

This Target Article has been accepted for publication and has not yet been copyedited and proofread. The article may be cited using its doi (About doi), but it must be made clear that it is not the final version.

THE MORAL PSYCHOLOGY OF OBLIGATION

Michael Tomasello

Max Planck Institute for Evolutionary Anthropology &

Duke University
Department of Psychology and Neuroscience
Duke University
Durham, NC 27708-0086

michael.tomasello@duke.edu

Long Abstract	=	242
Text	=	13,244
References	=	2,048
Total	=	15,534

Long Abstract

Although psychologists have paid scant attention to the sense of obligation as a distinctly human motivation, moral philosophers have identified two of its key features: first, it has a peremptory, demanding force, with a kind of coercive quality, and second, it is often tied to agreement-like social interactions (e.g., promises) in which breaches prompt normative protest, on the one side, and apologies, excuses, justifications, and guilt, on the other. Drawing on empirical research in comparative and developmental psychology, we provide here a psychological foundation for these unique features by showing that the human sense of obligation is intimately connected developmentally with the formation of a shared agent "we", which not only directs collaborative efforts but also self-regulates them. Thus, children's sense of obligation is first evident inside, but not outside, of collaborative activities structured by joint agency with a partner, and it is later evident in attitudes toward in-group, but not out-group, members connected by collective agency. When you and I voluntarily place our fate in one another's hands in interdependent collaboration - scaled up to our lives together in an interdependent cultural group - this transforms the instrumental pressure that individuals feel when pursuing individual goals into the pressure that "we" put on me (who needs to preserve my cooperative identity in this "we") to live up to our shared expectations: a we > me self-regulation. The human sense of obligation may thus be seen as a kind of self-conscious motivation.

Short Abstract

The human sense of obligation has two key features: (i) a peremptory, demanding force, with a coercive quality, and (ii) ties to agreement-like social interactions in which breaches prompt normative protest, on the one side, and apologies, excuses, justifications, and guilt on the other. Drawing on empirical research in comparative and developmental psychology, we provide here a psychological foundation for these unique features by showing that the human sense of obligation is intimately connected developmentally with the formation of a shared agent "we", which not only directs collaborative efforts but also self-regulates them.

KEY WORDS: Cooperation, Fairness, Morality, Motivation, Normativity, Obligation, Shared Intentionality

Humans often do things out of a sense of obligation. But obligation is not a major topic in modern psychology, not even in moral psychology. Presumably obligation is a kind of motivation, but, if so, it is a decidedly peculiar one. It has at least two distinctive features.

- **Special Force.** Obligation has a peremptory, demanding force, with a kind of coercive (negative) quality: I don't want to, but I have to. Failure to live up to an obligation leads to a sense of guilt (also demanding and coercive). Unlike the most basic human motivations, which are carrots, obligation is a stick.
- **Special Social Structure.** Obligation is prototypically bound up with agreements or promises between individuals, and so has an inherently social structure. It can even happen that an outside party judges that an agent is obligated to do something although the agent himself does not think so. Breaches of obligations often prompt normative protest, from the offended party, and apologies, excuses, and justifications, from the offender.

Our aim in this paper is to explain humans' sense of obligation, including these two special features, in moral psychological terms. Where did humans' presumably species-unique sense of obligation come from evolutionarily, and where does it come from ontogenetically? What are its functions, and from what more primitive components was/is it built? Our larger goal in answering these questions is to fit obligation into a larger picture of human sociality.

1. Philosophical Background

Despite its almost total neglect in psychology, obligation has been a major topic in moral philosophy for centuries, albeit most often in the context of normative discussions attempting to tell us what we are obligated to do. But there have been some influential descriptive approaches as well (in so-called metaethics) that attempt to discern what human obligation is in the first place. We consider in this section two such approaches, and then, in the sections that follow, attempt to improve on them through modifications based on empirical results from the fields of developmental and comparative psychology.

The first philosophers to consider obligation from a more or less psychological point of view were the Enlightenment moral philosophers, especially David Hume (1751/1957). Although contemporary research in moral psychology has focused mainly on what Hume dubbed the natural virtues, such as sympathy, just as important in his overall account are what he dubbed the artificial virtues, such as justice. The natural

virtues are so-called because they have a more or less direct grounding in natural (presumably evolved) human motives or passions, especially sympathy. But the artificial virtues such as justice (and also what we would today call fairness) have no such direct grounding in natural motives; they govern human behavior only through a deliberative process in which many factors are considered and weighed.

Hume argued that the artificial virtues such as justice (and fairness) are grounded in human convention. Societies have formulated rules for how individuals should behave when their various interests are potentially in conflict. In many such cases, individuals are willing to curtail their own self-interest and cooperate for some overall good, but only under the condition that others cooperate as well (what today we call contingent cooperation). Hume's main example is private property: each individual is willing to forgo any motivations to appropriate the property of others, given that others are willing to forgo their motivations to appropriate his. This cooperative attitude is grounded at least indirectly in sympathy, as individuals feel that for the good of society as a whole - a kind of generalized sympathy for the social group - individuals are inclined to follow the rule. Hume's other main example is promising. While there are many occasions on which breaking one's promise would be advantageous to the self or others, and cause no harm to anyone, sympathy and concern for the smooth functioning of society as a whole recommend always keeping one's promises. But in neither property nor promises is sympathy for the group sufficient. People who respect property and keep promises (and expect others to do so) also give a judgment of *approval* to the rule - reflectively, from "the general point of view" - and so they do not just follow the rule but feel obliged to do so.

Hume's account does not have much to say about obligation's special motivational force. Strawson (1962) argued that approaches in which the individual simply sits back and observes others, either approvingly or disapprovingly, are too passive to account for humans' normative relations to one another. He pointed out that humans have and express participant reactive attitudes toward others, such attitudes as resentment and blame, when others treat them badly, or gratitude and forgiveness, when others treat them well. Scanlon (1998) argues that this more active way of relating to others falls within a delimited domain of human morality having to do with "what we owe to each other". This appellation seems appropriate in the current context because it matches our intuition that the sense of obligation is a feeling of owing someone something (we even describe an obligation as being in someone's debt). A growing number of philosophers have focused on this delimited domain in recent years, most prominently Darwall (2006; 2012) (see also Wallace, 1994; in press).

In focusing specifically on obligation, Darwall (2006; 2012) points out the existence of what have been called bipolar moral attitudes (see also Thompson, 2004). If doing X would break my promise to you, for example, then I have an obligation not to do it, and, correspondingly, you have a right to expect me not to do it. The *obligation* from me to you and the *right* from you to me together define one another (see Hohfeld, 1923). And it is an empirical fact that in human moral communities individuals actively

assert their rights; they actively hold one another accountable for keeping their obligations. They do this by expressing their reactive attitudes, such as resentment and blame, to those who have reneged in acts of so-called moral protest, to which they expect/demand an appropriate response in return (Smith, 2013). Then, if the offender wishes to keep her good standing as a moral being (in both her own eyes and those of others), she must in some way recognize the legitimacy of the protest, for example, by apologizing or making an excuse (demonstrating shared values despite the breach).

Darwall's account is especially useful from a psychological point of view because he makes some specific claims about the way that individuals enter into these kinds of bipolar relations. For individuals in a community to hold one another morally accountable in the human-like way, they must possess what he calls second-personal competence, and they must recognize one another as having such competence in more or less equal measure. Thus, each must recognize and respect others in their community "as someone 'just like me'" (2013, p.29). Each individual sees herself as "one mutually accountable agent among others" (2013, p.7). Second-personal agents display "mutual respect between mutually accountable persons" (2006, p. 36). As such, a second-personal agent can expect another second-personal agent to hold himself accountable, such that he will respond appropriately to blaming (e.g., by giving a good excuse or by apologizing). And of course a second-personal agent at the same time gives other persons the authority to call him out when he transgresses. To participate in such a system of mutual accountability, Darwall (2013, p. 47) claims that the individual must be a rational agent with the capacity for reactive attitudes (such as blame) and also basic empathy (to put oneself in others' shoes), as well as "the capacities to make and regulate oneself by normative judgments about what attitudes are warranted" [e.g., by accepting blame or feeling guilty]. Household pets and young human infants do not have such competencies and therefore do not participate in a system in which individuals hold one another mutual accountable.

So long as a being has the capacity to take a second-person standpoint toward others and himself, make judgments about what demands would be warranted from this perspective, and (self-)regulate his conduct through making the relevant demands of himself [e.g., through feelings of guilt; MT], the being counts as second-personally competent. (2013, p. 47)

He immediately then claims (following Strawson) that "second-personal competence is both necessary and sufficient for moral obligation".

Darwall's account thus explains obligation's special coercive quality as deriving directly from the subject's recognition of a legitimate protest or claim (or an imagined protest or claim) from a social-interactive (second-personal) partner. And the connection to guilt follows directly from this: individuals internalize a second-personal interactant's blame or protest (or anticipate such blame or protest), and, to the degree that they find it warranted, apply it to themselves. Individuals use this internalized process to self-regulate their social behavior. The sense of obligation to an interactive partner is thus not simple approval of the governing societal rule (à la Hume), but rather it is the force

and legitimacy of the claim that a second-personal agent makes, or could make, on me.

With respect to the special social structuring of the sense of obligation, Darwall recognizes societal conventions as prototypical situations of obligation, but only if second-personal agents are involved – and such conventions are not strictly necessary. Thus, he posits both “directed obligations”, which concern what one person owes to another as a result of their second-personal statuses (e.g., even in the absence of conventions, they owe one another respect and fair treatment), and “obligations period”, which also rely on general normative principles of second-personal agency but are not directed at particular others.

Despite the advance that such social-interactive accounts of obligation represent – as compared with the older “spectator” accounts based on disinterested judgments of approval and disapproval – they leave open many important questions from a psychological point of view. Most important, they do not address such basic social-psychological questions as:

- (i) Where do the reactive attitudes (such as resentment and blame) come from, and why do human individuals (but not other animals) care that others are blaming them or making claims on them in the first place?
- (ii) Where do second-personal agents come from, such that other animals and young infants, for example, are not second-personally competent?
- (iii) When individuals accept a claim from another as legitimate or warranted – by, for example, apologizing, justifying, making excuses, or feeling guilty – in what does their understanding of such legitimacy consist?

To answer these questions is to give a more secure psychological foundation to Darwall’s and similar philosophical accounts of the human sense of obligation.

2. A Shared Intentionality Account

Our central claim here is that the human sense of obligation is part and parcel of humans’ ultra-cooperative nature. Evolutionarily, it emanated from the process by which collaborative partners assessed one another – and worried about being assessed – for their cooperativeness. This occurred in a socio-ecological context in which exclusion from collaboration meant death. Ontogenetically, young children first feel a sense of obligation to collaborative partners: they act more prosocially toward collaborative partners than toward others, they show resentment and protest when their collaborative partners treat them badly, and they formulate excuses and justifications when they treat their collaborative partners in objectionable ways. They are acting in these special ways toward collaborative partners – and not toward others – based not on societal rules or generalized principles of second-personal agency, but rather on the normative bonds that interdependent collaboration creates (at the same time and through the same processes that it creates second-personal agents; see below). We will argue, at the end of the paper, that classic tit-for-tat reciprocity does not create normative bonds of this kind

and so cannot account for humans' sense of obligation.

What is crucial for a sense of obligation psychologically, in this view, is a sense of shared agency, a sense that "we" are acting together interdependently. We have put our fates in one another's hands and so hold one another responsible for appropriate respect and treatment. The ability to form shared agencies derives from a uniquely human social psychology of shared intentionality, as described by philosophers of action such as Bratman (1992, 2014), Gilbert (1990, 2014), and Searle (1995, 2010), and as applied empirically by psychologists such as Tomasello (2014, 2016). In both phylogeny and ontogeny this unique social psychology unfolds in two key steps. The first step concerns how individuals relate to one another in the context of collaboration: joint intentionality, which creates a joint agent "we" and a dyadic, second-personal morality between collaborative partners. The second step concerns how individuals relate to their cultural group more generally: collective intentionality, which creates a collective agent "we" and a norm-based, "objective" morality in the cultural group. These two steps correspond, in a general way, to Darwall's distinction between directed obligation (to an individual) and obligation period (to no individual in particular).

2.1. Joint Intentionality and Second-Personal Responsibility

In some groups of chimpanzees, individuals hunt monkeys together in groups. But they do not call their partners out for poor performance, or apologize or make excuses or feel guilty for their own poor performance, or feel obligated to share the spoils in a fair manner among participants. They are basically using one another as social tools for their own individual ends (Tomasello et al., 2005), and therefore do not seem to be operating with a sense that they owe things to one another. But at some point in human evolution, a new form of social engagement emerged: collaborative interactions initiated by the joint agent "we", which self-regulated the individual partners "I" and "you" (perspectively defined). We may call this new form of collaborative interaction the joint intentional schema, or dual-level collaboration (Tomasello, 2014, 2016, 2019). The essential point in the current context is that the joint intentional schema created the possibility of a morality of fairness among coequal second-personal agents, who felt a sense of normative obligation to treat one another in mutually respectful ways.

2.1.1. *Evolutionary Background*

The specifics of how humans evolved their new forms of collaborative interaction are not important for current purposes. What is important is that at some point in human evolution the feeding ecology changed, and individuals were forced to collaborate with others to obtain food or else perish. The skills and motivations of good collaborators were thus naturally selected. In this context, a key part of the process was partner choice. Individuals were looking for the best partners, and of course they had to

be good partners themselves if they were to be chosen (Baumard et al., 2013). Partner choice is thus fundamentally a process of social selection in which individuals are both judge and judged, and the only ones who survive are those who get good partners because those good partners evaluate them positively. It is thus important that individuals have a kind of cooperative identity in the eyes of collaborative partners and themselves (“we”).

In addition to partner choice - in which individuals attempt to identify and choose the best partners - early humans also engaged in partner control in which individuals attempt to transform existing partners into better partners. In the evolutionary literature, the most common form of partner control is punishment for noncooperation (e.g., in meerkats' so-called ‘pay to play’), which hopefully makes individuals cooperate in better ways. But early humans developed a unique form of partner control that plays a key role in the way they relate to one another: they protested against non-cooperative behavior, giving the non-cooperator a chance to mend her ways voluntarily of her own accord. If she did not, her cooperative identity with partners would be damaged, and she could be hit with the ultimate sanction of being partner choiced out of existence.

The distinction between partner choice and partner control is important in the current context because we might say that whereas Hume focuses mainly on processes of partner choice, in which individuals simply judge who does or does not have the virtues of a good cooperative partner, Darwall focuses mainly on processes of partner control, in which individuals demand that their partner shape up (or else they will ship out). Partner control within the context of human collaborative interaction - which we will characterize below as a kind of “we” self-regulating each partner “I” and “you” - may thus be seen as the evolutionary basis for many of the most important participant reactive attitudes (and how recipients respond to them). Partner choice is more of a unilateral decision, whereas partner control - and so the sense of obligation - is bilateral negotiation. Thus, for instance, ownership is not just about how I relate to my iPhone, but rather about how you and I relate to one another with respect to my iPhone. And promising is not just about my behavior, but rather about how you and I relate to one another with respect to my behavior. This triadic structure - you and I relating to one another *about* some external object or action - is the defining organization of social activities structured by shared intentionality.

2.1.2. Ontogeny

Humans' adaptations for interacting with others in these new ways - culminating in a sense of fairness and obligation to treat others cooperatively (and to expect such treatment from others) - comes into being gradually during ontogeny via a kind of maturationally guided learning (Tomasello, 2019). It begins to emerge initially in young children from 1 to 3 years of age. Of crucial importance for the current argument, this initial emergence takes place mostly, or only, within joint intentional collaboration, and

can be seen in the species-unique ways in which human children, in contrast to other apes, relate to their collaborative partners. What we are looking for is the mutually respectful behavior toward partners characteristic of second-personal agents (à la Darwall), such things as persisting in commitments even when one does not want to, dividing the spoils “fairly” even when one does not want to, asking permission to break a commitment, respectfully protesting a partner’s uncooperative behavior, and justifying or making excuses for one’s own non-cooperative behavior.

Joint Goals and Commitments. In terms of motivation, the most basic comparative fact is that, in situations of free choice with rewards for both partners identical, 3-year-old children mostly choose to collaborate with a partner, whereas chimpanzees mostly choose to go it alone (Rekkers et al., 2011). Children are so motivated to collaborate that they actively attempt to reengage a recalcitrant partner, whereas chimpanzees ignore a recalcitrant partner and, again, attempt to go it alone (Warneken et al., 2006). Indeed, children are so motivated that they attempt to reengage a recalcitrant partner even when they know they could act alone and reach the same result (Warneken et al., 2012). Humans have a species-unique motivation and preference, at least among great apes, for pursuing goals by collaborating with others. But more than this preference, children also relate to one another inside these collaborative activities in unique ways suggestive of a sense of obligation to treat their partner respectfully (as an equal).

Within joint intentional activities, children go to some pains to make sure that their partner gets her reward. Thus, when two 3-year-olds are working together toward a joint goal, and, seemingly by accident, one child has access to her reward first, they nevertheless persist with the collaboration until the partner obtains her reward as well. They do not do this nearly as often in a control condition in which another child lacks her reward in the same way but they are not collaborating (Hamann et al., 2012). In contrast, chimpanzees in the same two situations persist with their partner at the same (low) rate, whether they are collaborating with her or not (Greenberg et al., 2010). Children's deferential behavior to partners within the collaborative condition of this study could conceivably be interpreted simply as an enhanced motivation/preference to help their partner. But it could also be interpreted as collaborative activities generating a normative sense of responsibility to the partner and her welfare.

Consistent with this latter interpretation, children at around three years of age begin to appreciate the normative force of joint commitments to collaborate.¹ Thus, when adults propose a joint commitment to two children (“Why don’t you guys X together, OK?”) and the children agree explicitly (“OK”), this has huge consequences in how they treat their partner. For example, after an adult has orchestrated a joint commitment between 3-year-old peers, if one of them is tempted away by a more

¹ Evolutionarily, joint commitments to collaborate are a key way of mitigating risk, as partners make sure that they both know in common ground that they are depending on one another, and so they assure one another that they can be depended upon to behave in expected ways.

rewarding "bribe" that benefits only her, they quite often resist this bribe; that is, they resist this bribe more often than if they are simply playing side-by-side without the joint commitment (Kachel & Tomasello, 2019). In another experimental paradigm, when 3-year-old children form a joint commitment with a partner, again as opposed to merely playing beside her, they more often do such things as wait for their partner when she is delayed, repair some damage done by their partner, refrain from tattling on their partner, and perform their partner's role for her when she is unable (Graefenhain et al., 2013). Although again possible, it is more difficult to interpret all of these behaviors as resulting simply from an enhanced prosocial motivation toward the other child; why should simply agreeing to do something have these various effects if not because it creates some sense of responsibility to one's partner?

In further support of the view that children understand joint commitments to have normative force (and not to just enhance motivation), when a child's partner fails to play his role in a joint commitment in the way they both know he should, the child protests (but not if the partner fails out of ignorance; Kachel et al., 2018). It seems implausible that children are protesting that the other child has not enhanced her prosocial motivation. Much more likely is that they are protesting the breaking of a joint commitment. And crucially, the child's protest in such situations is not articulated as a personal preference or desire (e.g., "I do not like it when you X"), but rather as a normative requirement of anyone who would play that role, with the child saying such normative things as: "One must pull on the rope", "That's not how it is done", etc. Thus, beyond the likes and dislikes of the individuals involved, there are for the children normative standards of how collaborative partners play their roles, which apply to anyone in the appropriate role. Connecting back to the previously cited studies, if young children see their collaborative partners as normatively committed to them in this way in such situations, then it is likely that their resisting bribes to defect on their partner, and similar behaviors, are underlain by a similar sense of them being normatively committed to their partner as well (see also Sipisova et al., 2018).

Finally, perhaps the strongest evidence that young children understand joint commitments as normatively binding comes from their understanding of how to break one. Thus, when an adult proposes a joint commitment to a 3-year-old to play a game together (which the child accepts), and the child is lured away to a more fun game, most of them acknowledge in some way their breaking of the commitment (in a way that 2-year-olds do not). They "take leave", sometimes by explicitly asking permission (e.g., "I'll go over there now, OK?"), sometimes by apologizing (e.g., "Sorry, but I'm going over there now"), sometimes by simply announcing their departure (e.g., "I'm going over there now"), and sometimes nonverbally by hesitating and looking anxiously to the partner before leaving. They do not do any of these things if they are just playing with the adult spontaneously, in the absence of a joint commitment, and then are lured away (Graefenhain et al., 2009). And in another experimental paradigm, when the child's partner to a joint commitment breaks it by leaving, it matters how he takes leave. If he asks permission or gives notice, the child allows him to leave without protest and waits

longer for him to return, as compared to the situation in which he leaves peremptorily without asking permission or giving notice (Kachel et al., 2019). These studies thus support the conclusion that 3-year-old children understand the normative force of joint commitments with respect to both their own and their partner's actions, based on behaviors such as leave-taking and normative protest that do not emanate simply from an enhanced preference.

The natural conclusion from all of these studies is that young children feel some kind of normative connection to their collaborative partner that they do not feel toward others outside of collaboration. It is thus collaboration among individuals, not societal convention, that is the first and most natural home of the sense of obligation (or responsibility; see below) directed to one's partner.

Sharing the Spoils "Fairly". If an ape has food resources in its possession, it very seldom gives up any of them to anyone else - and certainly not for no reason. Young children are a bit more generous, but not much; on average, in dictator games where they are free to share what they will, 3-year-olds across cultures offer peers about one in four items in their possession (Ibbotson, 2014).

But when the resources to be divided are the fruits of a collaborative effort, we see a very different pattern. When chimpanzees pull in a board together with food clumped in the middle, typically the dominant individual simply takes it all, and collaboration breaks down over trials (Melis et al., 2006). In contrast, human 3-year-olds in the same situation divide the spoils more or less equally on more or less every trial, and they can continue to collaborate in this manner indefinitely (Warneken et al., 2011). Most dramatically, when 3-year-old peers collaborate to pull in resources and, by "luck", one of them ends up with more than the other, the unlucky child often verbally notes the inequity (e.g., "I only have one"), and the lucky child often (about three-quarters of the time) hands over the extras so as to equalize the rewards among partners (Hamann et al., 2011). They almost never do this in a control condition with no collaboration, suggesting that the sense of shared agency in producing the rewards is crucial. In contrast, chimpanzees, in a study designed to be as similar as possible to this one, shared rewards (i.e., allowed the partner to take them) equally often inside and outside the context of a collaboration, presumably because they have no sense of shared agency in producing the spoils. In a related set of studies, children who received all of the rewards from pulling in a board with sweets on it, shared those sweets more often with a collaborative partner than with a peer who was simply nearby (i.e., was a free rider to the spoils; Melis et al., 2015). Chimpanzees in the same situation shared equally infrequently with partners and free riders alike (Melis et al., 2011).

Thus, in the context of joint intentional collaboration, but not in non-collaborative contexts, children, but not chimpanzees, are motivated to share equally with, and only with, their partner. This is not just enhanced generosity to one's collaborative partner: in the Hamann et al. (2011) study, children did not give the collaborative partner more than half but only the amount needed to equalize the rewards. Indeed, when motives for generosity and equality are explicitly pitted against one another in this experimental

paradigm, children who have been in a collaboration do not accept distributions that are generous either to their partner or themselves, but only accept distributions readily when the rewards are distributed equally (Corbit et al., 2017) – and they do not behave in this way outside of collaboration. Joint intentional collaboration does not just generate an enhanced generosity toward one's partner, but rather an enhanced normative sense that “we” should share “fairly” (i.e., equally).

Of crucial importance to this normative interpretation is the phenomenon of social comparison in distributive situations. The main point is that the sense of distributive fairness toward a collaborative partner implies a social comparison of the spoils obtained by self and partner and, critically, a judgment that we both *deserve* the same. This suggests that the sense of fairness is less about the absolute amount of resources distributed than about how one treats one's partner relative to the self (Englemann & Tomasello, in press). Thus, when children receive, for example, one piece of candy, they are content, but when a collaborative partner at the same time receives 5 pieces of candy, they are not content and often register normative protest (e.g., "It's not fair." See Rakoczy et al., 2016). Importantly, children feel this aversion to inequity also in the opposite direction: they are happy to receive 5 candies on their own, but they are unhappy if their partner at the same time receives only one, and indeed in this case they often share with the partner in order to equalize (in the vernacular, they show an aversion even to advantageous inequity). In contrast, in virtually identical situations, chimpanzees react to the absolute amount they receive irrespective of how much the partner receives; they do not engage in social comparison at all (see Ulber & Tomasello, 2017, for a comparative study of children and chimpanzees).²

Further supporting this view of distributive fairness as an issue of equal respect for one's partner is the phenomenon of procedural fairness (Shaw & Olson, 2014). In a small group of children, 5-year-olds do not like to receive a smaller share of the spoils than their partners. But if the children agree ahead of time on a fair procedure for distributing resources (rolling a die, drawing straws, etc.), then they are all content with the outcome even if they end up with less than the others (Grocke et al., 2015). The point is not to get the same amount, but to be treated fairly, as an equal. Children are also content to receive less than others if they have been given a voice in making the distributive decision (Grocke et al., 2018), again suggesting that the main issue is being treated fairly and with respect. And in a mini-ultimatum game, 5-year-old children will even sacrifice resources to punish a partner who does not share with them equally (Wittig et al., 2013) - presumably out of resentment for being treated as less than equal (i.e., unfairly) - which chimpanzees do not do (Jensen et al., 2007).

The natural conclusion from these sharing studies is that young children feel a kind of normative force to share fairly with collaborative partners that they do not feel

² Brosnan and de Waal (2003) claimed that capuchin monkeys have a sense of fairness. There are six published failures to replicate their results (with both monkeys and chimpanzees) using appropriate controls from five other laboratories (see Tomasello, 2016, pp. 32-34, for a review).

toward non-partners (whereas chimpanzees do not discriminate). They feel they should treat their partner, and be treated by their partner, as equally deserving, which is also the case in situations of procedural fairness even if the resources end up unequal. Again, then, it is joint intentional collaboration that is the first and most natural home of a sense of obligation (or responsibility; see below) directed to one's partner: in this case, to be respectful and fair to one's partner in dividing collaboratively produced resources.

2.1.3. *An Interim Explanation*

It is unlikely that the explanation for these experimental results is that collaboration induces an enhanced sympathy for the partner. Enhanced sympathy could potentially explain why children persist in collaboration in the face of outside temptations only after they have made a joint commitment but not otherwise, and why they share more with their partner when they are dividing the spoils of collaboration than otherwise. But enhanced sympathy alone could not, under any plausible scenario, generate children's tendency to share *equally* with their partner only inside, and not outside, of collaborative activities, much less their satisfaction with procedural fairness. It could also not explain children's normative protest against unfair treatment – either when the partner defects from the collaboration or shares the spoils unfairly – using normative language such as *must*, *should*, and *ought*, nor could it explain children's various leave-taking behaviors (including apologies and excuses) when they are breaking a commitment. The most plausible explanation is that children are feeling some sort of obligation or responsibility to treat their partner respectfully and fairly.

We will give our overall explanation for the developmental emergence of children's sense of obligation only after we have also examined older children's sense of collective agency and "objective" morality. For now, we simply wish to highlight the indispensable role of the joint agent "we", which not only conducts but also self-regulates the collaboration. Even 18-month-old toddlers seem to have formed a joint agent "we" with their partner, as evidenced by the fact that they attempt to reengage recalcitrant partners even when they do not need them for goal success, using cooperative means such as beckoning and encouragement (which apes do not do; Warneken et al., 2006; 2012). These communicative behaviors may be seen as attempts by the child to reconstitute our lost "we". And by the time they are three years of age, children are able to actually constitute a normatively structured joint agent "we" by forming with a partner a joint commitment to jointly self-regulate the collaborative process, in the sense that it gives each party to the agreement the standing to protest or rebuke non-cooperative behavior. Darwall (2006) characterizes the basis for such protest as the parties to the agreement giving one another the "representative authority" of the cooperating body ("we") to call the other out for non-cooperative behavior. And when children protest a partner's non-cooperative behavior, the offending partner typically recognizes and accedes to this protest because she views it as legitimate or warranted, based on their status as equally deserving second-personal agents who have thrown in together to form a self-regulating "we".

Taking the protest of collaborative partners seriously, as a legitimate grievance from "we", amounts to self-regulative pressure from "we". The fact that 3-year-old children do in fact feel this way when they offend their partner, at least sometimes, is apparent in their feelings of guilt when they let their partner down. In experimental studies, after children have semi-inadvertently ruined their play partner's creation, they go to much trouble to make reparations (Vaish et al., 2016). Guilt thus represents a kind of after-the-fact collaborative self-regulation, as it also constitutes pressure from "we" (our common ground standards for role performance) on both "I" and "you": what Tomasello (2016) calls a *we > me* moral attitude. More generally, what is being self-regulated either during or after the collaboration is my cooperative identity: how "we" evaluate "me" as a partner.

Key in all of this is the notion of *role* that joint intentional collaboration creates. In collaborative activities performed by a "we", partners each have their role to play. As both play their roles, they come to see one another as equally deserving individuals, equally worthy of respect. This recognition is based on a dawning understanding of self-other equivalence (Nagel, 1970): I and my partner are equivalent in all important respects in this collaborative context. Most basically, as children participate in joint intentional collaboration, they see that: (i) both participants are equal causal forces in producing the mutually intended outcome; (ii) both partners could switch roles as needed; and, most crucially, (iii) the standards of performance for each role (so-called "role ideals", the first social shared normative standards) are impartial in the sense that they apply to anyone in that role. By the time they are 3 years of age, children thus come to understand a kind of self-other equivalence in the context of joint intentional collaboration to the extent that they now view their partner as a mutually deserving second-personal agent to whom they owe respect and fairness.³

And so, the interim hypothesis is that participation in joint intentional collaboration, especially as initiated by a joint commitment, is the earliest source of children's (and early humans') feeling of obligation to their partner. Critical to the process is the child's judgment that she deserves to be criticized if she does not live up to the joint commitment. This judgment of legitimacy is based on (i) an understanding of the partner as an equally deserving second-personal agent who deserves to be treated cooperatively (based on a sense of self-other equivalence); and (ii) an understanding of a kind of *we > me* self-regulation of the collaborative activity in which the "we" is the joint agency to which the child has jointly committed and which she must, to maintain her cooperative identity in the partnership, respect. Nevertheless, because this early normative sense is delimited - directed to and only to a collaborative partner - let us call it a sense of second-personal *responsibility* directed to the partner (as alluded to above). To get to Darwall's sense of obligation period - not directed to any particular individual - we will need to get to a fully-fledged "objective" morality, and for that we need a more

³ Of course children this young do not view the adults with whom they are interacting as co-equals in general. But when they are rolling a ball back and forth or building a tower together they are, in the context of that play activity, co-equal play partners.

expansive form of cooperative social engagement.

2.2. Collective Intentionality and "Objective" Obligation

Second-personal responsibility to a collaborative partner is a real, but nevertheless circumscribed, form of normative obligation. To get to a more universal form, we need a larger social context with universalizing elements. Specifically, we need a human culture with which individuals identify: "we" Erewhonians (or whatever). Ontogenetically, it is only children after 3 years of age who identify with their culture in this manner and so not only favor in-group members over others, but also subordinate their own preferences to the cultural group's collective expectations for individual behavior, a.k.a. social norms, some of which evoke a sense of obligation to "objective" moral standards (violation of which can result in one losing one's moral identity within the group).

The most fundamental assumption guiding this second step of our account is that a cultural group - that is, evolutionarily, a hunter-gatherer group with clear demographic boundaries as characteristic of humans for most of their history - is nothing more or less than one big collaborative activity in which "we" as a people operate with a collective commitment to the group's surviving and thriving. Each individual has her role(s) to play in this collective commitment - both as a member of the group in general and, possibly, as a person playing some more specific division of labor role - and this generates, in a scaled-up manner, more universal normative expectations. Extending the proposal that commitment and fairness accompanied by a sense of obligation first apply to, and only to, collaborative partners, the proposal now is that the sense of obligation at this second step applies to, and only to, one's cultural compatriots. One may have sympathy for suffering out-group members, but it is not clear that one has obligations to be fair to them (unless one has an expanded sense of all of humanity as in one's in-group moral community).

2.2.1. *Evolutionary Background*

The emergence of modern humans, some 150,000 years ago, is marked by growth in the size of social groups, leading eventually to tribal organization - involving multiple semi-independent "bands", united into a larger tribe - and to competition with other tribal groups. What this meant was that one's group now contained a new category of individuals - in-group strangers - who had to be distinguished from out-group competitors. For individuals, this meant that it was important to identify who was and was not in the tribal group, and it was important that they themselves be identified by others as a member of the tribal group (a.k.a. culture) as well. To be identified as a group member what was most important was conformity, since the most reliable way to identify members of the cultural group was by commonalities of behavior - speech and other conventional cultural practices - and, at some point, by appearance in terms of

dress, cultural markers, etc. Nonconformists were suspect and at risk of being excluded from the group – a scaled up process of partner choice.

But there were also issues of partner control. One needed to be able to coordinate with all and only in-group members, even if one did not know them personally. For example, anyone who grew up in this group must know how to net-fish or worship with others in the conventional way. There was thus pressure to conform to the group's conventional cultural practices as the way we Erewhonians do things (where doing otherwise risks disrupting things for one's compatriots). And so arose social norms: because we all value the group's smooth functioning, we all must do things in the ways that we all expect us to do them. In addition, to be a good group member we must also make sure that others follow these norms as well (especially by normatively protesting violations). The enforcement of social norms is thus a kind of scaled-up, third-party version of the second-personal protest characteristic of joint intentional collaboration: it is a new, group-level form of partner control comprising group-level protest backed by a threat of exclusion (partner choice). When individuals deviate, other group members call them out for nonconformity, with the enforcer acting as a kind of representative of the larger cultural "we". Third-party enforcement on nonconformists, with the implicit backing of the group, makes everything much less personal: anyone who did what you did would be called out for it. It thus represents a first step toward the objectification of norms.

Cultural practices and social norms in large part identify our group as who we are: "we" are those people who talk, think, dress, and eat in these particular ways. Being a member of the group means identifying with these ways (begun by our revered ancestors), such that the group comprises not a finite number of individuals but a universalizing description of identity: "anyone who would be one of us". Even though the group's social norms existed before I was born, I feel myself to be, in an important sense, a co-author: "we" Erewhonians created these norms for the good of the group and everyone who would be one of us agrees. This creates the most basic – and sometimes pernicious – distinction in humans' group-minded existence: the distinction between those of us who, by following and enforcing the practices and beliefs of our culture, are rational/moral beings, and those from alien groups (barbarians) who are not rational/moral beings at all. This universalization of identity legitimates our ways of thinking and acting as objectively rational and moral (especially since those who would not be one of us seem to be incapable of behaving rationally and morally). "We" are therefore justified or warranted in coming down on "me" if I transgress, since all group members, including me, *should* be subject to these norms because they are the legitimate, indeed objectively true and valid, ways that rational/moral beings act.

In this context, a kind of new reality emerges: institutional reality (Searle. 1995). Some cultural practices become institutionalized, as the common ground assumptions and interdependencies involved are made explicit and public. Thus, mating behavior becomes marriage, leaders become chiefs, and items used to establish equivalences in trade become money. This process brings into existence new types of agents with new

types of deontic relationships to others (both rights and responsibilities) that are conferred upon them, as it were, by a declaration of the group. Chiefs have conferred upon them a new status that entitles them to perform marriages and declare war, but at the same time obliges them to consult with the elders before acting. Membership in the cultural group itself becomes a status: those who identify with the group - by affirming and conforming to its ways - are recognized as group members, often after passing some "rites of passage" at around adolescence. This cultural identity is something valuable to individuals, who seek to maintain it (e.g., via acts of impression management).

In this cultural-institutional context, then, individuals continue to self-regulate in a we > me manner, but now the "we" is our culture - i.e., those who would be one of us - and each of us must conform to the group's "objective" normative standards specifying the right and wrong ways to think and act. In this cultural-institutional context, individuals with a moral identity in the group feel an obligation to either conform to the group's ways or else to justify themselves to others by explaining their deviance as resulting from values that we all still share (e.g., I neglected my cultural duties because I had to go save my drowning child). Individuals thus feel that to maintain their cultural identity in the group, they are obliged to do (or justify not doing) the "objectively" right things, which are experienced (in Darwall's terms) as obligations period.

2.2.2. *Ontogeny*

Humans are thus adapted not just for cooperation with individuals, but also for life in a cultural group. These adaptations for group life begin to emerge ontogenetically at around the third birthday, as children begin to display an emerging group-mindedness and sense of "objectivity" in many domains of life (Tomasello, 2018, 2019). Of special importance in the current context, it is at this age that children's cooperative interactions with others begin to take a "normative turn": they begin to discern how "we" in this group *should* act, that is, how a collective commitment to the group - created through identification with the group and its social norms - *obligates* us to act. As development proceeds through the preschool years, children's emerging group-mindedness enables them to engage in various new forms of maturationally guided learning, in which the group and its normative concerns play an ever-increasing role.

Group-mindedness. Human children socially interact with others from early in infancy, sometimes even in the midst of multiple other people. But from 3 to 5 years of age, children begin to understand social groups as such. Astoundingly, during this age range children's social behavior is even affected in significant ways by simply being assigned to a "minimal group" established only verbally and with arbitrary supports, for example, as they don a green T-shirt and are told that they belong to the green group (along with other similarly clad children; Duhham, 2018). The result is that children's sense of being in a group is not based simply on physical proximity and/or familiarity (as it is, arguably, in other primates), but rather on the idea of a social group based on similarity alone.

Thus, 4- to 5-year-old children who have been assigned to a minimal group show loyalty by preferring to stick with the group even when it means losing rather than winning a game (Misch et al., 2014). Preschool children preferentially help in-group over out-group members (Over, 2018). When a member of a 5-year-old's minimally established in-group does something mean to a victim, she feels an in-group responsibility to make amends to that victim – whereas she does not feel a responsibility if the perpetrator was an out-group member (Over et al., 2016). When children in this age range are given a chance to share with children in their in-group they do so relatively generously, whereas they do not do so with children from an out-group (Fehr et al., 2008). In general, preschool children begin to show an understanding of the group as a kind of collective agency, as they judge that being a member of a task group means both that the individual wants to be a member and that the group wants her to be a member as well (Noyes et al., 2017). And, critically, even children's basic sense of social identity is group-minded, as (i) they care more about their individual reputations with in-group than with out-group members (Engelmann et al., 2013), and (ii) they engage in active attempts to manage other people's evaluative judgments not just of themselves but of their in-group as such (Engelmann et al., 2018).

These findings and others like them (see Dunham et al., 2008, 2018, for reviews) establish that children after three years of age are tuning in to the group-level organization within which they live, and, arguably, this underlies their newfound understanding of and relating to things in an objective and/or normative manner: this is how things are (for “us”) and this is how things are done (by “us”). Note that, in this view, in-group favoritism (and out-group disfavoritism) is basically a scaled-up version of children's tendency to favor collaborative partners over non-collaborators (i.e., free riders).

Social Norms. A basic requirement of membership in a cultural group is conformity, including to its social norms. From early in development young children imitate the actions of others, but by three years of age they are actually conforming the group by overriding their own individual preferences to do what others are doing (which other apes do not do; Haun & Tomasello, 2011; Haun & Over, 2014). This conformity often takes on a kind of objectifying or normative quality. For example, 3-year-old children engage in so-called overimitation, in which they copy aspects of adult behavior that are clearly not related to goal attainment (Lyons et al., 2007). One interpretation of this behavior is that when children see an adult performing an instrumental action with extra unnecessary flourishes, they do not interpret this as an individual idiosyncrasy of the actor, but rather as a manifestation of “how it is done” in the culture (Keupp et al., 2013).

An especially important indicator of preschoolers' emerging understanding of the group-mindedness of social norms is their proclivity not just to conform to them but to enforce them on others. From around 3 years of age, when children detect a social norm or rule violation, they protest, often normatively (Rakoczy et al., 2008; Schmidt et al., 2011, 2016a; see Schmidt & Tomasello, 2012, for a review). They presumably are motivated by something like a concern for how things are going in the group in general, as indicated by their use of normative language. This normative language takes one of

two forms: (i) an expression that this is how "one" should do it; or (ii) an even more objectifying expression that "this is how it is done" (Koymen et al., 2015). This language makes it clear that children are not just expressing their personal preference or desire, but rather they are referencing the group's normative standards that apply to all group members alike. Interestingly, when adults teach children about the world (in natural pedagogy; Csibra & Gergely, 2009), this same generic, universalizing mode operates (e.g., "To open these kinds of things, you must twist them like this"), and so children take the pedagogy to apply not just to the items indicated but generically to all kind-relevant agents and actions (Butler & Markman, 2014; Butler & Tomasello, 2016). The protesting child is, as it were, representing the group and its interests.

Importantly, recent research has found that preschool children also understand the group-relativity of some types of social norms. That is, specifically, when a perpetrator breaks a conventional norm, children enforce the norm if and only if the actor is an in-group member, since in-group members are within the purview of the norm and should know better (whereas out-group members are not; Schmidt et al., 2012). But when a perpetrator breaks a moral norm – specifically, by harming an in-group member – these same children enforce the norm on both in- and out-group individuals alike. Presumably, this differentiation of norm types (see also Turiel, 2006) reflects children's understanding that breaking moral norms represents a threat to the well-being of the group, whereas conventional norms are just how those who identify with the group behave to coordinate and self-identify. Interestingly, when children explain to an in-group peer the reasons for a third-party's norm violation, they explain a conventional violation by attempting to justify the rule, but they explain a moral violation by just stating the offense, assuming that it is clear that it violates a value that "we" all already share as members of the group (Mamman et al., 2018).

Finally, by around 5 years of age, children in small groups can create novel social norms for themselves by agreement (i.e., in the context of a novel game; Goeckeritz et al., 2014; Hardecker et al., 2017). Having created a norm, when new children come to play with the same materials, the creators enforce these self-invented rules on the new children normatively (e.g., "You have to play it this way"). And when someone agrees to a norm but then breaks it, children protest normatively, whereas when someone who has not agreed to the norm does exactly the same thing they do not protest (Schmidt et al., 2016b). The ways that young children understand and deal with the norms that they have created themselves - simply on the basis of agreement with peers with no authority figure involved - suggests an understanding that the agreements that constitute social norms produce individual rights and obligations. Such collective agreements or commitments may be seen as scaled-up versions of the joint commitments characteristic of joint intentional collaboration at the first step of our account.

Obligation. Together, these various lines of research suggest that young children's developing sense of belonging to a cultural group and their developing sense of morality, including a sense of obligation, are all of a piece. Children are loyal to the in-group, they feel guilty and make amends for the acts of the in-group, they share fairly

with the in-group, they conform to the in-group, they enforce conventional norms especially on the in-group, and they care more about their reputation with the in-group - as opposed to their general lack of interest in any out-group. In-group members thus constitute the child's sociomoral world, and others are simply outside that world. And so children make "objective" normative judgments about how things should be done, with the reference group being the members of the sociomoral world that they inhabit. They are objective about their known worlds. Later, school-age children come to understand that there may be other sociomoral worlds with their own norms (Schmidt et al., 2017).

With regard to the understanding of obligations specifically, preschool children appear to believe that members of social groups have obligations to one another that they do not have to out-group members. For example, Rhodes (2012) introduced 3- to 4-year-old children to two novel groups of characters (*flurps* and *zazes*). She then asked them to predict each group's behavior both toward in-group and out-group members. The children expected the in-group members not to harm one another (whereas it was less bad for them to harm out-group members), the authors' interpretation being that children judge that in-group members have an obligation not to harm one another that they do not have to outsiders. As a kind of control observation, children did not have different expectations for within and between group behaviors with regard to helping. So it is not just that they think individuals are "nicer" to in-group members, it is specifically about their being obliged to treat in-group members in special ways. Related studies have found similar results when children are asked to explain (rather than predict) the behavior of individual *flurps* and *zazes* toward in-group and out-group members (e.g., Chalik & Rhodes, 2015; Rhodes, 2014; Rhodes & Chalik, 2013). In a review, Rhodes and Wellman (2016, p. 195) state:

On this account, it is only those behaviors that children construe as obligatory that are shaped by their representations of social groups. Thus, children view people as obligated not to harm members of their own group, and because this is about an obligation, they do not extend this notion across group boundaries. In contrast, they fail to – at least at early ages – view pro-social actions as falling under the same scope of obligation, and thus, do not make group-based predictions about these types of behavior.

The hypothesis is thus that preschool children are beginning to do the same thing at this second step in our account, with regard to groups, that they were doing at the first step, with regard to collaborative partners. Namely, they are judging that obligations apply to those with whom they can and do form a "we", either within a collaborative partnership or within a cultural group. The fact that these same children do not expect in-group members to help one another more than out-group members provides a kind of control observation that it is not just more positive feelings toward in-group members; it is about obligation specifically. As development proceeds, children's understanding of "objective" standards that should apply universally to all rational/moral beings persists; it is just that, for some individuals at least, their

understanding of who is included in the class of rational/moral beings widens and relativizes.

Cross-cultural Variation. Recent research has begun to explore cross-cultural differences in the development of children's cooperation and morality. The overall pattern with respect to prosocial sharing, for example, is that children are quite similar at younger ages, and then during middle childhood (roughly, early school-age), their behavior begins to diverge based on the different social norms of the different cultural groups to which they belong (House et al., 2012; 2013; House & Tomasello, 2018). As another example, Schaefer et al. (2015) found that 4- to 11-year-old children from three different cultural groups (one WEIRD and two small-scale African cultures) took into account work-based merit (i.e., who produce the most resources) in culturally specific ways (some giving it more weight than others) when deciding how to divvy up collaboratively acquired resources. And Kanngießer et al. (in press) found that while children from several very different cultural contexts all respected the property of others, there was at the same time cultural variation in how much they did so.

These cross-cultural differences are consistent with the current ontogenetic account because they only concern children at our second, group-minded step as they gradually tune into and learn the particular social norms of their particular culture (leading to such things as “guilt cultures” vs. “shame cultures”). If studies were conducted that found cultural differences in children’s moral actions and judgments prior to three years of age, those would call into question the current ontogenetic story. However, there is very little cross-cultural data on children this young, as in many small-scale cultures toddlers are notoriously shy with adults of all kinds. In any case, the current ontogenetic account cannot be considered to be a universally valid account of the human species so long as the only children for whom we have rich data across ontogeny are from WEIRD cultures.

2.2.3. *A Second Interim Explanation*

In the case of joint intentionality and second-personal responsibility, our developmental explanation was about the most basic foundations of human morality, namely, how, through joint intentional collaboration, young children: (i) construct a sense of self-other equivalence, leading to the judgment that their partner is an equally deserving individual worthy of equal respect; and (ii) self-regulate the collaboration with their partner collaboratively (as a joint agent), leading to a kind of we > me valuation, creating for both partners the possibility of second-personal protest that both see as legitimate since they both affirmed it initially as part of their we-constituting joint commitment. The internalization of this dynamic constitutes the individual's feeling of second-personal responsibility to their collaborative partner (and not to non-collaborating outsiders). Then, in the second step of our account, 3- to 5-year-old children develop another way of relating to others cooperatively, namely, they begin to have a sense of themselves as members of the cultural group with which they identify.

That is, they have developed a second sense of “we”, applying not just to their partner of the moment, but also to their cultural group. Tomasello (2019) argues that there is very likely a strong evolutionary/maturational component to the emergence of this new group-minded form of social cognition at around 3 years of age.

In joint intentional activities, young children learn role ideals and standards which they strive to live up to. In a cultural group, while there are many specific cultural roles, the basic role is as a group member, which implies ideals and standards that individuals must live up to in order to have the status of members of the group in good standing. And just as in joint intentional collaboration there is collaborative self-regulation, in cultural groups there is a kind of group-level self-regulation via social norms. The crucial question is why the child buys into these social norms, which she had no part in creating, as legitimate. One possibility – noted above – is that the child identifies with her cultural group; she values being seen by others, and seeing herself, as a Erewhonian (or whatever), constituted by those of us, going back to our ancestors, who have certain distinctive ways of doing things. The child, in a sense, comes to consider herself a kind of co-author of the social norms (analogous to her co-authorship of a joint commitment): “we” Erewhonians created these social norms.

This moral identity view fits very well with the analysis of Korsgaard (1996), who proposes that “an obligation always takes the form of a reaction against the threat of a loss of identity” (p. 102). Indeed, one can imagine that this threat to self contributes to the sense of “objectivity” that goes along with a mature sense of obligation: I am obliged to conform and identify with those around me or else I really and truly, objectively, will cease to be who I am in the group. The social reference point for universal obligations is thus not actual human beings, however numerous, but the universalization “anyone who would be one of us”. Although this takes place within the cultural group, the psychology is still a universalizing psychology because in the ancestral state early humans thought of themselves and their compatriots as the only true persons, with outsiders being barbarians. This means that the mature sense of obligation is tied to shared agency even in this case. Thus, we may ask about our obligation (as opposed to our sympathy) to individuals from outside our moral community. Do we have obligations to invading Martians? We may feel sympathy for them in some situations, but do we feel a sense of obligation to them? If we fail to cooperate with them, do we owe them an apology or excuse? And what about other animals? Some people include them in the moral community and feel that we have obligations to them, but if we fail to cooperate with an animal do we owe it an apology or excuse? The point is that the sense of obligation (in contrast to the sense of sympathy) only operates within one's moral community, which is best thought of, in evolutionary perspective, as a kind of collective agency (which different cultural groups may construct for themselves somewhat differently).

Throughout this whole process, the role ideals and standards that individuals strive to meet must be viewed as legitimate, as warranted, or else meeting these ideals and standards is simply strategic. And so when children follow the ways of the cultural

group they are not just conforming strategically; and indeed they do not conform to the behavior of peers when that behavior does not accord with their own understanding of who they are as moral beings in the culture. For example, when a 5-year-old child observes three other children being callous to a needy peer, many of them do not follow along but behave prosocially in a way that confirms their own moral identity (Engelmann et al., 2016). Further, when children do not meet the group's normative ideals or standards, they do not always try to cover up (though they may on occasion), but rather they feel guilty and attempt to make reparations (Vaish et al., 2016). Guilt, as opposed to embarrassment or shame or regret, derives precisely from the fact that one knows one deserves the censure from the point of view of "we" in the group. The internalization of this social-interactive process is what creates the human sense of obligation - with a kind of universal, "objective" application - which derives its special force from its self-affirmed legitimacy rooted in one's sociomoral identity.

3. Obligation as Collaborative Self-Regulation

Some evolutionary theorists (e.g., de Waal, 2006) have proposed that tit-for-tat reciprocity between individuals - you scratch my back and I'll scratch yours - is the ultimate source of the human sense of fairness or justice (and so, presumably, obligation). But while cooperative acts may engender a sense of gratitude, and so motivate a return cooperative act, it is difficult to see how this process could engender a sense of obligation - unless, of course, the initial cooperator feels that his cooperative act gives him a claim on the recipient, and the recipient feels the force of this claim in the sense of its moral legitimacy. In the current view, a sense of legitimacy can only arise if the two of them feel that they are interdependent parts of some larger social partnership or group ("we") that is regulating the interaction. Even when young children practice reciprocity in the sharing of resources (e.g., Warneken & Tomasello, 2013), they do not apologize, justify, or make excuses if they do not reciprocate - they do not self-regulate the interaction morally - suggesting that they are motivated by something more like gratitude than obligation specifically. Thus, although tit-for-tat reciprocity can, if the partners think of themselves as an interdependent "we", play a role in generating obligations, it is neither necessary nor fundamental.⁴

The social exchange theory of Cosmides and Tooby (2004) locates the origin of humans' sense of obligation in the evolutionary psychology of humans trading of goods with one another. Their theory goes beyond simple tit-for-tat reciprocity in positing an agreement-like structure that underlies obligations between partners ("the obligation schema"). But the theory does not explore in any depth the origin of a sense of obligation, in that it does not seek to identify the psychological components out of which it is built either phylogenetically or ontogenetically. Why does social exchange generate obligations in the first place? Our interpretation, again, similar to that for simple reciprocity, is that social exchange creates obligations precisely to the extent that the parties to the exchange see it as a cooperative

⁴ Also important is the fact that there is much cultural variability in how humans view reciprocity: some so-called gift cultures act as though every prosocial act creates an obligation in return, whereas many other cultures do not. The tie to obligation is thus culturally contingent.

agreement that “we” have created and that “we” cooperatively self-regulate (e.g., by protest, justification, and apology).

In contrast, in our shared intentionality hypothesis we have focused on where the sense of obligation comes from phylogenetically and ontogenetically, that is, the more basic psychological components from which it is built. In support of the hypothesis we have presented two sets of experimental findings. First, children approaching their third birthdays treat collaborative partners especially prosocially, which other apes do not do. When their collaborative partner reneges on a joint commitment, children protest using normative language, and when the child herself wants to break a joint commitment, she knows that she must acknowledge the breaking or provide a good excuse (and she requires her partner to do the same) - because the joint commitment binds her obligatorily. In sharing resources, 3-year-olds feel obligated to share fairly with a collaborative partner but not with other children, and this is not just enhanced generosity, but a sense of equality and mutual respect. Thus, it would seem that from around their third birthdays, young children feel that they are obliged to treat their collaborative partners, but not others, with fairness and mutual respect.

Second, from soon after their third birthdays, children begin to conform to the practices of their cultural group, copying an adult’s idiosyncratic behavior even when they know it is not instrumentally necessary - arguably because they think it is normative for the group. This deference to the group also includes conformity to the group's social norms, but, beyond this, preschoolers also insist on conformity to the group's social norms from others (again using normative language). And they enforce conventional norms selectively on in-group rather than out-group members, demonstrating an understanding that such norms are group relative. By five years of age, children enforce norms even when they have created them themselves, demonstrating an understanding that the normative force emanates not from adult authority (as suggested by Piaget, 1932), but rather from the process of social agreement. Children of this age identify with their cultural group and thereby feel its normative force, and they judge that in-group compatriots are obliged to one another in ways that they are not to outsiders.

The most general point is that the conditions which give rise to a feeling of obligation are social in some very special ways. If I have an obligation to you then you have a right to expect me to fulfill it: you have a claim on me, and I feel the possibility of my debt to you if I do not meet it. *But you cannot make this claim on me unilaterally; simply saying you have a claim on me does not create an obligation.* To feel a sense of obligation I must recognize the legitimacy of your claim. In the current account this recognition of legitimacy occurs in, and only in, in the context of some social body, “we”, that unites us and supports this legitimacy normatively - because we not only collaborate to pursue shared goals but we also collaborate to self-regulate the collaboration. The prototypical situation for a feeling of obligation is thus one in which you and I voluntarily enter into an joint agency (including in creating and self-regulating an agreement) in which we both recognize the legitimacy of the claim-obligation relationship that is thereby created. In the legal world, a contract is a “chosen obligation”.

But why cannot the individuals of other animal species - including those, such as chimpanzees, with very complex social lives - create with one another a shared agency and so a claim-obligation relationship? Our hypothesis is that humans, and only humans, have evolved to collaborate both with other individuals as partners and within a cultural group with its various social roles. Participation in such cooperative activities leads individuals to view their collaborative partners and cultural compatriots, and only their collaborative partners and

cultural patriots, as mutually respected equals (i.e., as second-personal agents) in a shared agency, whose claims must be respected if the partnership or cultural membership is to be maintained. The social comparison in which self and other are seen as equivalent parts of a larger “we” - each worthy of equal respect in a way that free riders or out-group members are not - is not a social preference but the recognition of a fact (i.e., for individuals who collaborate in the requisite ways). Moreover, human individuals tend to identify with this “we” such that their role in the partnership or cultural group is at the core of their basic sociomoral identities, which makes respecting the partners’ claims vital to one’s very self. The sense of obligation to treat others fairly is nothing more or less, in our view, than the motivational force that accompanies the creation of a joint or collective agent “we”, cementing the partners’ or compatriots’ cooperative or cultural identities together in a normative bond.

The upshot is that to describe the architecture of the human sense of obligation in the mature individual, we need two key components. The first begins with the general primate sense of “instrumental pressure”, in which the individual feels a kind of rational force to do what it needs to do to achieve its goal: to get that apple, I *must* climb that tree. But what has happened in humans is that, in the context of their species-unique forms of interdependent collaboration, this kind of individual practical rationality has been transformed into a new kind of cooperative rationality. Within a collaborative partnership, I have not only instrumental goals toward the outside world, but also interpersonal goals with respect to my valued partner and our partnership. And, in the context of an interdependent partnership in which we both understand that the normative standards of role performance apply to us both equally, I cannot help but see my partner as an equal, and she me, engendering the sense of mutual respect that generates excuses, justifications, and apologies for breaches. For us to harvest those apples together cooperatively, and to maintain our cooperative relationship over time, we must each play our mutually known roles up to standard, while at the same time respecting one another by living up to our commitments, by asking permission to break our commitment, and by sharing the spoils fairly.

The second component begins with the general primate process of self-regulation, in which the individual monitors its own perceptions and actions from an executive level in order to learn and adapt. But what has happened in humans is that individuals living in an interdependent world must care about what others think of them - as collaborative partners and cultural compatriots - and so there occurs a kind of social self-regulation, which is unique to the species (Englemann et al., 2012). The type of social self-regulation that is most critical for understanding the human sense of obligation is conducted by the “we” that an individual forms with a collaborative partner or cultural compatriot. Critically, as opposed to reputation-based theories of the evolution of morality, the key is not just what “they” will think of me if I do X (which is strategic and associated with the emotion of shame), but what “we” will think of me if I do X based on our shared values (which is moral and associated with the emotion of guilt). One’s sociomoral identity derives from participation in and contributions to a “we”, and to maintain that identity one simply *must* respect the judgment of that “we” over my individual judgments (a we > me valuation). The sense of obligation has a coercive (negative) quality because it is a response to a claim, nay a threat, to my identity from the valued partners and/or compatriots with whom I am interdependent. I internalize this normative pressure, in Vygotskian fashion, and use it to avoid feel guilty and losing my sense of who I am socially.

Human individuals in shared agencies thus feel both and at the same time (i) a self-other equivalence to partners/compatriots (you = me) and (ii) a subordination to the “we” involved

(we > me). And so, in highly summarized form we may say that the human sense of obligation is the internalized social/rational pressure from a joint or collective self-regulating agent "we" – which comprises myself in my sociomoral identity and one or more respected second-personal agents with legitimate claims on me as an interdependent collaborative partner or cultural compatriot - to do what "we" expect me to.

4. Conclusion

Hume was the first to recognize that the morality of justice/fairness, and the sense of obligation that goes along with it, is structured by a certain kind of social interaction and relationship. He thought that this was societal conventions, but it turns out there are other less conventionalized forms of collaborative social engagement that still represent enough of an agreement, even if implicit, to induce a feeling of obligation to one's partner. Darwall and other social-relational theorists have focused on the "spectator" aspect of Hume's approach, and its inability to account for some of the key features of a sense of obligation, including its pre-emptiveness, its negative valence, and its imprimatur of legitimacy. They have attempted to explicate various features of the second-personal standpoint, and the nature of second-personal agents, that correct this shortcoming. However, these social-relational theorists have not given much attention to the underlying psychology that gives rise to the forms of social engagement requisite for the individual to feel a sense of obligation to other persons.

Here we have retained the social-relational philosophers' insights about the second-personal standpoint, but have tried to dig deeper into the social-interactive situations in which the human sense of obligation arises and operates. We have done that by focusing on processes of shared intentionality as a mode of social engagement that creates a joint or collective agent "we", especially as young children are first learning to collaborate. Convention or explicit agreement can play a role in interactions structured by shared intentionality, but they need not.

Even 3-year-old children behave in ways that evidence a sense of responsibility to their collaborative partner - not only deferring to their partner but also by "taking leave", making excuses, and feeling guilty, as appropriate - since members of a joint agency both see one another as equals, and they subordinate their own goals to those of the partnership. After three years of age children's emerging group-mindedness brings with it the universal and objective sense of obligation characteristic of adults, whose social identity comes from their role as members of their culture and/or moral community. Only in such situations of shared agency with others do individuals imbue the claims of others with the sense of legitimacy characteristic of a sense of obligation.

We began our account with the observation that the sense of obligation is presumably a motivation (what else?), but that it is a decidedly peculiar one. Let us end with a characterization of this peculiarity by means of a classification. In the current view, the human sense of obligation is best considered as a kind of self-conscious motivation, analogous to the self-conscious emotion of guilt. And indeed obligation is intimately related to guilt, as most often guilt is about not living up to one's obligations. Both guilt and obligation have to do with one's sense of who one is: "I must do this because I could not live with myself if I did not"; "I must not do it because that would just not be me". The sense of obligation may thus be considered as a self-conscious motivation because it derives from a kind of threat from a "we", into which one has entered, that one might lose one's cooperative or moral identity within that "we". It is not clear whether there are other motivations that one might also want to call self-conscious, other

than closely related motives such as a sense of responsibility or the like. But in either case, this would seem to be an apt categorization of the peculiar nature of the sense of obligation as one of humans' most important moral motivations.

In all, recognizing the insights of the second-personal philosophers and related approaches, and undergirding them with the psychological foundation of shared intentionality, provides, we would argue, the most comprehensive account to date of the underlying psychology of the human sense of obligation.

References

- Baumard, N., André, J. B., & Sperber, D. (2013). A Mutualistic Approach to Morality. *Behavioral and Brain Sciences*, 36(1), 59–122.
- Bratman, M. (1992). Shared co-operative activity. *Philosophical Review*, 101(2), 327–341.
- Bratman, M. E. (2014). *Shared Agency: A Planning Theory of Acting Together*. New York, NY: Oxford University Press.
- Brosnan, S. F., & de Waal, F. B. M. (2003). Monkeys reject unequal pay. *Nature*, 425, 297–299.
- Butler, L. & Markman, E. (2014). Preschoolers use pedagogical cues to guide radical reorganization of category knowledge. *Cognition*. 130(1), 116-127.
- Butler, L. and Tomaello, M. (2016). Two- and 3-year-olds integrate linguistic and pedagogical cues in guiding inductive generalization and exploration. *Journal of Experimental Child Psychology*, 145, 64-78.
- Chalik, L. & Rhodes, M. (2015). The communication of naive theories of the social world in parent–child conversation. *Journal of Cognition and Development*, 16, 719–741.
- Corbit, J., McAuliffe, K., Callaghan, T., Blake, P.R., & Warneken, F. (2017). Children’s collaboration induces fairness rather than generosity. *Cognition*, 168, 344-56..
- Cosmides, L., & Tooby, J. (2004). Social Exchange: The Evolutionary Design of a Neurocognitive System. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences*. MIT Press.
- Csibra, G., & Gergely, G. (2009). Natural pedagogy. *Trends in Cognitive Sciences*, 13(4), 148–153.
- Darwall, S. (2006). *The Second-Person Standpoint: Respect, Morality, and Accountability*. Cambridge, MA: Harvard University Press.
- Darwall, S. (2012). Bipolar obligation. In *Morality, Authority, & Law*. Oxford University Press.
- Darwall, S. (2013). *Essays in Second-Personal Ethics*. [Vol. 1: Morality, Authority, and Law]. Oxford, UK: Oxford University Press.
- de Waal, F. B.M. (2006). *Primates and Philosophers: How Morality Evolved*. Princeton, NJ: Princeton University Press.
- Dunham, Y. (2018). Mere membership. *Trends in Cognitive Sciences*, 22, 780 -793.
- Dunham, Y., Baron, A. S., & Banaji, M. R. (2008). The development of implicit intergroup cognition. *Trends in Cognitive Sciences*, 12(7), 248–253.
- Engelmann, J. M., Herrmann, E., & Tomasello, M. (2012). Five-year olds, but not chimpanzees, attempt to manage their reputations. *PLoS ONE* 7(10): e48433.
- Engelmann, J. M., Over, H., Herrmann, E., & Tomasello, M. (2013). Young children care

more about their reputation with ingroup members and potential reciprocators. *Developmental Science*, 16(6), 952–958.

Engelmann, J., Rapp, D., Herrmann, E., & Tomasello, M. (2016). Young children (sometimes) do the right thing even when their peers do not. *Cognitive Development*, 39, 86–92.

Engelmann, J., Rapp, D., Herrmann, E., & Tomasello, M. (2018). Concern for group reputation increases prosociality in young children. *Psychological Science*, 29, 181–90.

Engelmann, J. & Tomasello, M. (in press). Children's sense of fairness as a sense of equal respect. *Trends in the Cognitive Sciences*.

Fehr, E., Bernhard, H., & Rockenbach, B. (2008). Egalitarianism in young children. *Nature*, 454, 1079–1083.

Gilbert, M. (1990). Walking together: A paradigmatic social phenomenon. *Midwest Studies in Philosophy*, 15(1), 1–14.

Gilbert, M. (2014). *Joint Commitment: How We Make the Social World*. New York, NY: Oxford University Press.

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

Gräfenhain, M., Behne, T., Carpenter, M., & Tomasello, M. (2009). Young children's understanding of joint commitments. *Developmental Psychology*, 45(5), 1430–1443.

Gräfenhain, M., Carpenter, M., & Tomasello, M. (2013). Three-year-olds' understanding of the consequences of joint commitments. *PLoS One* 8(9): e73039.

Greenberg, J.R., Hamann, K., Warneken, F., & Tomasello, M. (2010). Chimpanzee helping in collaborative and noncollaborative contexts. *Animal Behaviour*, 80(5), 873–880.

Groce, P., Rossano, F., & Tomasello, M. (2015). Procedural justice in children: Preschoolers accept unequal resource distributions if the procedure provides equal opportunities. *J. of Experimental Child Psychology*, 140, 197–210.

Groce, P., Rossano, F., & Tomasello, M. (2018). Young children are more willing to accept group decisions in which they have had a voice. *Journal of Experimental Child Psychology*, 166, 67–78.

Hamann, K., Warneken, F., Greenberg, J., & Tomasello, M. (2011). Collaboration encourages equal sharing in children but not chimpanzees. *Nature*, 476, 328–331.

Hamann, K., Warneken, F., & Tomasello, M. (2012). Children's developing commitments to joint goals. *Child Development*, 83(1), 137–145.

Hardecker, S., Schmidt, M., & Tomasello, M. (2017). Children's Developing Understanding of the Conventionality of Rules. *Cognition and Development*, 18(2), 163–188.

- Haun, D. B. M., & Tomasello, M. (2011). Conformity to peer pressure in preschool children. *Child Development*, 82(6), 1759–1767.
- Haun, D., & Over, H. (2014). Like me: A homophily-based account of human culture. In P. J. Richerson & M. Christiansen (Eds.), *Cultural Evolution* (pp. 75–85). Cambridge, MA: MIT Press.
- Hohfeld, WN. (1923). *Fundamental legal concepts as applied in judicial reasoning and legal essays*. New Haven: Yale University Press.
- House, B.R., Henrich, J., Brosnan, S.F., & Silk, J.B. (2012). The ontogeny of human prosociality: behavioral experiments with children aged 3 to 8. *Evolution and Human Behavior*, 33, 291–308.
- House, B.R., Silk J. B., Henrich J., Barrett H.C., Scelza B.A., Boyette A.H., Hewlett, B.S., & Laurence, S. (2013). Ontogeny of prosocial behavior across diverse societies. *Proceedings of the National Academy of Sciences*, 110, 14586–14591.
- House, B. & Tomasello, M. (2018). Modeling social norms increasingly influences costly sharing in middle childhood. *Journal of Experimental Child Psychology*, 171, 84–98.
- Hume, D. (1751/1957). *An Enquiry concerning the Principles of Morals*. New York: Bobbs-Merrill.
- Ibbotson, P. (2014). Little dictators: a developmental meta-analysis of prosocial behavior. *Current Anthropology*, 55(6), 814–821
- Jensen, K., Call, J., & Tomasello, M. (2007). Chimpanzees are rational maximizers in an ultimatum game. *Science*, 318(5847), 107–109.
- 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., Svetlova, M., & Tomasello, M. (2019). Three- and 5-year-old children's understanding of how to dissolve a joint commitment. *Journal of Experimental Child Psychology*, 184, 34–47.
- Kachel, U., & Tomasello, M. (2019). 3- and 5-year-old children's adherence to explicit and implicit joint commitments. *Developmental Psychology*, 80–88.
- Kanngasser, P., Rossano, F., & Tomasello, M. (in press). Children, but not great apes, respect ownership. *Developmental Science*.
- Keupp, S., Behne, T. & Rakoczy, H. (2013). Why do children over-imitate? Normativity is crucial. *Journal of Experimental Child Psychology*, 116, 392–406
- Korsgaard, C. (1996). *The Sources of Normativity*. Cambridge, UK; New York, NY: Cambridge University Press.
- Koymen, B., Schmidt, M., Lieven, E., & Tomasello, M. (2015). Teaching versus enforcing norms in preschoolers' peer interactions. *J. of Experimental Child Psychology*, 135, 93–101

- Lyons, D.E., Young, A.G., & Keil, F.C. (2007). The hidden structure of overimitation *Proceedings of the National Academy of Sciences*, 104(50), 19751–19756
- Mamman, M., Koymen, 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.
- Melis, A. P., Hare, B., & Tomasello, M. (2006). Engineering cooperation in chimpanzees: tolerance constraints on cooperation. *Animal Behaviour*, 72(2), 275–286.
- Melis, A. P., Schneider, A.-C., & Tomasello, M. (2011). Chimpanzees share food in the same way after collaborative and individual food acquisition. *Animal Behaviour*, 82(3), 485–493.
- Melis, A., Floedl, A., Tomasello, M. (2015). Non-egalitarian allocations among preschool peers in a face-to-face bargaining task. *PLOS One*,10(3):e0120494.
- Misch, A., Over, H., & Carpenter, M. (2014). Stick with your group: Young children's attitudes about group loyalty. *Journal of Experimental Child Psychology*, 126, 19-36.
- Nagel, T. (1970). *The possibility of altruism*. Princeton, NJ: Oxford University Press.
- Noyes, A., & Dunham, Y. (2017). Mutual intentions as a causal framework for social groups. *Cognition*, 162, 133-142.
- Over, H. (2018). The influence of group membership on young children's prosocial behaviour. *Current Opinion in Psychology*, 20, 17-20.
- Over, H., Vaish, A., & Tomasello, M. (2016). Do young children accept responsibility for the negative actions of ingroup members? *Cognitive Development*, 40, 24-32.
- Piaget, J. (1932). *The moral judgment of the child*. New York: The Free Press.
- Rakoczy, H., Warneken, F., & Tomasello, M. (2008). The sources of normativity: Young children's awareness of the normative structure of games. *Developmental Psychology*, 44(3), 875–881.
- Rakoczy, H., Kaufman, M., Lohse, K. (2016). Young children understand the normative force of standards of equal resource distribution. *Journal of Experimental Child Psychology*, 150, 396-403.
- Rekers, Y., Haun, D., & Tomasello, M. (2011). Children, but not chimpanzees, prefer to forage collaboratively. *Current Biology*, 21, 1756-58
- Rhodes, M. (2012). Naive theories of social groups. *Child Development*, 83, 1900–1916
- Rhodes, M. (2014). Children's explanations as a window into their intuitive theories of the social world. *Cognitive Science*, 38, 1687–1697.
- Rhodes, M., & Chalik, L. (2013). Social categories as markers of intrinsic interpersonal obligations. *Psychological Science*, 24(6), 999–1006.
- Rhodes, M., & Wellman, H. M. (2017). Moral learning as intuitive theory revision.

Cognition, 167, 191–200.

- Scanlon, T. (1998). *What We Owe to Each Other*. Cambridge, MA: Belknap Press of Harvard University Press.
- Schaefer, M., Haun, D., & Tomasello, M. (2015). Fair is not fair everywhere. *Psychological Science*, 26, 1252-1260.
- Schmidt, M., Rakoczy, H., Tomasello, M. (2011). Young children attribute normativity to novel actions without pedagogy or normative language. *Developmental Science*, 14, 530-539.
- Schmidt, M., Butler, L., Heinz, J., & Tomasello, M. (2016a). Young children see a single action and infer a social norm. *Psychological Science*, 27, 1360-1370.
- Schmidt, M., Cabrera, I., & Tomasello, M. (2017). Children's developing metaethical judgments. *Journal of Experimental Child Psychology*, 164, 163-77.
- Schmidt, M., Rakoczy, H., & Tomasello, M. (2012). Young children enforce social norms selectively depending on the violator's group affiliation. *Cognition*, 124, 325-33.
- Schmidt, M., Rakoczy, H., & Tomasello, M. (2016b). Young children understand the role of agreement in establishing arbitrary norms – but unanimity is key. *Child Development*, 87, 612-626
- Schmidt, M., & Tomasello, M. (2012). Young children enforce social norms. *Current Directions in Psychological Science*, 21(4), 232–236.
- Searle, J. (1995). *The construction of social reality*. New York, NY: The Free Press.
- Searle, J. (2010). *Making the social world: The structure of human civilization*. New York: Oxford University Press.
- Shaw, A., & Olson, K. (2014). Fairness as partiality aversion: The development of procedural justice. *Journal of Experimental Child Psychology*, 119, 40–53.
- Siposova, B., Tomasello, M., & Carpenter, M., (2018). Communicative eye contact signals a commitment to cooperate in young children. *Cognition*, 179, 192-201.
- Smith, A. (2013). Moral Blame and Moral Protest. In D. J. Coates, & N. A. Tognazzini (Eds.), *Blame: Its Nature and Norms* (pp. 27–48). New York, NY: Oxford University Press.
- Strawson, P. F. (1962). Freedom and Resentment. *Proceedings of the British Academy*, 48, 1–25
- Thompson, Michael (2004). "What Is It To Wrong Someone?: A Puzzle About Justice," in *Reason and Value: Themes from the Philosophy of Joseph Raz*, eds., J. Wallace, P. Pettit, S. Scheffler, M. Smith. Oxford: Oxford U. Press.
- Tomasello, M. (2014). *A natural history of human thinking*. Cambridge, MA: Harvard University Press.

- Tomasello M (2016) *A natural history of human morality*. Cambridge, MA: Harvard University Press.
- Tomasello, M. (2018). The normative turn in early moral development. Special issue of *Human Development*, 61, 248–263.
- Tomasello, M. (2019). *Becoming human: A theory of ontogeny*. Harvard University Press.
- Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences*, 28, 675 – 691.
- Turiel, E. (2006). The development of morality. In W. Damon, & R. M. Lerner (Eds.), *Handbook of Child Psychology III: 13*. New York, NY: Wiley.
- Ulber, J., & Tomasello, M. (2017). Young children, but not chimpanzees, are averse to disadvantageous and advantageous inequities. *Journal of Experimental Child Psychology*, 155, 48–66
- Vaish, A., Carpenter, M., & Tomasello, M. (2016). The early emergence of guilt-motivated prosocial behavior. *Child Development*, 87, 1772–1782.
- Wallace, R. Jay (1994). *Responsibility and the Moral Sentiments*. Cambridge, MA: Harvard University Press.
- Wallace, J. (in press). *The Moral Nexus*. Princeton: Princeton University Press
- Warneken, F., Chen, F., & Tomasello, M. (2006). Cooperative Activities in Young Children and Chimpanzees. *Child Development*, 77(3), 640–663.
- Warneken, F., Lohse, K., Melis, A. P., & Tomasello, M. (2011). Young children share the spoils after collaboration. *Psychological Science*, 22(2) 267–273.
- Warneken, F., Gräfenhain, M., & Tomasello, M. (2012). Collaborative partner or social tool? New evidence for young children’s understanding of joint intentions in collaborative activities. *Developmental Science*, 15, 54–61.
- Warneken, F. & Tomasello, M. (2013). The emergence of contingent reciprocity in young children. *J. of Experimental Child Psychology*, 116, 338–350.
- Wittig, M., Jensen, K., & Tomasello, M. (2013). 5-year-olds understand fair as equal in a mini-ultimatum game. *Journal of Experimental Child Psychology*, 116(2), 324–337.