did give him breakfast.”
Story 9: Equal Desire	“Here are two girls. This girl likes cookies a lot. This girl also likes cookies a lot.”

Introduction to the stories about retributive fairness	
<p>“These kids are at school. Some of these kids have been behaving badly, and they’re going to get in trouble with the teacher. The kids don’t like getting in trouble with the teacher. We have to help the teacher decide who should get in trouble. Does that make sense? OK.”</p>	
<p>Story 10: Equal Outcomes</p>	<p>“Here are two boys. They both put their sister’s lunch in the trash, which made their sister sad. This boy put his sister’s sandwich in the trash. This boy also put his sister’s sandwich in the trash.”</p>
<p>Story 11: Unequal Intentions</p>	<p>“Here are two girls. They both put their brother’s lunch in the trash, which made their brother sad. This girl put her brother’s lunch in the trash because she was mad at her brother. This girl put her brother’s lunch in the trash because his lunch was in a paper bag that looked like trash.”</p>
<p>Story 12: Unequal Outcomes</p>	<p>“Here are two boys. They both stole food from their friend. This boy stole one granola bar from his friend. This boy stole five granola bars from his friend.”</p>

The host told the stories one at a time. At the end of each story about distributive fairness, the host asked the child: “So, should we give one of these boys/girls more cookies or both boys/girls the same amount of cookies?” If the child said to give more to one of the boys/girls, the host followed up by asking: “Which boy/girl should we give more cookies to?” The host then asked the child to justify their decision: “Why do you think we should give (the same)/(more to that boy/girl)?” Following this, the host then said: “OK. Let’s hear the next story.”

The phrasing of the questions about retributive fairness was slightly different. At the end of each story about retributive fairness, the host asked the child: “So, should one

of these boys/girls get in more trouble, or should both boys/girls get in the same amount of trouble?" If the child said one of the boys/girls should get in more trouble, the host followed up by asking: "Which boy/girl should get in more trouble?" The host then asked the child to justify their decision: "Why do you think (they should get in the same amount of trouble)/(that boy/girl should get in more trouble)?" Following this, the host then said: "OK. Let's hear the next story." After Story 12, a slide with a smiling face appeared, and the host thanked the child for participating.

4.3.3 Coding and Reliability

Children's allocation decisions for the twelve stories were coded as to whether children gave equally to the two recipients or more to one of the recipients (and, in the latter case, which of the two recipients children gave more to). For interrater reliability, two research assistants each coded 13% ($n = 17$) of children's allocation decisions for the twelve stories (26% in total). They attained Cohen's κ values of 0.99 and 1.00, respectively, with reference to the complete dataset.

Children's justifications for their allocations were also coded. The coding scheme was based on an inductive assessment of the data and contained twelve types of codes, as shown in Table 5. The twelve codes fell into two larger categories: valid forms of reasoning and invalid responses. The valid forms of reasoning were legitimate justifications for either allocating equally or allocating more to one of the recipients. For example, one participant who allocated the same amount of clean-up tasks to the two

boys on Story 6 justified their decision by noting that “they both cleaned two tables.” By referencing the fact that the two boys were identical on a relevant dimension (the amount of work they had already done), the participant was indeed providing a valid justification for treating the boys the same (by giving them the same amount of additional work). In another case, a different participant justified their decision to allocate more cookies to one of the boys on Story 7 by noting that “he didn’t have breakfast.” By referencing the boy’s need, the participant was making a valid appeal to the common ground principle that those in greater need of a resource for the sake of their well-being may have a legitimate claim to taking more of that resource compared to others.

Table 5: The coding scheme for children’s justifications of their allocations.

	Valid Justifications	Example Excerpts
It’s Not Fair	<ul style="list-style-type: none"> • Saying it’s not fair or wouldn't be fair for someone to have, get, or do more of X. • Saying that if someone got more X, the other person would get less X. 	<p>“that won't be fair if somebody else got more cookies”</p> <p>“if one girl gets more cookies the second girl will get none”</p>
Both	<ul style="list-style-type: none"> • Referencing how both recipients are similar: Both are, like, want, need, have, can have, should have, did, can do, or should do X or the same amount of X. 	<p>“they both cleaned two tables”</p> <p>“they both like cookies”</p> <p>“they both did something bad”</p>
Need / Deficit	<ul style="list-style-type: none"> • Referencing a state of need, such as pain, sadness, hunger, or lack of food. • Referencing that someone has none of, has less of, has only a little bit of, or never got to do X. 	<p>“he never played with that game before”</p> <p>“he didn't have breakfast”</p> <p>“he's hungry and the blue one isn't”</p>

Merit	<ul style="list-style-type: none"> Referencing how much work someone did or how hardworking or lazy someone is. 	<p>"he only cleaned 1 table"</p> <p>"she worked hard"</p>
Cognitive State	<ul style="list-style-type: none"> Indicating that someone was mad or angry. Indicating what someone thought, knew, or didn't know. 	<p>"she was mad at her brother"</p> <p>"she didn't know that her brother's lunch was inside the paper bag"</p>
Stealing More	<ul style="list-style-type: none"> Indicating that one recipient stole more than the other recipient. 	<p>"he stole more granola bars"</p> <p>"he stole 5 and he stole 1"</p>
Other	<ul style="list-style-type: none"> Giving a coherent justification not included in the other valid codes. 	<p>"then they can trade"</p> <p>"she threw away a whole bag filled with things"</p>
Invalid Justifications		Example Excerpts
Fair / Share / Same	<ul style="list-style-type: none"> Referencing "sharing" or "being fair" without explaining further. Saying that the recipients have, got, want, need, or eat "the same" X without explicitly mentioning "both" recipients. 	<p>"they like to share"</p> <p>"it's fair"</p> <p>"they like to have the same"</p> <p>"it makes them have the same amount of play time"</p>
Want / Like / Have to	<ul style="list-style-type: none"> Saying that someone wants, likes, should get, would be happy to get, or has to do X without explaining why. Saying that someone can have as much as they want of X. Saying that one recipient wants X more than the other recipient without explaining why. 	<p>"she likes cookies a lot"</p> <p>"they're supposed to get in trouble"</p> <p>"they would be happy"</p> <p>"I think she wants to put the blocks in more"</p>
Transgression	<ul style="list-style-type: none"> Saying that someone transgressed or is "bad" but without comparing the two transgressors. 	<p>"they put it in the trash"</p> <p>"he stole a lot of granola bars"</p>
Uninformative	<ul style="list-style-type: none"> Saying something unclear, uninformative, or irrelevant, or simply repeating oneself. 	<p>"I want to"</p> <p>"it's not nice to steal it"</p>
Empty	<ul style="list-style-type: none"> Not providing reasoning. 	

The invalid responses, in contrast, were responses that were tautological, unclear, or uninformative. For instance, one child responded to the question “Why do you think we should give the same?” by saying “it makes them have the same amount.” But this response was in effect only a restatement of the decision, not an explanation for why the decision was justified. Responses were also coded as invalid if they only referenced platitudes (e.g., “it’s fair,” “they would be happy,” or “it’s not nice”) without further explanation. For the stories about retributive fairness, responses were coded as invalid if they only referenced the fact that someone committed a transgression but did not otherwise compare the two transgressors—either by noting that both transgressors did something bad, which would justify punishing the transgressors the same, or by mentioning a feature of the story that would justify punishing one of the transgressors more (e.g., the cognitive state of the transgressor or the fact that one transgressor stole more than the other). In the absence of further elaboration, the mere fact that someone committed a transgression was not itself a sufficient justification for either punishing both transgressors the same or punishing one transgressor more than the other. For interrater reliability, a research assistant coded 25% ($n = 33$) of children’s reasoning based on the twelve reasoning codes. Cohen’s κ was 0.86.

4.4 Results

There were three primary sets of analyses. The first set of analyses investigated children’s allocation decisions. The aim here was to compare how effective the different

kinds of social experiences would be at leading children to overcome their equality bias and attend to aspects of deservingness. A series of multiple regressions was conducted, with the outcome measure being how frequently children allocated more to the more deserving recipient on the trials in which one recipient was more deserving than the other.

The second set of analyses investigated children's reasoning in conjunction with their allocation decisions, again focusing on the trials in which one recipient was more deserving than the other. One aim here was to examine whether children's reasoning was correlated to their allocation decisions. Another aim was to examine how frequently children, comparing across conditions, gave responses in which they both allocated more to the more deserving recipient and also justified their allocation with a valid, not invalid, reason. This type of response represented the most mature expression of moral reasoning as a social activity that our study measured—the capacity to not only make an appropriate moral judgment but also justify one's judgment to others with reference to the moral common ground one shares with others. With this type of response as the outcome measure, a series of multiple regressions was conducted.

The third set of analyses investigated the relation between children's false belief competence and the types of responses analyzed in the preceding two sets of analyses. In one set of regressions with the false belief composite score as the predictor and age included as a covariate, children's allocation decisions were analyzed as the outcome. In

another set of regressions with the false belief composite score as the predictor and age included as a covariate, children's allocations in conjunction with valid reasoning were analyzed as the outcome.

In addition to these three primary sets of analyses, ancillary analyses were also conducted. The ancillary analyses investigated the relation between children's age and their reasoning, the effects of the condition manipulation on children's responses on Story 2 given their responses on Story 1, and the patterns of results that were obtained when only examining the subset of children who showed an equality bias on Story 1.

4.4.1 Analyses of Allocations

The analyses of children's allocations were conducted in three steps. The first analysis examined children's allocations during the training trials (Stories 2, 4, and 5). In this first analysis, Story 1 was not included because children's responses on this trial preceded any condition manipulation. The second analysis examined children's allocations in the test trials about distributive fairness (Stories 7 and 8). The third analysis examined children's allocations in the test trials about retributive fairness (Stories 11 and 12).

4.4.1.1 Training Trials

A 2 x 2 multiple regression crossing Disagreement (Agree, Disagree) and Justification (Do Not Justify, Justify) found a main effect of Disagreement ($b = 0.68$, $SE = 0.22$, $t = 3.05$, $p = 0.003$), a main effect of Justification ($b = 0.86$, $SE = 0.22$, $t = 3.86$, $p =$

0.0002), and an interaction between Disagreement and Justification ($b = -0.85$, $SE = 0.32$, $t = -2.71$, $p = 0.008$), as shown in Figure 4. To follow up on the interaction, pairwise t -tests with Bonferroni-corrected p -values were conducted. The post hoc tests revealed that when the puppet agreed with the child, asking for justification led children to allocate more to the more deserving recipient ($M = 1.42$, $SD = 1.06$) compared to not asking for justification ($M = 0.56$, $SD = 0.84$), $p = 0.001$. But when the puppet disagreed with the child, asking for justification had no effect, $p = 1.00$. The interaction could also be described from a different angle. When the puppet did not ask for justification, disagreement led children to allocate more to the more deserving recipient ($M = 1.24$, $SD = 0.83$) compared to agreement ($M = 0.56$, $SD = 0.84$), $p = 0.02$. But when the puppet did ask for justification, disagreement had no effect, $p = 1.00$. Moreover, an additional finding was that when the puppet both disagreed with the child and asked for justification, children allocated more to the more deserving recipient ($M = 1.25$, $SD = 0.84$) compared to when the puppet both agreed with the child and did not ask for justification ($M = 0.56$, $SD = 0.84$), $p = 0.02$.

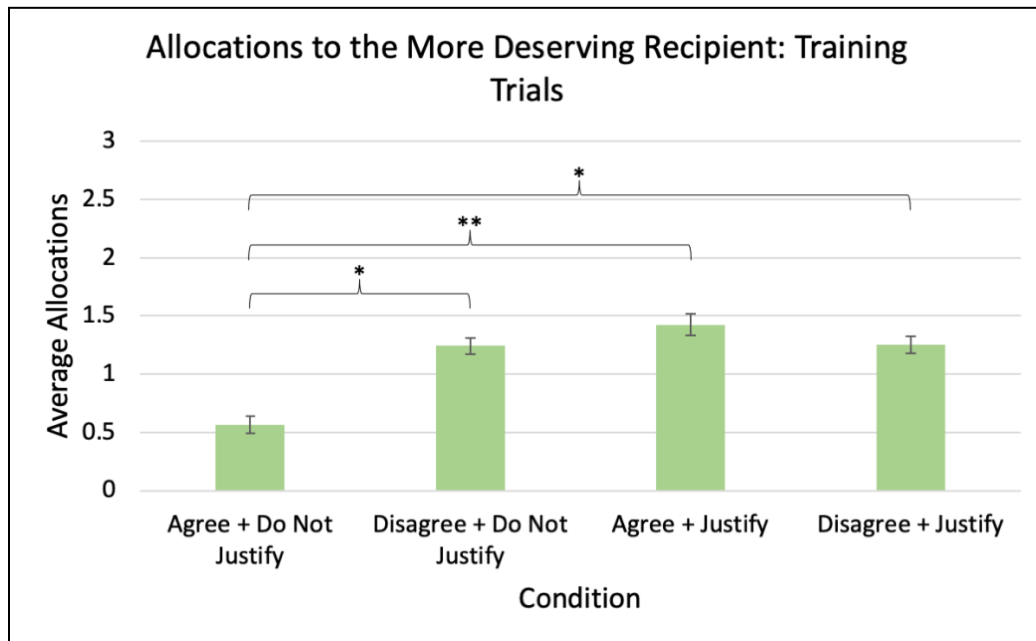


Figure 4: Children's allocations to the more deserving recipient on the training trials. Error bars represent standard errors. Asterisks represent significant differences (* = $p < 0.05$, ** = $p < 0.01$).

4.4.1.2 Distributive Fairness Test Trials

A 2 x 2 multiple regression crossing Disagreement (Agree, Disagree) and Justification (Do Not Justify, Justify) found a main effect of Disagreement ($b = 0.56$, $SE = 0.16$, $t = 3.50$, $p = 0.0006$), a main effect of Justification ($b = 0.38$, $SE = 0.16$, $t = 2.37$, $p = 0.02$), and an interaction between Disagreement and Justification ($b = -0.66$, $SE = 0.23$, $t = -2.92$, $p = 0.004$), as shown in Figure 5. To follow up on the interaction, pairwise t -tests with Bonferroni-corrected p -values were conducted. The post hoc tests revealed that when the puppet did not ask for justification, disagreement led children to allocate more to the more deserving recipient ($M = 0.97$, $SD = 0.68$) compared to agreement ($M = 0.41$,

$SD = 0.61$), $p = 0.004$. But when the puppet did ask for justification, disagreement had no effect, $p = 1.00$.

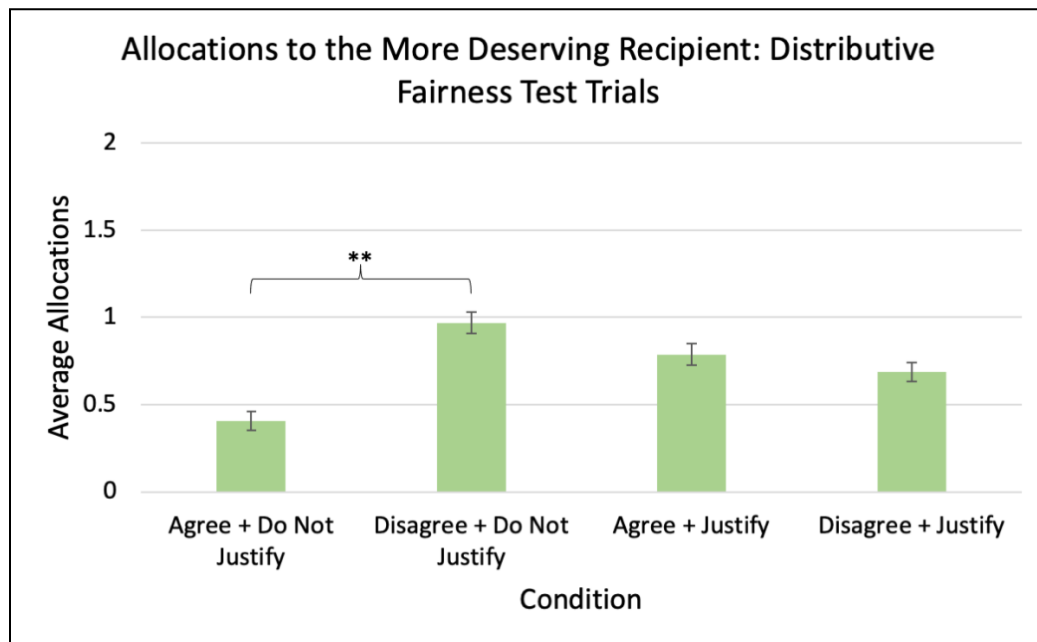


Figure 5: Children’s allocations to the more deserving recipient on the distributive fairness test trials. Error bars represent standard errors. Asterisks represent significant differences (= $p < 0.01$).**

4.4.1.3 Retributive Fairness Test Trials

A 2 x 2 multiple regression crossing Disagreement (Agree, Disagree) and Justification (Do Not Justify, Justify) found no significant effects, $ps > 0.05$. However, this may have been due to children’s particularly poor performance on Story 11, the scenario involving an accidental transgression. On Story 11, the average rates of allocating more punishment to the more intentional transgressor were 0.19 in the Simple Agreement condition, 0.15 in the Simple Disagreement condition, 0.18 in the Justifying After

Agreement condition, and 0.19 in the Resolving Disagreement condition. In other words, less than one fifth of the children in any condition allocated more punishment to the more intentional transgressor. The children's uniformly poor performance was not altogether surprising, given that children find such morally-relevant theory of mind scenarios challenging even up to age 7 or 8 (Killen et al., 2011). As such, a separate analysis focusing only on children's responses on Story 12 was conducted. With children's responses on Story 12 as the dependent variable, a 2 x 2 multiple regression crossing Disagreement (Agree, Disagree) and Justification (Do Not Justify, Justify) found a significant main effect of Disagreement ($b = 0.29$, $SE = 0.12$, $t = 2.37$, $p = 0.02$), as shown in Figure 6. Namely, disagreement led children to allocate more punishment to the more deserving recipient ($M = 0.58$, $SD = 0.50$) compared to agreement ($M = 0.42$, $SD = 0.50$).

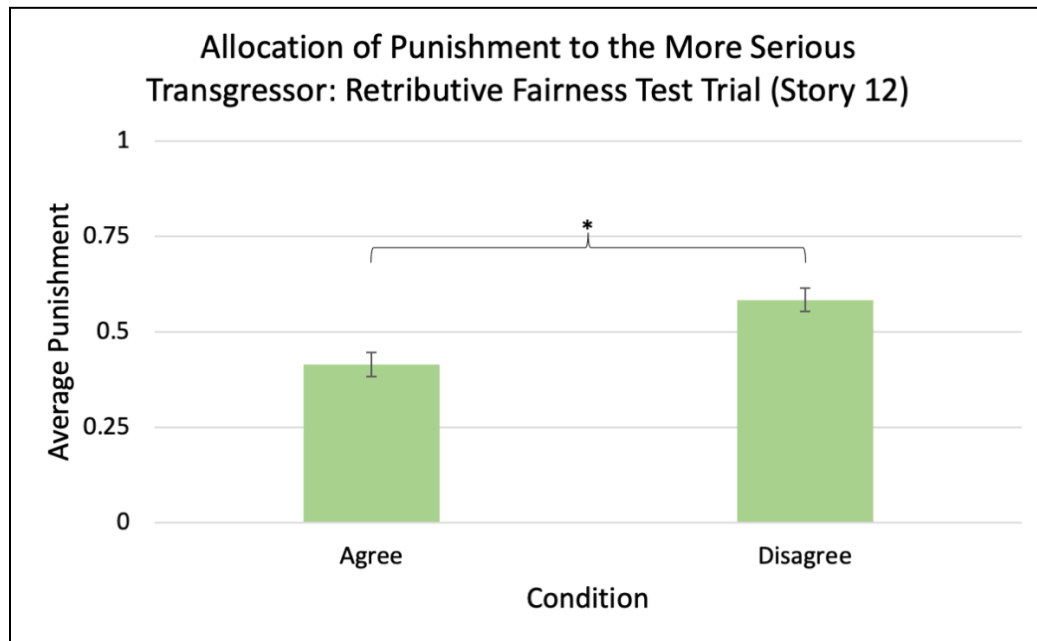


Figure 6: Children’s allocations of punishment to the more serious transgressor on the final retributive fairness test trial (Story 12). Error bars represent standard errors. Asterisks represent significant differences ($* = p < 0.05$).

4.4.2 Analyses of Reasoning in Conjunction with Allocations

To begin, two correlation analyses were conducted. The first correlation analysis examined whether children’s reasoning related to their allocation decisions during the training trials (Stories 1, 2, 4, and 5). This first analysis only examined the data from the two conditions in which children were asked for justification during the training phase. The second correlation analysis examined whether children’s reasoning related to their allocation decisions during the test trials (Stories 7, 8, 11, and 12). This analysis included the data from all four conditions, given that all of the children were asked for justification during the test trials.

Next, a set of three analyses examined children's responses in which they both allocated more to the more deserving recipient and were able to justify their allocation with a valid reason. The first analysis examined children's responses during the training trials (Stories 1, 2, 4, and 5). This time, Story 1 was included, since the provision of reasoning occurred after the condition manipulation. Note that this first analysis only included the data from the two conditions in which children were asked for justification during the training phase. The second analysis examined children's responses in the test trials about distributive fairness (Stories 7 and 8). The third analysis examined children's responses in the test trials about retributive fairness (Stories 11 and 12). These second and third analyses included the data from all four conditions.

4.4.2.1 Correlations Between Allocations and Reasoning

In the training phase, there was a significant correlation between whether children allocated more to the more deserving recipient on a given trial and whether they provided valid reasoning on that trial, $r = 0.34, p < 0.001$. Similarly, in the test phase, there was a significant correlation between these two measures, $r = 0.35, p < 0.001$.

4.4.2.2 Training Trials

A regression with the sole predictor of Disagreement (Agree, Disagree) found no difference between the Justifying After Agreement condition ($M = 1.24, SD = 1.17$) and the Resolving Disagreement condition ($M = 0.88, SD = 1.04$) in the frequency with which

children allocated more to the more deserving recipient in conjunction with giving a valid justification for their allocation, $p = 0.19$.

4.4.2.3 Distributive Fairness Test Trials

A 2×2 multiple regression crossing Disagreement (Agree, Disagree) and Justification (Do Not Justify, Justify) found a main effect of Disagreement ($b = 0.50$, $SE = 0.16$, $t = 3.18$, $p = 0.002$), a main effect of Justification ($b = 0.35$, $SE = 0.16$, $t = 2.22$, $p = 0.03$), and an interaction between Disagreement and Justification ($b = -0.70$, $SE = 0.22$, $t = -3.12$, $p = 0.002$), as shown in Figure 7. To follow up on the interaction, pairwise t -tests with Bonferroni-corrected p -values were conducted. The post hoc tests revealed that when the puppet did not ask for justification, disagreement led children to more frequently allocate more to the more deserving recipient in conjunction with giving a valid justification for their allocation ($M = 0.88$, $SD = 0.70$) compared to agreement ($M = 0.38$, $SD = 0.61$), $p = 0.01$. But when the puppet did ask for justification, disagreement had no effect, $p = 1.00$.

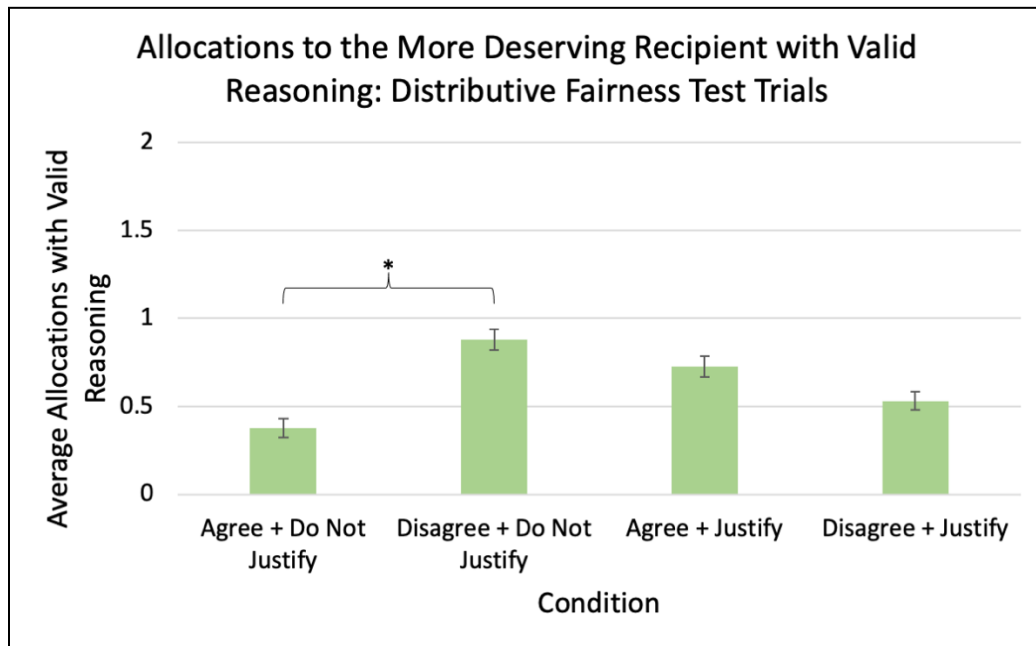


Figure 7: Children’s allocations to the more deserving recipient in conjunction with providing valid reasoning on the distributive fairness test trials. Error bars represent standard errors. Asterisks represent significant differences ($* = p < 0.05$).

4.4.2.4 Retributive Fairness Test Trials

A 2 x 2 multiple regression crossing Disagreement (Agree, Disagree) and Justification (Do Not Justify, Justify) found no significant effects, $ps > 0.05$. To account for the possibility that this result was simply due to children’s particularly poor performance on Story 11, another 2 x 2 multiple regression with the same predictors was conducted but with children’s responses on Story 12 as the dependent variable. However, this multiple regression also found no significant effects, $ps > 0.05$.

4.4.3 Analyses of False Belief

All of the outcome measures analyzed in the regressions from the preceding two sections were reanalyzed in a new series of regressions that only included the two

predictors of false belief competence and age. Children's false belief competence was represented by their false belief composite score, which ranged from 0 to 2.

4.4.3.1 Allocations

Children's false belief competence was not related to their allocations on the training trials (and this was the case whether Story 1 was included or not), $ps > 0.05$. However, children's false belief competence was related to their allocations on the distributive fairness test trials ($b = 0.19$, $SE = 0.07$, $t = 2.55$, $p = 0.01$). Namely, children with high false belief scores allocated more to the more deserving recipient compared to children with low false belief scores (false belief score of 0: $M = 0.56$, $SD = 0.64$; score of 1: $M = 0.56$, $SD = 0.55$; score of 2: $M = 0.96$, $SD = 0.73$). As for the retributive fairness test trials, there was no significant relation between children's false belief competence and their allocation decisions, and this was the case whether the two test trials were considered together or analyzed separately, $ps > 0.05$. Note that in the analysis of the retributive fairness test trials (considered together), there was a significant effect of age ($b = 0.39$, $SE = 0.13$, $t = 2.92$, $p = 0.004$), such that children allocated more punishment to the more deserving recipient with age.

4.4.3.2 Reasoning in Conjunction With Allocations

When analyzing the training trials (including Story 1) of the two conditions in which children were asked to justify themselves in the training phase, children's false belief competence was related to their frequency of allocating more to the more

deserving recipient in conjunction with providing a valid justification ($b = 0.42$, $SE = 0.18$, $t = 2.34$, $p = 0.02$). Namely, children with high false belief scores produced more of these responses compared to children with low false belief scores (false belief score of 0: $M = 0.59$, $SD = 0.80$; score of 1: $M = 1.04$, $SD = 1.11$; score of 2: $M = 1.45$, $SD = 1.22$). The relation between these two measures was not significant in the analysis of the distributive fairness test trials, which included the data of all the children from all four conditions, $p > 0.05$.

As for the analysis of the retributive fairness test trials, which also included the data from all four conditions, there was a significant relation between children's false belief competence and their frequency of allocating more to the more deserving recipient in conjunction with providing a valid justification ($b = 0.16$, $SE = 0.06$, $t = 2.79$, $p = 0.006$). Again, children with high false belief scores produced more of these responses compared to children with low false belief scores (false belief score of 0: $M = 0.21$, $SD = 0.47$; score of 1: $M = 0.32$, $SD = 0.47$; score of 2: $M = 0.60$, $SD = 0.61$). Note that in the analysis of the retributive fairness test trials (considered together), there was also a significant effect of age ($b = 0.24$, $SE = 0.11$, $t = 2.16$, $p = 0.03$), such that children allocated more punishment to the more deserving recipient in conjunction with providing valid reasoning with age. Also, note that when considering Story 11 and Story 12 separately, the effect of false belief was only significant in Story 12 ($b = 0.13$, $SE = 0.05$, $t = 2.73$, $p =$

0.007; false belief score of 0: $M = 0.13$, $SD = 0.34$; score of 1: $M = 0.27$, $SD = 0.45$; score of 2: $M = 0.42$, $SD = 0.50$), not Story 11, $p > 0.05$.

4.4.4 Ancillary Analyses

To begin, two correlation analyses were conducted. The first correlation analysis examined whether children's reasoning related to their age during the training trials (Stories 1, 2, 4, and 5). This first analysis only examined the data from the two conditions in which children were asked for justification during the training phase. The second correlation analysis examined whether children's reasoning related to their age during the test trials (Stories 7, 8, 11, and 12). This second analysis included the data from all four conditions, given that all of the children were asked for justification during the test trials.

The next set of analyses examined the effects of the condition manipulation on children's responses on Story 2 given their responses on Story 1. Two groups of children were of interest: those who had allocated equally on Story 1 ($n = 96$) and those who had allocated more to the more deserving recipient on Story 1 ($n = 28$). Within each of these two groups, the effects of the condition manipulation on children's allocations on Story 2 were examined. Accordingly, this analysis helped clarify the immediate impact of the kind of social interaction children experienced from the puppet. There was also a group of children who had allocated more to the less deserving recipient on Story 1, but the sample size ($n = 6$) was too small for analysis.

A third set of analyses repeated the first two primary sets of analyses reported in the Results section but only within the subset of children ($n = 96$) who had allocated equally on Story 1. The aim of these analyses was to examine the effects of the condition manipulation on children who had tangibly displayed an initial equality bias.

4.4.4.1 Correlations Between Age and Reasoning

There was not a significant correlation between children's age and whether they provided valid reasoning on the training trials, $r = 0.05$, $p = 0.45$. Similarly, there was not a significant correlation between children's age and whether they provided valid reasoning on the test trials, $r = 0.02$, $p = 0.57$.

4.4.4.2 Responses on the First and Second Stories

Within the subset of children who had allocated equally on Story 1, a 2×2 multiple regression crossing Disagreement (Agree, Disagree) and Justification (Do Not Justify, Justify) found a main effect of Disagreement ($b = 0.49$, $SE = 0.13$, $t = 3.74$, $p = 0.0003$) and an interaction between Disagreement and Justification ($b = -0.40$, $SE = 0.18$, $t = -2.19$, $p = 0.03$). To follow up on the interaction, pairwise t -tests with Bonferroni-corrected p -values were conducted. The post hoc tests revealed that when the puppet did not ask for justification, disagreement led children to allocate more to the more deserving recipient ($M = 0.57$, $SD = 0.51$) compared to agreement ($M = 0.08$, $SD = 0.28$), $p = 0.002$. But when the puppet did ask for justification, disagreement had no effect, $p = 1.00$. Within the subset of children who had allocated more to the more deserving

recipient on Story 1, another 2 x 2 multiple regression with the same predictors was conducted. However, this multiple regression found no significant effects, $ps > 0.05$.

4.4.4.3 Examining Children With an Equality Bias

Focusing on the subset of children who had allocated equally on Story 1, the first two primary sets of analyses reported in the Results section were repeated. All of the significant effects that were observed in the full sample were observed in the subset except in two cases. In these two cases, the p -value of an effect that was significant in the full sample was no longer significant in the subset, but the direction of the effect was consistent with what was observed in the full sample. Firstly, the main effect of Disagreement in the multiple regression examining only children's allocations on Story 12 was no longer significant, $p = 0.17$. Secondly, in the analysis of responses in which children both allocated more to the more deserving recipient and gave a valid justification for their allocation in the test trials about distributive fairness, the Bonferroni-corrected p -value for the contrast between the Simple Agreement condition and the Simple Disagreement condition was no longer significant, $p = 0.19$.

4.5 Discussion

The aim of the present study was to experimentally assess which features of social interactions have a positive impact on children's moral development. Children discussed what to do in simple moral scenarios with a puppet interlocutor, who either agreed or disagreed with the child's ideas and, moreover, either asked the child to

justify their ideas or not. The moral scenarios were about issues of fairness, that is, how to allocate various things between different recipients. Because children have a known developmental starting point in their thinking about fairness (namely, an inflexible bias to always prefer equal allocations, regardless of context), moral development could be operationalized concretely as a shift towards being able to make appropriate allocations and articulate valid justifications in cases where it would be fair, according to common ground norms and values, to give more to a more deserving recipient instead of giving equally. Children's allocations and reasoning were assessed both during the ongoing social interactions with the puppet (the "training phase") and in a different context that was identical for all children (the "test phase").

The results suggested that social experiences of disagreement and social experiences of being asked for justification can both be helpful for moral development. Experiences of disagreement had a positive effect on moral development not only during the training phase but also during the test phase. That is, children who had experienced disagreement in one social context (the training phase) showed improved moral development even in a different context that no longer featured disagreement (the test phase). In contrast, the positive effect of being asked for justification was limited to the training phase and did not carry over to the test phase. Potentially, this absence of a carryover effect may have been due to a methodological feature of the study: All children were asked for justification during the test phase. As such, the carryover effect

of training phase justification on test phase performance may have been obscured by the fact that all children experienced some effect of justification during the test phase.

The positive effect of disagreement as a catalyst of moral development accords with the Piagetian notion of disequilibrium (Walker et al. 2000; Piaget, 1954). According to the Piagetian view, when the equilibrium between one's knowledge about the world and the information coming in from the world itself is disrupted (i.e., disequilibrated) by novel information that contradicts one's prior knowledge, then one is prompted to revise one's schemas in order to make sense of reality. Our findings suggest that simple social interactions can be an effective vehicle of disequilibrium, which in turn stimulates development in moral thought. Indeed, children benefited from merely encountering someone disagree with them even in the absence of further engaging with that person to uncover their rationale or co-construct a consensus. This highlights how the mere experience of encountering disagreement from another person could already be a sufficient trigger of disequilibrium, possibly because it signals that there may be information that one has not yet considered.

The value of disagreement as a catalyst for development also accords with recent research from educational psychology. According to Kuhn and colleagues, one effective way for students to improve their critical thinking skills is to engage in argumentation against others who have opposing beliefs (Kuhn, 2015, 2018, 2019; Papathomas & Kuhn, 2017; Zillmer & Kuhn, 2018). Whereas students have a natural tendency to focus on

reasons and justifications that support what they already believe, the presence of an opposing viewpoint motivates students to look beyond what they already think and instead consider possible rebuttals to both their own and others' positions (Kuhn, 2019). Although this body of research focused on non-moral domains of critical thinking in older children and adolescents, it would be interesting to further explore whether argumentation could also be helpful for the development of moral reasoning in younger children.

One unexpected finding of our study was that whereas experiences of disagreement and experiences of being asked for justification were each independently helpful for moral development, they did not seem to be especially conducive when combined. Contrary to our hypotheses, the Resolving Disagreement condition (in which the puppet both disagreed with the child and asked the child for justification) did not result in the greatest moral development relative to the other conditions. The unexpectedly weak effect of this condition might be better explained by practical rather than theoretical reasons. One possibility is that this condition was simply too overwhelming for the children. In this condition, the puppet asked the child to justify themselves immediately after expressing disagreement with the child (in a relatively lengthy way that explicitly contrasted their own opinion with the child's opinion). This type of discourse may have been too taxing for the children, who may have needed

more time to process the puppet's disagreement before switching over to articulating a justification for their own opinion.

Aside from the social interaction conditions, false belief competence was also related to children's moral judgments and reasoning. In the training phase, children with better false belief competence were more likely to give more to the more deserving recipient in conjunction with providing a valid justification compared to children with worse false belief competence. In the test phase, false belief competence was also related to giving more to the more deserving recipient (and providing a valid justification for doing so). Given that children's justifications were situated in an inherently social context (in which children were justifying their views *to* another person), our results indicate that there is at least some relation between false belief competence and the capacity to participate in morality as a social activity (between-minds). This makes sense on conceptual grounds, since the process of reasoning with another person presupposes an understanding that the other person could be mistaken (i.e., have a false belief). To be sure, further research is needed on how false belief competence relates to the capacity to participate in morality as a social activity.

4.5.1 Limitations

There were some methodological and conceptual limitations of the present study. One limitation was that the social interactions involved a puppet interlocutor rather than a human peer or adult. The advantages of the puppet were that it enabled us

to precisely control the interaction as well as reduce perceptions of adult authority, but a disadvantage was that it lessened the ecological validity of the study. Another limitation of the study was the short time interval between the training phase and the test phase, which both occurred within the same session. Future studies that assess children over longer time periods would help clarify how social interactions influence moral development over larger timescales. A third limitation of the study was that its examination of moral development was restricted to issues of fairness, such as distributive fairness (e.g., how to allocate resources) and retributive fairness (e.g., how to allocate punishment). Although fairness is a central moral concern in children's lives, it is by no means the only domain of morality that children encounter. Future research could also examine, for instance, how social interactions affect children's moral judgments and reasoning about social inclusion and exclusion, group norms and practices, or aggressive behavior (see Gray & Graham, 2018).

4.6 Conclusion

Morality is a social activity that humans co-construct and participate in cooperatively (Li & Tomasello, 2021). Moral development, in turn, is the process of learning how to participate in the social activity that is morality. Many theorists have proposed that social interactions are a key influence on children's moral thought, but there has been a surprising lack of experimental research examining which particular features of social interactions are effective at promoting moral development. Using an

experimental design that manipulated the types of social interactions children experienced with a puppet while discussing what to do in simple moral scenarios, the present study found evidence that social experiences of being disagreed with, as well as social experiences of being asked to justify oneself, are both helpful for moral development (in the form of overcoming an inflexible equality bias and acting in accordance with common ground norms and values of fairness). These findings may help shed light on the broader question of how humans have been uniquely able to shape each other, via socialization, into cooperative and moral beings.

Chapter 5. Conclusion

The three research chapters of this dissertation were linked by a unifying theme: the interconnectedness of language and morality during ontogeny. As described in Chapter 2, language and morality had shared evolutionary origins as forms of cooperative social action. Both are expressions of the human social cognitive capacity to engage in shared intentionality (i.e., to align, exchange, and interact with others' mental states), which evolved as an adaptation for the evolutionary demands of obligate collaboration. Chapter 2 also described previous empirical studies into four moral functions of language, which are operative even in young children: initiating morality (e.g., forming joint commitments), preserving morality (e.g., teaching norms to novices), revising morality (e.g., changing existing norms), and acting on morality (e.g., engaging in reason-giving and justification about moral issues). Building on the theoretical foundations established in Chapter 2, the next two chapters described novel empirical investigations into specific moral functions of language.

Chapter 3 focused on the use of language to signal what is socially normative behavior. The study reported in Chapter 3 revealed that young children conformed more to the choices of another person when those choices were framed as conventional norms, as opposed to personal preferences. Notably, this effect was found in children who were only 3.5 years old, suggesting that linguistic cues can signal normativity even at a very young age. Moreover, this effect was significant whether the model who stated

the norms and preferences was an adult or a 6-year-old girl, which suggests that it was truly the normativity of the message that mattered, not just the authority of the messenger. This study may help motivate further inquiries into the essential function of language as a means of signaling normativity, a research topic that has been relatively understudied.

Chapter 4 focused on a domain of morality that is inherently linguistically mediated: moral reasoning as a social activity. Moral reasoning, in this sense, refers to how individuals engage in reason-giving and justification when discussing what would be the right thing to do. The aim of the study reported in Chapter 4 was to identify features of social interactions that may be helpful for children's moral development (operationalized in terms of children's abilities to make fair moral decisions and justify those decisions to others). Two helpful features were identified. Specifically, it was found that (i) experiences of being disagreed with and (ii) experiences of being asked to justify oneself were both helpful for prompting moral development. This study may help motivate further inquiries into how exactly social interactions facilitate moral development.

Overall, Chapter 3 and Chapter 4 both contributed findings in support of the theme established in Chapter 2: the interconnectedness of language and morality during ontogeny. At bottom, language and morality are essentially elaborated versions of the more fundamental human capacities to communicate (i.e., influence one another's

mental states) and engage in social normativity (i.e., form expectations about how individuals should treat one another). These two capacities, simple as they may seem, underlie the most monumental achievements of human culture, from the physical and social realities that humans have constructed to the vast bodies of knowledge and wisdom that humans have stored in their linguistic and moral traditions. The developmental trajectories of these two capacities are inherently interrelated, as language functions in important ways to sustain moral cognition and behavior. Ultimately, what this dissertation aims to have shown is that language and morality—two of the most magnificent expressions of human cooperation—have intertwined roots in evolution and ontogeny.

Appendix A. Children’s Initial Preferences and Subsequent Choices

According to the data, children’s indications of which items they felt like using seemed to correspond to their actual preferences. For each item option, we report the number of children who initially indicated that they felt like using that option as well as the number and percentage of those children who actually chose that option, consistent with their indication (Table 6). For instance, 8 children indicated that they felt like using the zebra plate. Of those 8 children, 5 children (63%) actually chose the zebra plate. Children chose what they indicated 62% of the time, which was a majority of the time and higher than expected at a chance level of 25%, $\chi^2(1) = 304.05, p < 0.01$. Item options are listed in the order in which they were presented by the informant.

Table 6: Children’s initial preferences and subsequent choices.

Option (* indicates option endorsed by informant)	# of children who indicated a preference for the option	# of children who chose what they indicated (per option)	% of children who chose what they indicated (per option)
Plates			
zebra	8	5	63%
rainbow	83	62	75%
round white	10	6	60%
square white*	2	2	100%
forgot to ask	1	N/A	N/A
Cups			
smiley face	26	10	38%
Frozen	52	36	69%
red	12	4	33%
blue*	14	11	79%

Teas			
apple	56	37	66%
orange	31	18	58%
celery	8	2	25%
potato*	9	6	67%
Snacks			
donut	48	35	73%
cookie	41	18	44%
egg	11	4	36%
veggie*	4	2	50%

Appendix B. Protest Measure

Method

In addition to the conformity measure, our study also included a second measure assessing whether children would protest when a squirrel puppet deviated from the informant's endorsements. The host first introduced the child to the squirrel puppet during their warm-up in the greeting room prior to going to the tea party room together. Later, the assessment of protest occurred after all the conformity trials were completed. During this protest phase, the host informed the child that a squirrel puppet (operated by the host) would also set up for the tea party. The puppet then chose items that differed from the informant's prior endorsements (as well as from the child's choices in cases in which the child had deviated from the informant's endorsement). The puppet gave the child two opportunities to object—first by saying, "Maybe I'll use this one. . ." and then, if the child had not objected, "Should I use this one?"

A score of 0 (representing no protest) was given if the child did not protest. Children also scored 0 if they protested but endorsed an item other than what the informant had endorsed. We reasoned that such cases may not accurately represent protest in service of upholding a norm, since the child could have simply wanted the puppet to fulfill their own preferences. A score of 1 (representing protest) was given only if the child advised the squirrel puppet to choose the item that the informant had endorsed.

Results

For inter-rater reliability, a second coder viewed 25% of the sessions ($n = 26$) and coded whether the child protested or not. This second coder had 100% agreement with the previously coded data. Rates of protest were very low. In only 23 instances (6% of 416 possible opportunities to protest) did children advise the puppet to choose the item that the informant had endorsed. These 23 cases included 16 protests against deviations from norms and 7 protests against deviations from preferences. The difference between the number of protests against deviations from norms and the number of protests against deviations from preferences was not significant, as indicated by a two-tailed binomial test, $p = 0.09$.

Discussion

Overall, our protest measure appeared to have not worked, given that children rarely protested. This finding was not too unexpected, given that our task manipulation was a subtle linguistic framing. It was interesting, then, that the same linguistic cues that were not strong enough to lead children to correct how others behave were nonetheless strong enough to sway children's own behavior. Here, we discuss three plausible interpretations for why protest behaviors were low compared to in previous studies. One interpretation of this finding is that the perceived force of a norm may reside on a gradient, such that some norms exert enough pressure that one feels compelled to follow them but not enough pressure that one feels compelled to enforce them on others.

A second interpretation pertains to children's memory capacities. Children encountered the protest phase after having experienced four conformity trials in which the informant endorsed one of four options for a tea party item. That is, children observed the informant endorse four of sixteen different options for the tea party over four different trials. It is possible that by the time children encountered the protest phase, they did not clearly recall which items the informant had endorsed, let alone which items the informant had endorsed with norms and which with preferences. Future research could potentially reduce memory demands by presenting individual protest trials after individual conformity trials rather than together in a block at the end of the study.

A third interpretation pertains to a more practical—but, in our view, not insignificant—aspect of our study design. In most studies that include protest against a puppet as a dependent measure, there is a warm-up phase in which a puppet acts in ways that invite intervention (e.g., in a clumsy or clearly erroneous manner). In this warm-up phase, children have the opportunity to practice correcting the puppet. Accordingly, by the time children encounter the test phase, they are primed to think of the puppet as someone they can object to. However, our warm-up phase did not involve the puppet acting in ways that invited intervention. In our warm-up phase, the puppet simply introduced themselves to the child. As a result, the children in our study may

have been less prepared—compared to children in other studies—to object to the puppet when it acted in ways that deviated from the endorsements of the informant. Other researchers may learn from our experience and be sure to include a warm-up phase to encourage higher rates of protesting.

Appendix C. Additional Analyses

We conducted additional exploratory analyses to examine (i) the possible effects of counterbalancing Order (Preference First, Norm First) and (ii) whether the main effect of Endorsement held in both the adult informant and the child informant conditions separately. Overall, all analyses were consistent with the results that we reported in our main text. Order did have a significant interaction with Endorsement, but it was in a way that was still consistent with our interpretation of the main effect of Endorsement. Most importantly, Order did not interact with Informant and did not alter the (lack of an) interaction between Informant and Endorsement. In other words, our main conclusion (that children conformed more to norms than to preferences whether the informant was an adult or a child) remained intact even after accounting for the effects of Order. Moreover, the main effect of Endorsement was significant in both the adult informant and the child informant conditions separately.

Main Effect of Order

To assess the effects of Order, we ran and compared a series of models. Our baseline model that served as a basis for model comparison was a null model that only included the random intercept of Participant. The simplest model of interest added the main effect of Order to the null model. The condition “Preference First” was chosen as the reference level for Order, as we expected a priori that children would conform less in this case. However, this model containing the main effect of Order did not lead to a

significant improvement in fit compared to the null model, $\chi^2(1) = 2.59, p = 0.11$, and also did not find a significant main effect of Order ($b = 0.20, SE = 0.12, t = 1.62, p = 0.11$).

Crossing Order and Informant

In a second model, we examined whether Order interacted with Informant. In this model, we added the main effects of Order and Informant as well as their interaction to the null model. However, this model also did not lead to a significant improvement in fit compared to the null model, $\chi^2(3) = 6.02, p = 0.11$. What is more, this model showed no significant main effect of Order ($b = 0.24, SE = 0.17, t = 1.40, p = 0.17$), no significant main effect of Informant ($b = 0.25, SE = 0.17, t = 1.47, p = 0.14$), and no interaction between Order and Informant ($b = -0.06, SE = 0.24, t = -0.26, p = 0.79$).

Crossing Order and Endorsement

In a third model, we examined whether Order would interact with Endorsement. Thus, in this model, we added the main effects of Order and Endorsement as well as their interaction to the null model (Table 7a). Unlike the first two models, this model did lead to a significant improvement in fit compared to the null model, $\chi^2(3) = 12.24, p = 0.007$. There was a significant main effect of Order ($b = 0.36, SE = 0.14, t = 2.47, p = 0.01$): Children conformed more when they heard norms first than when they heard preferences first. There was also a significant main effect of Endorsement ($b = 0.36, SE = 0.11, t = 3.16, p = 0.002$): Children conformed more to norms than to preferences. These

main effects were qualified by a significant interaction between Order and Endorsement ($b = -0.32, SE = 0.16, t = -2.04, p = 0.04$).

The interaction revealed that when children heard norms first, they conformed at similar rates to the norms ($M = 0.57$) as to the subsequent preferences ($M = 0.54$). But when children heard preferences first, they conformed more to norms ($M = 0.54$) than to preferences ($M = 0.18$). To confirm whether the difference between conformity to norms and conformity to preferences was significant within the subset of children who heard preferences first, we ran another model examining only the main effect of Endorsement in the subset of children who had preferences first (Table 7b). This model led to a significant improvement in fit compared to a null model that only included the random intercept of participant (for the subset of children who had preferences first), $\chi^2(1) = 11.31, p = 0.0008$, and also confirmed that the main effect of Endorsement was significant within this subset ($b = 0.36, SE = 0.10, t = 3.56, p = 0.0008$).

Importantly, the interaction between Order and Endorsement was still consistent with our hypothesis that children are motivated to conform in response to cues of conventionality. That is, hearing norms first may have led children to construe the entire play situation as normative, including the preferences that were stated later. In such cases, when children heard the preferences stated by the informant who had previously established themselves as a cultural representative, they may have interpreted the informant's preferences to also be expressions of norms. But when children heard

preferences first and norms only later, they may not have begun to construe the situation as normative until after they had freely acted on their own preferences. Hence, conformity to norms was greater than conformity to preferences within this latter subset.

Table 7: Effects of Order and Endorsement. Panel A: Summary of the linear mixed effects model of conformity as predicted by Order (Preference First, Norm First) crossed with Endorsement (Preference, Norm). Panel B: Summary of the linear mixed effects model of conformity as predicted by Endorsement (Preference, Norm) within the subset of children who heard preferences first. * $p < 0.05$; ** $p < 0.01$.

A. Formula: Conformity ~ Order * Endorsement + (1 Participant)					
<i>Model fit:</i>	AIC	BIC	logLik	deviance	df.resid
	456.2	476.2	-222.1	444.2	202
<i>Random effects:</i>	Variance	Std. Dev.			
Participant	0.2163	0.4650			
Residual	0.3243	0.5694			
<i>Fixed effects:</i>	Estimate	Std. Error	df	t value	Pr(> t)
(Intercept)	0.1800	0.1040	179.2973	1.731	0.0851
Order [Norm]	0.3570	0.1443	179.2973	2.474	0.0143*
Endorsement [Norm]	0.3600	0.1139	104.0000	3.161	0.0021**
Order [Norm] x Endorsement [Norm]	-0.3230	0.1581	104.0000	-2.043	0.0435*
B. Formula: Conformity ~ Endorsement + (1 Participant)					
<i>Model fit:</i>	AIC	BIC	logLik	deviance	df.resid
	202.8	213.2	-97.4	194.8	96
<i>Random effects:</i>	Variance	Std. Dev.			

Participant	0.2028	0.4503			
Residual	0.2552	0.5052			
<hr/>					
<i>Fixed effects:</i>	Estimate	Std. Error	df	t value	Pr(> t)
(Intercept)	0.1800	0.0957	83.6073	1.881	0.0635
Endorsement [Norm]	0.3600	0.1010	50.0000	3.563	0.0008**

Crossing Order, Informant, and Endorsement

In a fourth model, we examined whether the (lack of an) interaction between Informant and Endorsement would still hold after accounting for Order. This was important to verify, as one of our key conclusions was that children conformed more to norms than to preferences whether the informant was an adult or a child. Thus, we included the main effects and interactions of all three variables: Order, Informant, and Endorsement. However, this fourth model did not lead to a significant improvement in fit compared to our third model (i.e., the model crossing Order and Endorsement), $\chi^2(4) = 4.22, p = 0.38$, so our third model remained the most parsimonious explanation of the data.

Nonetheless, the results of this fourth model were still consistent with our conclusions. The only significant effects in this model were a main effect of Order ($b = 0.43, SE = 0.20, t = 2.12, p = 0.04$) and a main effect of Endorsement ($b = 0.46, SE = 0.16, t = 2.80, p = 0.006$). Importantly, the interaction between Informant and Endorsement was not significant ($b = -0.19, SE = 0.23, t = -0.83, p = 0.41$), and the interaction between Order,

Informant, and Endorsement was also not significant ($b = 0.12$, $SE = 0.32$, $t = 0.37$, $p = 0.71$). Thus, even when accounting for Order and its interactions with the other variables, our main conclusion remained intact: Children conformed more to norms than to preferences whether the informant was an adult or a child.

The Adult and Child Informant Conditions

A final question of interest was whether the main effect of Endorsement would hold in the separate subsets of children who heard from the adult and child informants. Given that the main effect of Endorsement was found to be qualified by an interaction between Order and Endorsement in our third model, we limited our analyses to the subset of children who heard preferences first, since the subset of children who heard norms first had restricted variability in the differences between their rates of conformity to the norms and the preferences. Focusing on the subset of children who had the adult informant and preferences first, we ran a model that only included the main effect of Endorsement (Table 8). This model led to a significant improvement in fit compared to a null model that only included the random intercept of participant (for this subset of children), $\chi^2(1) = 4.07$, $p = 0.04$, and also confirmed that the main effect of Endorsement was significant within this subset ($b = 0.27$, $SE = 0.13$, $t = 2.10$, $p = 0.046$). Namely, these children conformed more to norms ($M = 0.62$) than to preferences ($M = 0.35$).

Table 8: Effect of Endorsement. Summary of the linear mixed effects model of conformity as predicted by Endorsement (Preference, Norm) within the subset of children who heard preferences first from the adult informant. * $p < 0.05$; ** $p < 0.01$.

Formula: Conformity ~ Endorsement + (1 Participant)					
<hr/>					
<i>Model fit:</i>	AIC	BIC	logLik	deviance	df.resid
	113.9	121.7	-52.9	105.9	48
<hr/>					
<i>Random effects:</i>	Variance	Std. Dev.			
Participant	0.3639	0.6032			
Residual	0.2138	0.4623			
<hr/>					
<i>Fixed effects:</i>	Estimate	Std. Error	df	t value	Pr(> t)
(Intercept)	0.3462	0.1491	37.2266	2.322	0.0258*
Endorsement [Norm]	0.2692	0.1282	26.0000	2.100	0.0456*

Next, focusing on the subset of children who had the child informant and preferences first, we ran a model that only included the main effect of Endorsement. However, the model failed to converge due to having an insufficient number of observations. Whereas in the adult informant condition, our sample included a perfect balance of children who heard preferences first ($n = 26$) and children who heard norms first ($n = 26$), in the child informant condition, we had slightly fewer children who heard preferences first ($n = 24$) than children who heard norms first ($n = 28$) due to imperfect counterbalancing during recruitment. Thus, in lieu of running a model, we conducted a

paired-samples Wilcoxon signed-rank test instead. This Wilcoxon signed-rank test nonetheless confirmed that children in this subset conformed significantly more to norms ($M = 0.46$) than to preferences ($M = 0.00$), $Z = -2.34$, $p = 0.02$, $r = 0.34$. Indeed, the children in this subset actually never conformed to the preferences at all.

References

- Acemoglu, D., & Robinson, J. A. (2012). *Why nations fail: The origins of power, prosperity, and poverty*. New York, NY: Crown Business.
- Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1–48.
- Bicchieri, C. (2006). *The grammar of society: The nature and dynamics of social norms*. New York, NY: Cambridge University Press.
- Bybee, J. L., & Beckner, C. (2010). Usage-based theory. In B. Heine & H. Narrog (Eds.), *The Oxford handbook of linguistic analysis* (1st ed., pp. 827–855). Oxford, UK: Oxford University Press.
- Carpenter, M., Akhtar, N., & Tomasello, M. (1998). Fourteen- through 18-month-old infants differentially imitate intentional and accidental actions. *Infant Behavior & Development*, *21*, 315–330.
- Carpenter, M., Uebel, J., & Tomasello, M. (2013). Being mimicked increases prosocial behavior in 18-month-old infants. *Child Development*, *84*, 1511–1518.
- Chomsky, N. (1967). Recent contributions to the theory of innate ideas. *Synthese*, *17*, 2–11.
- Chwe, M. S.-Y. (2001). *Rational ritual: Culture, coordination, and common knowledge*. Princeton, NJ: Princeton University Press.
- Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition*, *22*, 1–39.
- Clegg, J. M., & Legare, C. H. (2016a). A cross-cultural comparison of children’s imitative flexibility. *Developmental Psychology*, *52*, 1435–1444.
- Clegg, J. M., & Legare, C. H. (2016b). Instrumental and conventional interpretations of behavior are associated with distinct outcomes in early childhood. *Child Development*, *87*, 527–542.
- Damon, W., & Killen, M. (1982). Peer interaction and the process of change in children’s moral reasoning. *Merrill-Palmer Quarterly*, *28*, 347–367.

- DeScioli, P., & Kurzban, R. (2018). Morality is for choosing sides. In K. Gray & J. Graham (Eds.), *Atlas of moral psychology* (pp. 177–185). New York, NY: The Guilford Press.
- Diesendruck, G., & Markson, L. (2011). Children's assumption of the conventionality of culture. *Child Development Perspectives, 5*, 189–195.
- Dijker, A. J. M. (2018). Vulnerability-based morality. In K. Gray & J. Graham (Eds.), *Atlas of moral psychology* (pp. 430–439). New York, NY: The Guilford Press.
- Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*, 1149–1160.
- Fehr, E., Bernhard, H., & Rockenbach, B. (2008). Egalitarianism in young children. *Nature, 454*, 1079–1083.
- Gert, B. (2004). *Common morality: Deciding what to do*. New York, NY: Oxford University Press.
- Gibbard, A. (1990). Norms, discussion, and ritual: Evolutionary puzzles. *Ethics, 100*, 787–802.
- Gilbert, M. (1990). Walking together: A paradigmatic social phenomenon. *Midwest Studies in Philosophy, 15*, 1–14.
- Göckeritz, S., Schmidt, M. F. H., & Tomasello, M. (2014). Young children's creation and transmission of social norms. *Cognitive Development, 30*, 81–95.
- Goodman, N. D., & Frank, M. C. (2016). Pragmatic language interpretation as probabilistic inference. *Trends in Cognitive Sciences, 20*, 818–829.
- Gräfenhain, M., Behne, T., Carpenter, M., & Tomasello, M. (2009). Young children's understanding of joint commitments. *Developmental Psychology, 45*, 1430–1443.
- Gray, K., & Graham, J. (Eds.). (2018). *Atlas of moral psychology*. New York, NY: The Guilford Press.
- Grice, P. (1989). *Studies in the way of words*. Cambridge, MA: Harvard University Press.

- Groce, P., Rossano, F., & Tomasello, M. (2015). Procedural justice in children: Preschoolers accept unequal resource distributions if the procedure provides equal opportunities. *Journal of Experimental Child Psychology, 140*, 197–210.
- Haidt, J., & Graham, J. (2007). When morality opposes justice: Conservatives have moral intuitions that liberals may not recognize. *Social Justice Research, 20*, 98–116.
- Hamlin, J. K., & Van de Vondervoort, J. W. (2018). Infants' and young children's preferences for prosocial over antisocial others. *Human Development, 61*, 214–231.
- Hardecker, S., Schmidt, M. F. H., & Tomasello, M. (2017). Children's developing understanding of the conventionality of rules. *Journal of Cognition and Development, 18*, 163–188.
- Haun, D. B. M., Rekers, Y., & Tomasello, M. (2014). Children conform to the behavior of peers; other great apes stick with what they know. *Psychological Science, 25*, 2160–2167.
- Haux, L., Engelmann, J. M., Herrmann, E., & Tomasello, M. (2017). Do young children preferentially trust gossip or firsthand observation in choosing a collaborative partner? *Social Development, 26*, 466–474.
- Holtgraves, T. M. (2002). *Language as social action: Social psychology and language use*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Kachel, U., Svetlova, M., & Tomasello, M. (2018). Three-year-olds' reactions to a partner's failure to perform her role in a joint commitment. *Child Development, 89*, 1691–1703.
- Kachel, U., & Tomasello, M. (2019). 3- and 5-year-old children's adherence to explicit and implicit joint commitments. *Developmental Psychology, 55*, 80–88.
- Kalish, C. W., & Shiverick, S. M. (2004). Children's reasoning about norms and traits as motives for behavior. *Cognitive Development, 19*, 401–416.
- Kanngiesser, P., Köymen, B., & Tomasello, M. (2017). Young children mostly keep, and expect others to keep, their promises. *Journal of Experimental Child Psychology, 159*, 140–158.

- Kant, I. (2004). *Critique of practical reason* (T. K. Abbott, Trans.). Mineola, NY: Dover Publications. (Original work published 1788)
- Kenward, B. (2012). Over-imitating preschoolers believe unnecessary actions are normative and enforce their performance by a third party. *Journal of Experimental Child Psychology, 112*, 195–207.
- Keupp, S., Behne, T., & Rakoczy, H. (2013). Why do children overimitate? Normativity is crucial. *Journal of Experimental Child Psychology, 116*, 392–406.
- Killen, M., Breton, S., Ferguson, H., & Handler, K. (1994). Preschoolers' evaluations of teacher methods of intervention in social transgressions. *Merrill-Palmer Quarterly, 40*, 399–415.
- Killen, M., & Cords, M. (2002). Prince Kropotkin's ghost. *American Scientist, 90*, 208–210.
- Killen, M., & Dahl, A. (2018). Moral judgment: Reflective, interactive, spontaneous, challenging, and always evolving. In K. Gray & J. Graham (Eds.), *Atlas of moral psychology* (pp. 20–30). New York, NY: The Guilford Press.
- Killen, M., Mulvey, K. L., Richardson, C., Jampol, N., & Woodward, A. (2011). The accidental transgressor: Morally-relevant theory of mind. *Cognition, 119*, 197–215.
- Killen, M., Rutland, A., Abrams, D., Mulvey, K. L., & Hitti, A. (2013). Development of intra- and intergroup judgments in the context of moral and social-conventional norms. *Child Development, 84*, 1063–1080.
- Kohlberg, L. (1973). The claim to moral adequacy of a highest stage of moral judgment. *The Journal of Philosophy, 70*, 630–646.
- Köymen, B., Jurkat, S., & Tomasello, M. (2020). Preschoolers refer to direct and indirect evidence in their collaborative reasoning. *Journal of Experimental Child Psychology, 193*, 104806.
- Köymen, B., Mammen, M., & Tomasello, M. (2016). Preschoolers use common ground in their justificatory reasoning with peers. *Developmental Psychology, 52*, 423–429.
- Köymen, B., O'Madagain, C., Domberg, A., & Tomasello, M. (2020). Young children's ability to produce valid and relevant counter-arguments. *Child Development, 91*, 685–693.

- Köymen, B., Rosenbaum, L., & Tomasello, M. (2014). Reasoning during joint decision-making by preschool peers. *Cognitive Development, 32*, 74–85.
- Köymen, B., & Tomasello, M. (2020). The early ontogeny of reason giving. *Child Development Perspectives, 14*, 215–220.
- Kruger, A. C. (1992). The effect of peer and adult-child transactive discussions on moral reasoning. *Merrill-Palmer Quarterly, 38*, 191–211.
- Kruger, A. C. (1993). Peer collaboration: Conflict, cooperation, or both? *Social Development, 2*, 165–182.
- Kruger, A. C., & Tomasello, M. (1986). Transactive discussions with peers and adults. *Developmental Psychology, 22*, 681–685.
- Kuhn, D. (2015). Thinking together and alone. *Educational Researcher, 44*, 46–53.
- Kuhn, D. (2018). A role for reasoning in a dialogic approach to critical thinking. *Topoi, 37*, 121–128.
- Kuhn, D. (2019). Critical thinking as discourse. *Human Development, 62*, 146–164.
- Lane, J. D., Wellman, H. M., Olson, S. L., LaBounty, J., & Kerr, D. C. R. (2010). Theory of mind and emotion understanding predict moral development in early childhood. *British Journal of Developmental Psychology, 28*, 871–889.
- Langacker, R. W. (2013). *Essentials of cognitive grammar*. New York, NY: Oxford University Press.
- Legare, C. H., & Nielsen, M. (2015). Imitation and innovation: The dual engines of cultural learning. *Trends in Cognitive Sciences, 19*, 688–699.
- Li, L., Britvan, B., & Tomasello, M. (2021). Young children conform more to norms than to preferences. *PLOS ONE, 16*, e0251228.
- Li, L., Rizzo, M. T., Burkholder, A. R., & Killen, M. (2017). Theory of mind and resource allocation in the context of hidden inequality. *Cognitive Development, 43*, 25–36.

- Li, L., & Tomasello, M. (2021). On the moral functions of language. *Social Cognition, 39*, 99–116.
- Mammen, M., Köymen, B., & Tomasello, M. (2018). The reasons young children give to peers when explaining their judgments of moral and conventional rules. *Developmental Psychology, 54*, 254–262.
- Mammen, M., Köymen, B., & Tomasello, M. (2019). Children's reasoning with peers and parents about moral dilemmas. *Developmental Psychology, 55*, 2324–2335.
- McGuigan, N., & Robertson, S. (2015). The influence of peers on the tendency of 3- and 4-year-old children to over-imitate. *Journal of Experimental Child Psychology, 136*, 42–54.
- Mercier, H., & Sperber, D. (2011). Why do humans reason? Arguments for an argumentative theory. *Behavioral and Brain Sciences, 34*, 57–111.
- Mikhail, J. (2007). Universal moral grammar: Theory, evidence and the future. *Trends in Cognitive Sciences, 11*, 143–152.
- Nadelhoffer, T., Nahmias, E., & Nichols, S. (Eds.). (2010). *Moral psychology: Historical and contemporary readings*. Malden, MA: Wiley-Blackwell.
- Nichols, S. (2018). The wrong and the bad. In K. Gray & J. Graham (Eds.), *Atlas of moral psychology* (pp. 40–45). New York, NY: The Guilford Press.
- Nichols, S., Kumar, S., Lopez, T., Ayars, A., & Chan, H.-Y. (2016). Rational learners and moral rules. *Mind & Language, 31*, 530–554.
- Papathomas, L., & Kuhn, D. (2017). Learning to argue via apprenticeship. *Journal of Experimental Child Psychology, 159*, 129–139.
- Paulus, M. (2014). The emergence of prosocial behavior: Why do infants and toddlers help, comfort, and share? *Child Development Perspectives, 8*, 77–81.
- Pettit, P. (2018). *The birth of ethics: Reconstructing the role and nature of morality*. New York, NY: Oxford University Press.
- Piaget, J. (1932). *The moral judgment of the child* (M. Gabain, Trans.). New York, NY: Harcourt, Brace and Company.

- Piaget, J. (1954). *The construction of reality in the child* (M. Cook, Trans.). New York, NY: Basic Books.
- Pinker, S. (2012). *The better angels of our nature: Why violence has declined*. New York, NY: Penguin Books.
- Prinz, J. J. (2018). The history of moral norms. In K. Gray & J. Graham (Eds.), *Atlas of moral psychology* (pp. 266–276). New York, NY: The Guilford Press.
- Rai, T. S., & Fiske, A. P. (2011). Moral psychology is relationship regulation: Moral motives for unity, hierarchy, equality, and proportionality. *Psychological Review*, *118*, 57–75.
- Rakoczy, H. (2008). Taking fiction seriously: Young children understand the normative structure of joint pretence games. *Developmental Psychology*, *44*, 1195–1201.
- Rakoczy, H., & Schmidt, M. F. H. (2013). The early ontogeny of social norms. *Child Development Perspectives*, *7*, 17–21.
- Rhodes, M., Leslie, S.-J., & Tworek, C. M. (2012). Cultural transmission of social essentialism. *Proceedings of the National Academy of Sciences*, *109*, 13526–13531.
- Rizzo, M. T., Elenbaas, L., Cooley, S., & Killen, M. (2016). Children's recognition of fairness and others' welfare in a resource allocation task: Age related changes. *Developmental Psychology*, *52*, 1307–1317.
- Rizzo, M. T., Li, L., Burkholder, A. R., & Killen, M. (2019). Lying, negligence, or lack of knowledge? Children's intention-based moral reasoning about resource claims. *Developmental Psychology*, *55*, 274–285.
- Roberts, S. O., Ho, A. K., & Gelman, S. A. (2017). Group presence, category labels, and generic statements influence children to treat descriptive group regularities as prescriptive. *Journal of Experimental Child Psychology*, *158*, 19–31.
- Scanlon, T. M. (2008). *Moral dimensions: Permissibility, meaning, blame*. Cambridge, MA: Harvard University Press.

- Schein, C., & Gray, K. (2018). Moralization: How acts become wrong. In K. Gray & J. Graham (Eds.), *Atlas of moral psychology* (pp. 363–370). New York, NY: The Guilford Press.
- Schelling, T. C. (1980). *The strategy of conflict*. Cambridge, MA: Harvard University Press.
- Schmidt, M. F. H., 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*, 1360–1370.
- Schmidt, M. F. H., Gonzalez-Cabrera, I., & Tomasello, M. (2017). Children’s developing metaethical judgments. *Journal of Experimental Child Psychology, 164*, 163–77.
- Schmidt, M. F. H., Rakoczy, H., & Tomasello, M. (2011). Young children attribute normativity to novel actions without pedagogy or normative language. *Developmental Science, 14*, 530–539.
- Schmidt, M. F. H., Rakoczy, H., & Tomasello, M. (2012). Young children enforce social norms selectively depending on the violator’s group affiliation. *Cognition, 124*, 325–333.
- Schmidt, M. F. H., & Tomasello, M. (2012). Young children enforce social norms. *Current Directions in Psychological Science, 21*, 232–236.
- Searle, J. R. (2001). *Rationality in action*. Cambridge, MA: The MIT Press.
- Shafto, P., Goodman, N. D., & Griffiths, T. L. (2014). A rational account of pedagogical reasoning: Teaching by, and learning from, examples. *Cognitive Psychology, 71*, 55–89.
- Shaw, A., & Olson, K. R. (2012). Children discard a resource to avoid inequity. *Journal of Experimental Psychology: General, 141*, 382–395.
- Sigelman, C. K., & Waitzman, K. A. (1991). The development of distributive justice orientations: Contextual influences on children’s resource allocations. *Child Development, 62*, 1367–1378.
- Sinnott-Armstrong, W. (2018). Asking the right questions in moral psychology. In K. Gray & J. Graham (Eds.), *Atlas of moral psychology* (pp. 565–571). New York, NY: The Guilford Press.

- Siposova, B., Tomasello, M., & Carpenter, M. (2018). Communicative eye contact signals a commitment to cooperate for young children. *Cognition*, *179*, 192–201.
- Smetana, J. G., & Asquith, P. (1994). Adolescents' and parents' conceptions of parental authority and personal autonomy. *Child Development*, *65*, 1147–1162.
- 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). New York, NY: Psychology Press.
- Sripada, C. S. (2005). Punishment and the strategic structure of moral systems. *Biology and Philosophy*, *20*, 767–789.
- Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.
- Tomasello, M. (2008). *Origins of human communication*. Cambridge, MA: The MIT Press.
- Tomasello, M. (2014). *A natural history of human thinking*. Cambridge, MA: Harvard University Press.
- Tomasello, M. (2016a). *A natural history of human morality*. Cambridge, MA: Harvard University Press.
- Tomasello, M. (2016b). The ontogeny of cultural learning. *Current Opinion in Psychology*, *8*, 1–4.
- Tomasello, M. (2018). The normative turn in early moral development. *Human Development*, *61*, 248–263.
- Tomasello, M. (2019). *Becoming human: A theory of ontogeny*. Cambridge, MA: Harvard University Press.
- Tomasello, M., & Gonzalez-Cabrera, I. (2017). The role of ontogeny in the evolution of human cooperation. *Human Nature*, *28*, 274–288.
- Tomasello, M., Melis, A. P., Tennie, C., Wyman, E., & Herrmann, E. (2012). Two key steps in the evolution of human cooperation: The interdependence hypothesis. *Current Anthropology*, *53*, 673–692.

- Tomasello, M., & Vaish, A. (2013). Origins of human cooperation and morality. *Annual Review of Psychology, 64*, 231–255.
- Turiel, E. (2014). Morality: Epistemology, development, and social opposition. In M. Killen & J. G. Smetana (Eds.), *Handbook of moral development* (2nd ed., pp. 3–22). New York, NY: Psychology Press.
- Turiel, E. (2018). Reasoning at the root of morality. In K. Gray & J. Graham (Eds.), *Atlas of moral psychology* (pp. 9–19). New York, NY: The Guilford Press.
- Vaish, A., Missana, M., & Tomasello, M. (2011). Three-year-old children intervene in third-party moral transgressions. *British Journal of Developmental Psychology, 29*, 124–130.
- Vaish, A., & Tomasello, M. (2014). The early ontogeny of human cooperation and morality. In M. Killen & J. G. Smetana (Eds.), *Handbook of moral development* (2nd ed., pp. 279–298). New York, NY: Psychology Press.
- Walker, L. J., Hennig, K. H., & Krettenauer, T. (2000). Parent and peer contexts for children's moral reasoning development. *Child Development, 71*, 1033–1048.
- Warneken, F., Chen, F., & Tomasello, M. (2006). Cooperative activities in young children and chimpanzees. *Child Development, 77*, 640–663.
- Warneken, F., & Tomasello, M. (2006). Altruistic helping in human infants and young chimpanzees. *Science, 311*, 1301–1303.
- Wilson, D., & Sperber, D. (2012). *Meaning and relevance*. Cambridge, UK: Cambridge University Press.
- Wyman, E., Rakoczy, H., & Tomasello, M. (2009). Normativity and context in young children's pretend play. *Cognitive Development, 24*, 146–155.
- Yang, C., Crain, S., Berwick, R. C., Chomsky, N., & Bolhuis, J. J. (2017). The growth of language: Universal Grammar, experience, and principles of computation. *Neuroscience & Biobehavioral Reviews, 81*, 103–119.
- Yucel, M., & Vaish, A. (2018). Young children tattle to enforce social norms. *Social Development, 27*, 924–936.

- Zhao, X., & Kushnir, T. (2018). Young children consider individual authority and collective agreement when deciding who can change rules. *Journal of Experimental Child Psychology, 165*, 101–116.
- Zillmer, N., & Kuhn, D. (2018). Do similar-ability peers regulate one another in a collaborative discourse activity? *Cognitive Development, 45*, 68–76.
- Zlatev, J. (2014). The co-evolution of human intersubjectivity, morality, and language. In D. Dor, C. Knight, & J. Lewis (Eds.), *The social origins of language* (pp. 249–266). Oxford, UK: Oxford University Press.

Biography

Leon Li graduated with a B.S. in Psychology from the University of Maryland, College Park, in 2015. He obtained a M.A. from Duke University in 2019 as part of his doctoral studies in Psychology. Leon's publications include: (1) Of papers and pens: Polysemes and homophones in lexical (mis)selection, (2) Theory of mind and resource allocation in the context of hidden inequality, (3) Brain-to-speech decoding will require linguistic and pragmatic data, (4) Lying, negligence, or lack of knowledge? Children's intention-based moral reasoning about resource claims, (5) Envisioning intention-oriented brain-to-speech decoding, (6) Why does awe have prosocial effects? New perspectives on awe and the small self, (7) On the moral functions of language, and (8) Young children conform more to norms than to preferences. At Duke, Leon was part of the Society for Duke Fellows, the James B. Duke Fellowship, the University Scholars Program, the Vertical Integration Program, the Kenan Graduate Fellowship, and the Summer Research Fellowship. Leon also received support for travel and research from the Charles LaFitte Foundation.