

cambridge.org/bbs

Target Article


Cite this article: Tomasello M. (2020) The moral psychology of obligation. *Behavioral and Brain Sciences* 43, e56: 1–58. doi:10.1017/S0140525X19001742

Target Article Accepted: 19 May 2019
Target Article Manuscript Online: 28 May 2019
Commentaries Accepted: 20 September 2019

Keywords:

cooperation; fairness; morality; motivation; normativity; obligation; shared intentionality

What is Open Peer Commentary? What follows on these pages is known as a Treatment, in which a significant and controversial Target Article is published along with Commentaries (p. 12) and an Author's Response (p. 46). See bbsonline.org for more information.

Michael Tomasello^{a,b} 

^aMax Planck Institute for Evolutionary Anthropology, 04103 Leipzig, Germany and ^bDepartment of Psychology and Neuroscience, Duke University, Durham, NC 27708-0086

michael.tomasello@duke.edu

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, I 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 therefore be seen as a kind of self-conscious motivation.

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.

1. **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.
2. **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.

My aim in this target article 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? My 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. I 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

© The Author(s), 2019. Published by Cambridge University Press

CAMBRIDGE
UNIVERSITY PRESS

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. Although 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) argued 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; 2013a) (see also Wallace 1994; 2019).

MICHAEL TOMASELLO is professor of Psychology and Neuroscience at Duke University and emeritus director at the Max Planck Institute for Evolutionary Anthropology in Leipzig, Germany. His research interests focus on processes of cooperation, communication, and cultural learning in human children and great apes. He has been elected to the National Academy of Sciences of the United States, Germany, Hungary, and Sweden. His most recent books include *A Natural History of Human Thinking* (Harvard University Press, 2014), *A Natural History of Human Morality* (Harvard University Press, 2016), and *Becoming Human: A Theory of Ontogeny* (Harvard University Press 2019).

In focusing specifically on obligation, Darwall (2006; 2013a) has pointed 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’” (Darwall 2013c, p. 29). Each individual sees herself as “one mutually accountable agent among others” (Darwall 2013c, p. 7). Second-personal agents display “mutual respect between mutually accountable persons” (Darwall 2006, p. 36). Accordingly, 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 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 another's shoes), as well as “the capacities to make and regulate oneself by normative judgments about what attitudes are warranted” (Darwall 2013c, p. 47; 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 mutually 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], the being counts as second-personally competent. (Darwall 2013c, p. 47)

Darwall immediately then claims (following Strawson) that “second-personal competence is both necessary and sufficient for moral obligation” (Darwall 2013c, p. 47).

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 the following:

1. 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?
2. Where do second-personal agents come from, such that other animals and young infants, for example, are not second-personally competent?
3. 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

My 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 respectfully toward collaborative partners than toward others in various ways; 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). I will argue, at the end of the article, 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 (1995a; 2010), and as applied empirically by psychologists such as Tomasello (2014a; 2016; 2019). In both phylogeny and ontogeny, this unique social psychology unfolds in two key steps. The first 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 2014a; 2016; 2019). The essential point in the current context is that the joint intentional schema created the possibility of a morality of fairness among co-equal second-personal agents, who felt a sense of normative obligation to treat one another in mutually respectful ways.

2.1.1. Evolutionary background

The specific chronology 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 non-cooperation (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 I 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.

2.1.2.1. 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 (Rekers 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 just deserts. Thus, when two 3-year-olds are working together toward a joint goal, and, seemingly by accident, one child has access to a reward first, they nevertheless persist with the collaboration until the partner obtains a reward as well. They do not do this nearly as often in a control condition in which another child lacks a 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 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 the partner's welfare.

Consistent with this latter interpretation, children at around 3 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 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 the partner when she is delayed, repair some damage done by the partner, refrain from tattling on the partner, and perform the partner's role for her when she is unable (Gräfenhain 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 together 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 just to 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 his 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," and so on. 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 being normatively committed to their partner as well (see also Siposova 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 explicitly 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 (Gräfenhain 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.

2.1.2.2. 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 experimental 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 accept distributions readily only 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 (Engelmann & Tomasello 2019). Thus, when children receive, for example, one piece of candy, they are content, but when a collaborative partner at the same time receives five 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 five 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 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.

I will give an overall explanation for the developmental emergence of children's sense of obligation only after having examined older children's sense of collective agency and "objective" morality. For now, I 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 3 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) has called 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 1970b): 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: (1) both participants are equal causal forces in producing the mutually intended outcome; (2) both partners could switch roles as needed; and, most crucially, (3) 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.⁵

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 (1) 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 (2) 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 full-fledged "objective" morality, and for that we need a more 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, also known as 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 (read: 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, and so on. 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 1995a). 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” around the time of 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 – that is, 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, which we all agree was the right thing to do). 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.

2.2.2.1. 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; Dunham 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 & Dunham 2017). And, critically, even children’s basic sense of social identity is group-minded, as (1) they care more about their individual reputations with in-group than with out-group members (Engelmann et al. 2013); and (2) 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 (for reviews, see Dunham 2018; Dunham et al. 2008) establish that children after 3 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).

2.2.2.2. 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 3 years of age they are actually conforming to the group by overriding their own individual preferences to do what others are doing (which other apes do not do; Haun & Over 2014; Haun & Tomasello 2011). 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; for a review, see Schmidt & Tomasello 2012). 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: (1) an expression that this is how “one” should do it; or (2) 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 2007) 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 simply 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; Göckeritz 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 emerging 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.

2.2.2.3. 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 (and the in-group's reputation itself) – 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 in-group members as having an obligation not to harm one another, which 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, but that 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 (2017, 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 or especially. 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.

2.2.2.4. 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 pro-social 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 – Western, educated, industrialized, rich, and democratic – and two small-scale African cultures) took into account work-based merit (i.e., based on who produces the most resources) in culturally specific ways (some giving it more weight than others) when deciding how to divvy up collaboratively acquired resources. And Kanngisser et al. (2019) 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 concern children only 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 3 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: (1) construct a sense of self-other equivalence, leading to the judgment that their partner is an equally deserving individual worthy of equal respect; and (2) 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, although 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 an 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 to 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 of “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 to (as opposed to our sympathy for) 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. Interestingly, when young children practice reciprocity in the sharing of resources (e.g., Warneken & Tomasello 2013), they typically 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 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, namely, from the formation of shared agencies in collaboration and culture. In support of the hypothesis we have presented evidence that children approaching their third birthdays treat collaborative partners, as opposed to others, especially respectfully. They honor their joint commitments, ask permission of their partner to break them, respect protest against breaches, and provide a respectful excuse (and/or feel guilty) if they do break a norm. And they share collaboratively produced resources with their partner, but not others, with a sense of equality and mutual respect. And then, from soon after their third birthdays, children begin to conform to the practices of their cultural group and to insist on conformity to the group’s social norms from others (often using normative language). They enforce conventional norms selectively on in-group members, and, by 5 years of age, enforce even self-created norms on others, 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 judge that in-group compatriots are obliged to one another more than to outsiders. Other creatures do not experience a sense of obligation because they do not socially interact with one another in any of these ways.

The most general point is that the conditions which give rise to a feeling of obligation are interactive 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, 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 some kind of shared agency 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.”

And so, 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 individuals monitor their 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 (Engelmann 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 feeling guilty and losing my sense of who I am socially.

To summarize, we may thus 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. He thought that this was societal convention, 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.

In this article I 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. I 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 subordinate their own goals to those of the partnership. After 3 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.

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

Notes

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.


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 (for a review, see Tomasello 2016, pp. 32–34).

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.

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 pro-social act creates an obligation in return, whereas many other cultures do not. Reciprocity's tie to obligation is thus culturally contingent.

Open Peer Commentary

Differentiating between different forms of moral obligations

Rajen A. Anderson^a , Benjamin C. Ruisch^b
and David A. Pizarro^a

^aDepartment of Psychology, Cornell University, Ithaca, NY 14853-7601 and

^bDepartment of Psychology, Ohio State University, OH 43210.

raa255@cornell.edu <http://rajenanderson.com>

ruisch.3@osu.edu <http://benrusich.com>

dap54@cornell.edu <http://www.peezer.net>

doi:10.1017/S0140525X19002589, e57

Abstract

We argue that Tomasello's account overlooks important psychological distinctions between how humans judge different types of moral obligations, such as prescriptive obligations (i.e., what one *should* do) and proscriptive obligations (i.e., what one *should not* do). Specifically, evaluating these different types of obligations rests on different psychological inputs and has distinct downstream consequences for judgments of moral character.

Tomasello draws from research on nonhuman primates and human children to position moral obligation as a uniquely human motivation, arguing that the psychological construction of a shared agent "we" works to direct and self-regulate collaborative activities in humans, giving rise to a sense of moral obligation. However, we argue that this account, while illuminating, overlooks key distinctions between different kinds of moral obligations – and, importantly, the distinct psychological processes that they entail. Among the most central of these distinctions is that found between prescriptive obligations (i.e., obligations to engage in certain positive, beneficial behaviors; e.g., one should help others in need) and proscriptive obligations (i.e., obligations not to engage in certain negative, harmful behaviors; e.g., one should not hurt others). When we fulfill and adhere to prescriptive obligations, we generally earn praise and gratitude; when we violate proscriptive obligations, we generally earn blame and punishment. Additionally, failures to meet prescriptive obligations and non-violations of proscriptive obligations are relatively less relevant to observers (e.g., Haidt & Baron 1996).

Importantly, social psychological research has revealed differences between how prescriptive fulfillments and proscriptive violations are evaluated. Adults tend to judge proscriptive morality as concrete, mandatory, and duty-based, while viewing prescriptive morality as abstract, discretionary, and based in either duty or desire; even framing the same basic moral act as either prescriptive or proscriptive can change people's judgments (Janoff-Bulman et al. 2009). Additionally, adults treat beliefs about

immoral acts (i.e., violations of obligation) as more objective, agreed upon by others, and true than beliefs about moral acts (i.e., fulfillments of obligation; Goodwin & Darley 2012). On the whole, then, proscriptive obligations appear to have more psychological weight than prescriptive obligations.

It may be possible to reconcile these asymmetries within Tomasello's theoretical framework – for example, it may be that proscriptive obligations entail a greater sense of “we” in their formation and enforcement – but more work will be necessary to assess these possibilities.

Additionally, the process by which observers judge violations of obligation and fulfillments of obligation also appears to differ. For example, immoral, obligation-violating acts tend to elicit more causal attribution and counterfactual thinking than moral, obligation-fulfilling acts (Bohner et al. 1998; Bostyn & Roets 2016; Roesse & Olson 1997). In addition, although agent intentionality influences the perceived morality and immorality of both obligation-fulfilling acts and obligation-violating acts, intentionality appears to matter more for evaluating blameworthy acts than praiseworthy acts (Guglielmo & Malle 2019; Ohtsubo 2007; Pizarro et al. 2003). For example, both adults and children judge unintended, obligation-violating “side effects” (of a person's actions) to be blameworthy. However, they fail to judge unintended, obligation-upholding side effects to be equally praiseworthy (Knobe 2003a; 2003b; Leslie et al. 2006). People also more readily incorporate the magnitude of the consequences of proscriptive obligations than prescriptive obligations into their judgments of the agent (Gneezy & Epley 2014; Klein & Epley 2014). Together, this research suggests that judgments of proscriptive obligations incorporate factors related to causality, intentionality, and consequences more than do judgments of prescriptive obligations.

Importantly, research suggests that even young children have an awareness of this distinction between prescriptive and proscriptive violations. For example, children have been shown to exhibit better memory for negative, obligation-violating individuals than positive, obligation-fulfilling individuals (Barclay & Lalumiere 2006; Kinzler & Shutts 2008). Additionally, children as young as 14 months have greater difficulty following do's than don'ts (Kochanska et al. 2001). These findings suggest that these asymmetries do not simply reflect “second step” (sect. 2, para. 2) differences attributable to culture-specific learning, but rather emerge at an earlier ontogenetic stage.

These differences in judgments of prescriptive and proscriptive obligations suggest that there are important nuances to how developing humans learn about the moral obligations of their group and culture – and that the processes by which we learn about, represent, and evaluate prescriptive obligations (e.g., to obey our elders) may be different from those of proscriptive obligations (e.g., to not harm others).

Additionally, there is also evidence that the relevant weight accorded to each of these forms of obligation can differ as a function of the specific context in which they are embedded. For example, these two broad classes of moral obligation also seem to operate somewhat distinctly across intergroup boundaries. This appears to be an important caveat to Tomasello's claim that obligation “applies to, and only to, one's cultural compatriots” (sect. 2.2, para. 2). Whereas we agree that this claim generally holds true regarding *prescriptive* obligations (e.g., there is little expectation that one will help members of cultural outgroups), proscriptive obligations seem to be more common across intergroup boundaries (e.g., the obligation not to willingly harm members of cultural outgroups).

One apparent example of these asymmetries is the (rapidly growing) number of public apologies made by leaders of majority (racial/ethnic, religious, and/or cultural) groups for past injustices to minority groups – apologies which, almost without exception, center on violations of proscriptive obligations not to harm, rather than violations of prescriptive obligations (Blatz et al. 2009; Lazare 2004).

As Tomasello suggests, these cross-group apologies may indicate “an expanded sense of all of humanity as in one's in-group moral community” (sect. 2.2, para. 2). However, his account does not explain these apparent asymmetries in how prescriptive and proscriptive obligations operate across group boundaries.

We suggest that these prescriptive/proscriptive asymmetries may offer Tomasello a promising opportunity to refine his theory of moral obligation. These differences in how observers evaluate proscriptive and prescriptive obligations suggest that there may be different forms of obligation, each utilizing distinct psychological processes. At the very least, different moral obligations appear to engender different responses when they are upheld or violated. What psychologically distinguishes prescriptive obligations from proscriptive obligations? How could such differences emerge? For a complete account of the psychology of moral obligations, these questions ought to be addressed.

Obligations to whom, obligations to what? A philosophical perspective on the objects of our obligations

Kati Kish Bar-On 

The Cohn Institute for the History and Philosophy of Science and Ideas, Tel Aviv University, Tel Aviv Yafo, Israel, 6997801.

katikish@gmail.com <http://telaviv.academia.edu/KatiKish>

doi:10.1017/S0140525X19002395, e58

Abstract

Tomasello strives to understand the underlying psychology behind the human sense of obligation, but he only addresses a specific kind of obligation: to other human beings. We argue that in order to account for the psychological underpinning of human behavior, one should also consider people's sense of commitment to non-human entities, such as ideals, values, and moral principles.

Tomasello presents a compelling analysis of the motivating force behind humans' sense of obligation, focusing on the intersubjective structure of an obligation. However, obligations are not always immersed in agreements only *between* individuals. Occasionally, they are promises we make to *ourselves*, our commitments to a particular ideal of living or a moral value – entities that Tomasello does not discuss. These two different types of obligations can together explain individuals' behavior. If Bob is impudent to Charlie and Alice condemns Bob's behavior, Alice's criticism may derive from her heartfelt commitment to a certain principle of proper behavior to which she holds everyone accountable; she does not necessarily think that Bob has a sense of

obligation to Charlie or anyone else. She holds Bob accountable to the *standard itself*, not to other people. Alice *cares* about the ideal of proper behavior; therefore, it is *important* to her that people follow a certain behavioral standard when engaging with others.

The philosopher Harry Frankfurt argued that human behavior is shaped by the things we care about (Frankfurt 1982). Although ethical considerations regulate our relations with other people (using moral obligations), we often do not find the requirements of ethics to be the only things we care about. Even people with a strong sense of moral obligations to others can care equally about ideals such as being loyal to a family tradition or can devote themselves to ecological principles of fighting climate change (Frankfurt 2006).

Some people find moral obligations to be their most important obligations, above all other commitments. Others may choose to intentionally violate a moral obligation to another person not because there is a stronger moral obligation to which they are committed, but rather because they consider a certain value, ideal, or course of action to be more important *to them* than meeting the demands of moral obligations to others. Frankfurt's work draws our attention to the possibility that a unanimous hierarchical scale of obligations (where morality is superior to all other commitments) may not exist. An individual's sense of obligation to other people as well as to specific values or standards varies not only between people but also between situations: A person can choose to fulfill his obligation to another individual in one situation, but in a different circumstance he might choose to obey his commitment to a certain value or ideal over a moral obligation to a person.

Frankfurt has addressed this type of commitment as an integral part of one's inner identity, which serves as the reason and motivation for one's actions. He writes:

A person who cares about something is, as it were, invested in it. He identifies himself with what he cares about in the sense that he makes himself vulnerable to losses and susceptible to benefits depending upon whether what he cares about is diminished or enhanced. Thus he concerns himself with what concerns it, giving particular attention to such things and directing his behavior accordingly. (Frankfurt 1988, p. 83)

People whose behavior is governed by their commitments to the ideals they care about rather than by their moral obligations to others may be considered self-centered or judged as a person who lacks empathy. However, as Frankfurt has pointed out, the structure of people's personality is more complex than a one-dimensional scale of morality on one end and self-interest on the other. He argued that persons may feel committed to certain cultural or religious ideals that derive neither from moral nor egoistic considerations and may pursue those nonmoral ideals without considering their own personal goals (Frankfurt 2004).

Even though Frankfurt challenged the widely accepted authoritative nature of moral obligations, he did not claim that our relationships with other people or the moral requirements for maintaining those relationships are not important to us. He offered the possibility that other types of commitments may count as heavily or even more heavily with us, and he held that one should not assume that moral obligations always override them. For Frankfurt, what guides our behavior is not our moral or nonmoral obligations *per se*, but rather our attitude toward them: how much we care about these obligations and how important it is for us to carry them out.

According to Frankfurt, the fact that one cares about a certain thing is constituted by a complex set of cognitive dispositions, but

he did not account for the interpersonal factors that affect those dispositions. The role of an individual's social environment and group expectations in shaping human behavior is, to some extent, left out of Frankfurt's discussion. To fill the lacuna, let us return to Alice's disapproval of Bob's behavior: Even if Alice profoundly cares about a certain ideal of proper behavior, why should she hold Bob accountable to the same ideal? What urges her to condemn Bob's acts? Whereas Frankfurt did not engage in this type of question, Tomasello considers the societal aspects that are missing from Frankfurt's account. In Tomasello's view, the origins of Alice's expectations may derive from her view of herself, Bob, and Charlie, as members of the same cultural group. As such, all group members are obligated to conform to the group's ways or, if not, they must provide an explanation for their deviation. The object of Alice's sense of obligation may be a standard rather than a person, but in any case, her cultural identity and group affiliation will always play a significant role in the process of cultivating the things she cares about and her attitude toward transgression.

To conclude, human or non-human entities can be the objects of our feelings of obligation, and those, in turn, influence and shape our behavior. My intention in this commentary was to point toward an enhanced psychological account of humans' sense of obligation, including the approaches of both Frankfurt and Tomasello. Each theory portrays only part of the picture; together, they offer a more comprehensive account of individuals' feelings of obligation and the *objects* of those feelings.

Children's everyday moral conversation speaks to the emergence of obligation

Karen Bartsch 

Psychology Department, Wyoming University, Laramie, WY 82071-3415.
bartsch@uwyo.edu <http://www.uwyo.edu/psychology/faculty/bartsch.html>

doi:10.1017/S0140525X19002358, e59

Abstract

For Tomasello's proposed ontology of the human sense of moral obligation, observations of early moral language may provide useful evidence complementary to that afforded by experimental research. Extant reports of children's everyday moral talk reveal patterns of participation and content that accord with the proposal and hint at extensions addressing individual differences.

In his account of the ontogeny of our human sense of moral obligation, Tomasello pinpoints its origin in early joint intentional action and describes two developments in the preschool years: the first regarding interpersonal obligation between collaborative partners and the second regarding norm-based morality within a cultural group. Tomasello invokes extensive support from cleverly designed experimental studies of young children's behavior (e.g., Gräfenhain et al. 2009; Rekers et al. 2011), which have the dual virtues of bypassing young children's verbal limitations and facilitating cross-species comparisons. But Tomasello's

reliance on experimental findings raises the question of how well the account tallies with real-life observations.

I suggest that complementary evidence is available in observations of children's everyday interactions, especially conversations about morality. To be clear, Tomasello does incorporate research involving language, such as his own finding that children react to transgression with protest that reflects normativity (Rakoczy et al. 2008). But a different perspective is afforded by observations of children's real-life interactions. For instance, in exploring emerging moral sensitivity as revealed in two children's at-home talk with adults (Wright & Bartsch 2008), we noted patterns that accord intriguingly with Tomasello's account. We examined 1,333 conversations involving moral terms (e.g., "good," "wrong," "mean") sampled from "Abe" and "Sarah" as they aged from 2.5 to 5 years (Wright & Bartsch 2008), utilizing archived transcriptions from the CHILDES database (MacWhinney 2000). Surprisingly, such conversations were most frequent, relative to other talk, when children were 2.5–3, compared to 4–5, years of age. We speculated that the early intensity of such talk reflected the new autonomy, mobility, and active participation associated with the "terrible twos," noting also its coincidence with frequent talk about the desires of the children and adults, a recognized milestone in theory-of-mind development (e.g., Bartsch & Wellman 1995). Tomasello's proposal expands the explanation: Perhaps as 2- and 3-year-olds, Abe and Sarah were constructing with their collaborators (here, conversational partners) the hypothesized second-person morality that constitutes initial sensitivity to obligation. The ensuing decline in such conversation, relative to other talk, perhaps reflected the second phase when an acquired understanding of group norms reduced explicit reference to moral rules, now assumed to be universally recognized (e.g., Mamman et al. 2018).



Another characteristic of Abe's and Sarah's early moral conversations that tallies with Tomasello's account is the active role played by children in such discussions (Wright & Bartsch 2008). Far from being passive recipients, both children surpassed their adult interlocutors in initiating moral conversations and did so as though they were exploring the rules and roles of those involved. Even as 2- and 3-year-olds, Abe and Sarah used moral terms frequently to give and request reasons (e.g., "because he's nice to nice people," Wright & Bartsch 2008, p. 61) more than for other purposes, such as to communicate feelings or to disapprove of actions, a functional disparity that increased with age. Abe at age 3, for instance, said, "You could have put it on the floor for me. I asked you so you should have done it" (Wright & Bartsch 2008, p. 61).

Young children are not only surprisingly active in everyday moral conversation, as fits with Tomasello's proposition that they are constructing a sense of moral obligation, but their favored topics also accord with the account. Abe's and Sarah's conversations focused on mostly internal as opposed to external motivations, specifically on dispositions and behavior, primarily those of other people, and especially on bad behavior (Wright & Bartsch 2008). Other researchers have similarly reported young children's fascination with others' dispositions and transgressions, observing, for instance, that children focus more on a sibling's, than on their own, transgression (e.g., Dunn & Munn 1986; Ross & den Bak-Lammers 1998). Dunn (1987), who conducted comprehensive and systematic longitudinal studies of conversation among family trios consisting of a parent, a toddler, and an older sibling, reported that by age 2, children communicated openly about both obligation and blame with respect to family social rules and others' feelings. These observations accord with

Tomasello's characterization of emergent obligation, adding perspective and detail from the child's actual context and voice.

Observations of children's everyday conversations may also provide clues about individual differences in a developing sense of moral obligation. For example, although both Abe and Sarah initiated most moral discussions, Abe was active in 80% of conversations compared to Sarah at 60%, rates maintained throughout the several years (Wright & Bartsch 2008). Conversation content also differed: For Abe, feelings and others' welfare was the modal topic, characterizing 25% of moral conversations, while for Sarah issues of obedience and punishment were modal at 50% (Wright & Bartsch 2008). These differences hint at divergent paths in moral focus and maybe in a sense of obligation. Evidence across studies suggests that contexts may figure importantly in individual trajectories. For Abe and Sarah, most moral conversations concerned immediate interpersonal interests rather than impersonal and abstract topics such as social rules (Wright & Bartsch 2008), consistent with Tomasello's characterization of the earliest sensibility regarding obligation. Dunn (1987), however, reported an increase over toddlerhood in the frequency with which mothers and older siblings spoke to toddlers about social rules and broken or flawed objects. It may matter whether early interactions involve only the child and parent, as in the observations of Abe and Sarah, as opposed to involving the child, a parent, and an older sibling, as in Dunn's (1987) research. Such comparisons highlight the role of context, a factor acknowledged by Tomasello in his discussion of cross-cultural variation. Intensive observations of conversations suggest that, even across families, who is talking and what is talked about matter. For instance, pretend play has been observed to be a common context for moral discussion, although an activity that differs across families (e.g., Dunn 1987; Wright & Bartsch 2008). These examples suggest that extant and future observational studies can contribute not only specificity regarding the nature of critical early interactions but also to an understanding of emergent individual differences as they relate to a developing sense of moral obligation.

The role of affect in feelings of obligation

Stefen Beeler-Duden, Meltem Yucel 
and Amrisha Vaish 

Psychology Department, University of Virginia, Charlottesville, VA 22903.
sjb3px@virginia.edu
https://pages.shanti.virginia.edu/Social_Development_Lab_3/
nmy2bg@virginia.edu www.meltemyucel.com
 (corresponding author)
av8u@virginia.edu
https://pages.shanti.virginia.edu/Social_Development_Lab_3/

doi:10.1017/S0140525X19002449, e60

Abstract

Tomasello offers a compelling account of the emergence of humans' sense of obligation. We suggest that more needs to be said about the role of affect in the creation of obligations. We also argue that positive emotions such as gratitude evolved to

encourage individuals to fulfill cooperative obligations without the negative quality that Tomasello proposes is inherent in obligations.

Tomasello offers a compelling account of the evolutionary and ontogenetic emergence of the uniquely human sense of obligation. The target article alludes to some affective experiences in these “feelings of obligation” but does not directly address the role of emotions. Further, to the extent that Tomasello does consider the role of affect, he describes obligations as having a fundamentally coercive (negative) quality. In this commentary, we propose (1) a more central role of affect in the development of a sense of obligation, and (2) that positive emotions such as gratitude represent an evolutionary solution to the fundamentally negative quality of obligations.

First, Tomasello emphasizes the development of obligation through the me–you equivalence established first through collaborative activities and then in cultural groups. This account is certainly persuasive, but we think that it is not sufficient. Research shows that from the very start, children’s collaborative interactions are also heavily affect-laden, and this affect is vital for establishing their sense of obligation. According to the affect-as-information theory (Storbeck & Clore 2008), affective signals provide children with cues about what is safe and important in the environment. For example, visual cliff experiments show that even 1-year-olds take their parent’s nonverbal signals into account and use their emotional expressions to make decisions about their future behaviors (e.g., whether or not to crawl across the cliff; Sorce et al. 1985).

More pertinently, caregivers use affect to convey vital information about obligations to infants and young children. For instance, parents use distinct affect when communicating with infants about norm violations compared to pragmatic violations (Dahl & Tran 2016). Parents even use distinct affect when communicating about moral versus other kinds of norm violations: They use angry vocalizations when their infant violates a moral norm (e.g., hits another person) but use fearful vocalizations when their infant violates a prudential norm (e.g., harming themselves; Dahl et al. 2014). This affective information guides children’s attention to important distinctions between various types of social norms and the obligations that accompany those norms (Arsenio & Ford 1985). This in turn helps children to successfully navigate their obligations and avoid breaking commitments. Affect-laden testimonies, for example, moralize novel rule violations and elicit moral judgments from children (Rottman & Kelemen 2012; Rottman et al. 2017). Similarly, among young children as well as adults, witnessing violations of moral norms induces higher physiological arousal than witnessing violations of conventional norms (Yucel et al., *in press*). As mentioned by Tomasello, as children begin to differentiate these social norms from one another, they begin to understand *who* is obligated to adhere to particular norms and *when* an individual is obligated to do so. Affect-laden communication and children’s own affective experiences thus aid children in developing a sense of obligation and help children work out the nuances of the particular system of obligation in which they must function.

Second, Tomasello argues that obligation makes us do things we do not *want* to do but feel we *must* do. He emphasizes the negative emotions associated with obligation (guilt, blame, and resentment) and argues that positive emotions (such as gratitude) cannot account for a sense of obligation. We propose, however, that positive emotions such as gratitude, which motivate us to reciprocate favors to our benefactors, may have evolved precisely to counter

the fundamentally negative quality of obligation: to turn actions such as reciprocity into something we *want* to do, not only because it is right but also because it feels good. On its face, this goes against Tomasello’s argument that obligation cannot involve things we *want* to do. Yet perhaps the ultimate and proximate stories diverge on this point: In our evolutionary history, we needed psychological mechanisms to make us put aside our selfish interests and invest in our cooperative, interdependent relationships (as Tomasello himself has elegantly argued: e.g., Tomasello 2016a). Emotions are just such psychological mechanisms (Krebs 2008; Nesse & Ellsworth 2009). It is highly likely, then, that some emotions have evolved to help us fulfill our obligations to others. Some of these emotions are negative (resentment, guilt), but some are also positive (gratitude). At the proximate (ontogenetic) level, these positive emotions do not seem to carry the fundamentally coercive quality of obligations. However, we submit that positive emotions are precisely the evolutionary trick that allows us to happily carry out many of our social obligations. Positive emotions are part of the toolbox that allows us to meet our obligations, and to do so without resenting them.

Taken together, we seek to highlight the important functions that affect serves in the emergence of the human sense of obligation. We presented evidence from the affect literature and argued that affect plays an important role in how obligations are socialized and maintained during development. Moreover, we suggested that positive emotions may have evolved to counteract the negativity associated with the sense of obligation one feels to one’s collaborative partners and cultural group. Thus, we argue that a full account of the evolutionary and ontogenetic emergence of humans’ sense of obligation must include the vital role of affect in the creation, enforcement, and maintenance of these obligations.

NOTE

Beeler-Duden and Yucel contributed equally.

The sense of obligation is culturally modulated

Andrea Bender 

Department of Psychosocial Science & SFF Centre for Early Sapiens Behaviour, University of Bergen, N-5020 Bergen, Norway.

andrea.bender@uib.no <https://www.uib.no/en/persons/Andrea.Bender>

doi:10.1017/S0140525X19002371, e61

Abstract

Tomasello argues in the target article that, in generalizing the concrete obligations originating from interdependent collaboration to one’s entire cultural group, humans become “ultra-cooperators.” But are all human populations cooperative in similar ways? Based on cross-cultural studies and my own fieldwork in Polynesia, I argue that cooperation varies along several dimensions, and that the underlying sense of obligation is culturally modulated.

Tomasello presents a compelling account of how the human sense of obligation arises from shared intentionality and interdependent collaboration. Developing this concrete, collaboration-dependent

obligation into a generalized support of one's cultural group ("obligation period"), he argues, turns humans into a highly cooperative species. Although cultural variability in this second step is acknowledged, its profound impact appears to be underestimated.

Obligation is one of the four deontic modalities, along with ban, permission, and release from obligation (often expressed by way of modal verbs: "must," "must not," "may," and "need not"). These modalities are systematically interrelated in a "square of opposition," constituting the structure of the deontic system (Beller 2012). People are highly competent in drawing inferences based on these interrelations, which must remain in place to preserve the system's consistency and operability, such as that obligation entails permission, or that something forbidden cannot simultaneously be prescribed (Beller 2008; 2010). Due to these logical exigencies, deontic systems are invariable on an abstract level. What does vary across cultures is their content: the concrete things that are forbidden, prescribed, or permitted.

Promises are a textbook example at the intersection of the human sense of obligation and cooperation. Promises create a social contract between the two parties involved, cutting across five psychologically relevant levels: (1) As an attempt to induce a behavioral change in the addressee, promises reflect a specific *motivation*, the wish to initiate cooperation, and (2) they are formulated in a specific *linguistic* manner. (3) Promises entail *deontic* implications, generating an obligation for the speaker, contingent on a cooperative act of the addressee in the case of a conditional promise. For promises to work, it is essential that the ensuing obligation is perceived as binding, as the addressee has only the speaker's word to go on. (4) On the *behavioral* level, the involved parties decide whether or not to cooperate, with either outcome then (5) eliciting distinct affective responses (Beller et al. 2005).

Some core characteristics are implicated in the very concept of promises and hence should be universal. Their structure is determined by actions, goals, and expectations; they entail asymmetries in terms of ensuing obligations and the temporal sequence of decisions to be made; and their binding force should be independent of how close the relationship between the parties is. Still, a study with participants from Germany, China, and Tonga revealed cultural variation on almost all levels (Beller et al. 2009). For instance, the linguistic formulation considered most appropriate varied with culture-specific communication styles, and affective responses to a kept or broken promise depended on display rules for emotions. Crucially, even the deontic implications were modulated by cultural conventions. In Tonga, a Polynesian archipelago in the Western Pacific, the very utterance of a conditional promise was found to exert pressure on the addressee – that is, an obligation to be responsive to the promise – hence coercing an act of cooperation not necessarily volunteered otherwise. Interestingly, this was accompanied by a reduced obligation for the speaker to actually keep the promise. It is as if conditional promises were simply understood as a request for a favor ("I want you to do this for me"). Since the addressee is obligated to fulfill this wish, the promised return-favor may be discounted.


In other words, obligations originate not only from the deontic structure of promises, but also load on external rules arising from the cultural value system. In Tonga, this includes the prime virtue of reciprocal help (*fetokoni'aki*), which is grounded in respect and concern for others, is reflected in and reinforced by a strong orientation towards the group, and obligates people to provide mutual support and to share food and other resources (Bender 2007). Although this solidarity network incurs economic costs for the individual, it also generates an informal insurance system

for all (Bender et al. 2002). Similar patterns have been observed more widely both within and across the small island communities of the Pacific (e.g., Hage & Harary 1996; Petersen 2000). And yet, this is only *one* possible outcome of the cultural patterning of cooperation and obligation.

Two large-scale studies by Henrich et al. (2005; 2010) demonstrated astounding variability in what different cultural groups hold to be cooperative and fair. Using economic games to assess the willingness to share with anonymous members of one's group, participants were found to differ substantially in the share they were willing to offer, in the size of offers they were willing to accept, or in the expenses they were willing to invest in the punishment of those who violated norms of cooperation. In some groups, much lower shares were both offered and accepted than is typical for WEIRD (Western, educated, industrialized, rich, and democratic) populations, whereas other groups rejected even highly generous offers. Respective tendencies co-varied systematically with characteristics of the group (such as community size) and its cultural institutions, including economic organization, the structuring of social interactions, or the type of religion. Especially in small groups – the default state in large parts of human history – the concern with fairness appeared to be lower in situations lacking relationship information, suggesting that "modern prosociality is not solely the product of an innate psychology, but also reflects norms and institutions that have emerged over the course of human history" (Henrich et al. 2010, p. 1480).

In a nutshell, obligation period is modulated by cultural conventions and practices along several dimensions: by shaping the conceptualization of self-declared commitments, the kinds of obligations perceived as arising from them, and even the emotional responses linked to them. This impacts on the extent to which the sense of obligation is generalized to in-group strangers and on the scope of cooperation deemed appropriate, rendering culture one of the most powerful forces in forming human cognition (Bender & Beller 2019) and social behavior (Henrich et al. 2005). Considering this profound impact enables the surprising insight that culture itself evolved to stabilize and generalize the predisposition for cooperative relationships in our species by fostering and molding the sense of obligation, even if in culture-specific ways.

Tomasello on "we" and the sense of obligation

Michael E. Bratman 

Philosophy Department, Stanford University, Stanford, CA 94305.

bratman@stanford.edu

<https://philosophy.stanford.edu/people/michael-e-bratman>

doi:10.1017/S0140525X19002383, e62

Abstract

Tomasello explores four interrelated phenomena: (1) joint intentional collaboration; (2) joint commitment; (3) "self-regulative pressure from 'we'"; and (4) the sense of interpersonal obligation. He argues that the version of (1) that involves (2) is the "source" of (3) and so the source of (4). I note an issue that arises once we distinguish two versions of (3).

Tomasello in the target article explores four interrelated phenomena:

1. joint intentional collaboration
2. joint commitment
3. “self-regulative pressure from ‘we’” (sect. 2.1.3, para. 3)
4. sense of interpersonal obligation

His “shared intentionality hypothesis” aims to “undergird” (4) “with the psychological foundation of shared intentionality.”

Tomasello sometimes suggests that (4) “arises” from (1). However, he notes that a child can be “just playing with the adult spontaneously, in the absence of a joint commitment” (sect. 2.1.2.1, para. 5) But “playing with [an] adult” – even if spontaneous – is not merely playing alongside an adult. If the child is playing *with* the adult then there is a kind of joint intentional collaboration, and both the child and the adult may well think: “we are playing together.” There is a “we.” But, Tomasello notes, there is not yet “joint commitment”; and, as I understand Tomasello, there is not yet (4).

This complexity arises within the “interim hypothesis ... that participation in joint intentional collaboration, especially as initiated by a joint commitment, is the earliest source of children’s ... feeling of obligation.” (sect. 2.1.3, para. 5) This seems to say that (1) is the “source” of (4), and (2) is just a special case of (1). But this does not fit with the recognition of the possibility of (1) in the absence of (2) and (4). So I think the best reading of Tomasello’s hypothesis is that the “source” of the sense of obligation is (2), not simply (1).

What is joint commitment? Elsewhere Tomasello has said that joint commitment consists in “collaborative activity ... initiated by an overt and explicit act of cooperative communication” (Engelmann & Tomasello 2018, p. 440). But, first, I take it that the relevant overt acts will be, more specifically, forms of assurance. And, second, if it is a *joint* commitment, it is necessary that all parties assure. So, I take the view to be that jointly intentional collaborative activity that is initiated by such a web of assurances is the source of the sense of obligation.

Further, Tomasello’s proposal is that (2) is the source of (4) because (2) is the source of (3), and (3) in turn ensures (4). When he writes that “the sense of obligation ... is nothing more or less ... than the motivational force that accompanies the creation of a joint or collective agent ‘we’...” (sect. 3, para. 6) he is seeing “a joint or collective agent ‘we’” as involving “the sense of obligation,” and, I take it, supposing that this “we” arises from joint commitment.

We need to distinguish two versions of (3). Suppose the child plays spontaneously with the adult in the absence of joint commitment. Why is this a case of joint intentional collaboration, not just a case of each merely playing alongside the other? In (Bratman 2014), I propose (roughly) that in shared intentional activity each publicly and interdependently intends the joint activity and that it go by way of relevant intentions of each, mutual responsiveness, and meshing subplans. This structure of intentions of each constitutes their joint/shared intention so to act. Applied to our present case, this explains a kind of “self-regulative pressure from” the “we,” since both the child and the adult can reason instrumentally concerning what is needed for their jointly intended playing together. For example, the child can reason: “As one of the interdependent elements of our shared intention, I intend that we play together; for us to play together, I need to give him the ball; so I will.” In this sense, (1) ensures a version of (3). But this *weak* version of (3) simply involves instrumental reasoning concerning the jointly intended end of playing together; it need not involve a sense


of obligation to the other. The child can reason instrumentally concerning this jointly intended end of playing together without thinking that she owes it to the adult to give him the ball.

In contrast, in a *strong* version of (3) the self-regulation involves a sense of obligation to the other. And Tomasello’s thought is that joint commitment ensures this strong version of (3), and so thereby (4).

Given that joint commitment involves joint intentional collaboration, joint commitment ensures the weak version of (3). But why does adding the cited web of assurances ensure the *strong* version of (3), one that is not ensured simply by joint intentional collaboration?

Perhaps Tomasello’s view is simply that it is a feature of our human psychology that joint intentional collaboration that is initiated by such a web of assurances does in fact induce a sense of obligation; so, when such a web precedes and initiates joint intentional collaboration, it transforms the weak “we” of joint intentional collaboration into the strong, sense-of-obligation-loaded “we.” This would change the order of explanation and appeal directly to a connection between (2) and (4), and then thereby explain the connection between (2) and the strong version of (3). More importantly, it raises a further question. Does the structure of joint intentional collaboration that is initiated by such a web of assurances induce a sense of obligation *by virtue of the role of shared intentionality in that structure*? Well, it does not do this solely by virtue of the role of intentional joint collaboration: that only ensures a weak “we.” Instead, it does this in part by virtue of the web of assurances. It is that web of assurances that is doing the work in going from a weak to a strong “we.” But such assurances need not themselves be jointly intentional activities; and they can occur in the absence of relevant jointly intentional activities (think of a case of insincere assurances). So, what is undergirding (4) involves a crucial element that is separable from intentional collaborative activity. Is this compatible with the idea underlying the shared intentionality hypothesis?

The joy of obligation: Human cultural worldviews can enhance the rewards of meeting obligations

Emma E. Buchtel 

The Education University of Hong Kong, Hong Kong SAR, China.

buchtel@eduhk.hk

<https://repository.eduhk.hk/en/persons/emma-ellen-kathrina-buchtel>

doi:10.1017/S0140525X19002607, e63

Abstract

Is it particularly human to feel coerced into fulfilling moral obligations, or is it particularly human to enjoy them? I argue for the importance of taking into account how culture promotes prosocial behavior, discussing how Confucian heritage culture enhances the satisfaction of meeting one’s obligations.

Why do humans fulfill moral obligations? Tomasello persuasively argues that humans’ unusual ability to form a sense of

interdependence with others, therefore identifying with the group and its norms, allows children to develop a sense of obligation to follow cooperative social norms. Tomasello defines these obligations primarily in terms of their “coercive” power, which force us to subordinate selfish desires to the demands of the group, supported by social pressure and self-punishing guilt. Curiously (given Tomasello’s groundbreaking work, e.g., Tomasello 2016b; Tomasello et al. 1993), culture plays a small role in this article, beyond providing a group identity and, presumably, some of the content of the norms that become obligatory. In this commentary, I argue for the importance of studying the positive experience of fulfilling obligations, and the role of culture in enhancing it, to understanding why humans adopt moral behavior.

In our ratcheted up societies (Tomasello et al. 1993), culture is a key element in creating prosocial norms and making it likely that we will fulfill them. Culture determines the content of moral obligations (e.g., Miller & Bersoff 1992; Miller et al. 1990); consists of complex systems of beliefs and practices that uphold and spread cooperation (e.g., Norenzayan et al. 2016); changes the experience of fulfilling one’s obligations (Buchtel et al. 2018); and perhaps even changes whether or not absolute impermissibility is a central aspect of morality (Buchtel et al. 2015). Although cultural influences necessarily build on children’s initial propensities to learn social norms, the content of our particular normative cultures has enormous influence over whether and why we follow those norms. To understand why humans fulfill obligations and responsibilities to one another, we must study how specific cultures sustain those obligations.

For example, even moral impermissibility may be a more central feature of Western concepts of morality than Chinese ones, suggesting that the “coercive” nature of moral obligation might not be the key universal human feature to be explained (Anscombe 1958). In our study of Chinese lay prototypes of immoral behaviors, we found a puzzling disconnect between what was acknowledged to be most harmful (e.g., killing people) and what was called immoral; the Chinese word for immoral was most applicable to behaviors that were particularly uncultured or uncivilized, such as spitting in the street or disrespecting parents, and less applicable to criminally harmful behaviors such as stealing or killing (Buchtel et al. 2015). This suggests that the categories of antisocial behaviors are differently organized in Confucian heritage cultures than they are in Western heritage cultures, where instead behaviors have been divided into those that are impermissible (very harmful) versus not (Buchtel et al. 2015). Historically, morality in China has been based on the teaching of Confucian virtues such as benevolence and propriety, while criminals – presumably beyond the motivational reach of virtue – were dealt with through the criminal law system (Bakken 2000). Rather than emphasizing moral absolutism, Confucianism is instead described as a form of virtue ethics (Angle & Slote 2013), according to which morality guides character development so that the virtuous can respond appropriately to different situations.

Similarly, conflict between the self and group needs is a classically Western concern, but not a central theme in Confucianism (Buchtel et al. 2018). Instead, the obligations, duties, and responsibilities associated with one’s social role are perceived as an opportunity to realize and ennoble the self (Shun & Wong 2004). Consider the experience of not breaking one’s obligations – but meeting them. What allows us to feel *joy* by being a moral person? We’ve found that adult participants who are more influenced by Confucian heritage cultures (vs. Western heritage cultures) are more likely to associate an obligation to help others with positive

emotions and also with a sense of personal agency (Buchtel et al. 2018; see also Miller et al. 2011; Tripathi et al. 2018). It is likely that Confucian heritage moral cultures strengthen the experience of intra- and interpersonal rewards for responding to what others think one ought to do. This current cultural difference has strong parallels with philosophical differences between these two heritage cultures: A Confucian admiration of those who fulfill obligations sincerely (detailed in contemporary philosophers’ discussion of Confucian role ethics, e.g., Ames & Rosemont 2014), versus a Kantian concern about avoiding the coercive nature of such obligations. However, along with the cultural difference, we also found surprisingly strong evidence that Euro-Canadians also experienced both positive emotion and a sense of personal agency when they felt more obligation to help, suggesting that fulfilling duties may be universally rewarding.

The role of cultural evolution in the creation of cooperative humans and societies is a new and burgeoning research area (e.g., Norenzayan et al. 2016). Like the focus on guilt and punishment in the target article, much of the focus in cultural evolution has been on how the sanctioning of norm violators motivates cooperation and obedience to social norms. Yet, the above research suggests that following norms generates a sense of agency and positive emotions (Buchtel et al. 2018). Similarly, the pursuit of eudaemonic activities such as prosocial behavior has been linked to reward responses and increased well-being (e.g., Steger et al. 2008; Telzer et al. 2014). Although moral behavior is undoubtedly encouraged by punishment for norm violations, we also need to know more about how cultures enhance the rewards experienced when we meet prosocial norms.

Culture is a crucial feature of what enables humans to willingly follow social norms. The psychological or invisible cultural environment – values, philosophies, religion – has, because of humans’ unique ability to create and learn culture, become a vital tool in our historically endowed arsenal of survival mechanisms. Assuming that the human ability to create and absorb culture is a key feature of human evolutionary success, then knowledge of how different cultures encourage prosocial behavior is essential to understanding how human children develop a propensity towards goodness.

Tomasello’s tin man of moral obligation needs a heart

Jeremy I. M. Carpendale^a  and Charlie Lewis^b 

^aPsychology Department, Simon Fraser University, Burnaby, BC, V5A 1S6 Canada and ^bPsychology Department, Lancaster University, Fylde College, Lancaster, LA1 4YW United Kingdom.

jcarpend@sfu.ca

<https://www.sfu.ca/psychology/about/people/profiles/jcarpend.html>

c.lewis@lancaster.ac.uk

<https://www.lancaster.ac.uk/people-profiles/charlie-lewis>

doi:10.1017/S0140525X19002462, e64

Abstract

In place of Tomasello’s explanation for the source of moral obligation, we suggest that it develops from the concern for others already implicit in the human developmental system. Mutual

affection and caring make the development of communication and thinking possible. Humans develop as persons within such relationships and this develops into respect and moral obligation.

Tomasello has a great talent for selecting crucially important topics and carefully exploring them with brilliant experimental work in collaboration with his group of talented colleagues. He proposes collaborative interaction as a social interactive medium in which obligation develops because others are recognized as equals, and so equally deserving, and this leads to mutual respect. The second step is to move beyond responsibility to one's particular collaborative partners and to extend obligation to one's social group, as internalized pressure from a collective "we."

We support Tomasello's goal of explaining the developmental and evolutionary origins of the human sense of moral obligation within an interactive framework. Although his approach is social in the sense of focusing on individuals who interact, this already presupposes rather than explains the development of those individuals. A more thoroughly social approach from a process-relational perspective proposes that persons develop within interaction (Carpendale et al. 2013; Carpendale & Lewis 2004; 2006; 2015a; 2015b).

Instead of attempting to explain the emergence of a sense of obligation later in development, concern for others is already implicit in the typical human developmental system with its origins in mutual affection and caring. This is necessary to start the developmental process early in the infant's life, as seen in frequent dyadic exchanges and the emergence of intense attachments by 6 months of age. Humans develop as persons within relationships of mutual affection, and this transforms to mutual respect in the sense of treating those others with care. Morality is based on mutual affection (Piaget 1932/1965). Treating others as someone, not something (Spaemann 2006), is already embedded in the structure of communicative interaction that infants experience in development – the seeds for mutual respect (Carpendale 2018; Habermas 1990).

It is because human infants are born relatively helpless that there is so much potential for their development. A strong positive emotional connection is a foundation for the human developmental system in which infants develop as persons, and learn to communicate, which makes thinking possible. Within their intense social emotional relationships, infants first learn to communicate through coordinating their interaction with others. Such communication is the basis of language and results in the development of human thinking. All of this requires the social-emotional cradle in which humans develop. Moral obligation is not the result of realizing that others are equals and therefore should be treated with respect. Instead, it is a natural outcome of mutual affection and understanding. This caring and concern for others is what later develops into a sense of obligation, first to those close and, later, extended to others.

There is a missing link in Tomasello's explanation. He suggests that, in collaborating with others, children see them as equal and so equally deserving. But this does not explain why they feel obliged to them. We don't add moral obligation later in development – it is already implicit in the human developmental system as a result of the nature of early relationships. Infants are treated as persons, as participants in interaction. It is the product of treating others as persons and responding to them in everyday activity. Our interpretation of the research showing that 3-year-olds feel obligation to those they interact

with collaboratively is that children have experienced obligation within the communicative interaction they grow up in. Conversation is a special case of collaborative interaction in general, as Grice (1975) suggested, which is extended to the research settings involving collaboration. In conversation, failing to respond to others is morally accountable (Turnbull 2003). Some children may occasionally be prompted by caregivers to respond if they fail to do so on their own, but we suggest that this is unlikely to be the primary way that they learn about obligation in conversation. Instead children pick up on others' expectations of a response within many daily interactions. Gradually, children begin to recognize the consequences for others' feelings of not responding to them.

Tomasello's second step involves conformity, which he sees as a requirement for membership in a cultural group. He suggests that individuals feel social pressure as obligation, but we are not convinced that this can be a complete explanation for moral obligation. People sometimes feel a moral obligation to disobey the culture's (and our parents') ways of doing things if they are believed to be wrong and need to be changed. Tomasello does not explain this. For him, children buy into cultural norms without evaluating them and uphold such norms because of what others will think about them. But this is just conformity. It does not get us to right and wrong. Individuals may disagree with and oppose such norms leading to change. Although conformity is a dimension of human social life, Tomasello's approach is incomplete and leads to moral relativism. It cannot explain how the Greta Thunbergs of every generation challenge the status quo so early in their development.

We suggest that the second step Tomasello proposes beyond individuals' obligation to their collaborative partners is not just one step. Instead, it is a gradual process of including more perspectives on the moral issue in question, beginning with those in close relationships and extending to one's cultural group. But this can be further extended to other groups and to other animals.

Tomasello proposes taking a social approach to explaining the source of moral obligation, but there are three problems with his argument. First, there is still an implicit separation of emotions and cognition. Second, the process he describes begins with individuals who then cooperate and so feel social pressure as obligation, but we don't always feel obligation as onerous. Third, there is insufficient explanation of how, or why, obligation emerges so late in development. Mutual enjoyment in interaction makes human communication possible and then language and forms of thinking based on language. Caring and mutual affection are embedded in the structure of the human developmental system. These strong emotional bonds are the seed for mutual respect, which is already there in communication, and develops increasingly into moral obligation.

Intuitive theories inform children's beliefs about intergroup obligation

Lisa Chalik 

Yeshiva University, Stern College for Women, Department of Psychology, New York, NY 10016.

lisa.chalik@yu.edu <http://www.developingmindslab.com>

doi:10.1017/S0140525X19002516, e65

Abstract

In addition to emerging from children's direct experiences with collaborative partners and groups, children's beliefs about obligation also arise from a process of intuitive theory-building in early childhood. On this account, it is possible for at least some of children's beliefs to emerge in the absence of specific experiences where obligations are held among fellow members of a group "we."

In the target article, Tomasello argues that children initially view obligations as holding among collaborative partners, then they extend this knowledge to collaborative and cultural groups during the preschool years. Overall, this is a convincing argument that makes sense given the extensive literature on when and how children develop an understanding of norms.

Yet, the argument that children's sense of obligation is dependent on their developing concept of "we" implies that it is children's own direct experience in social groups that drives all of their moral understanding. Clearly, children's firsthand experiences with the social groups in their environment influence a range of social category-based processes, including obligation. Yet, this argument overlooks the possibility that at least some of children's beliefs about groups exist as part of a more abstract framework of how social groups function, beyond children's own experiences. For example, Tomasello reviews a great deal of work showing that preschool-aged children show preferential treatment for in-group over out-group members across a range of experimental paradigms (Fehr et al. 2008; Misch et al. 2014; Over 2018; Over et al. 2016). These findings are not trivial and provide strong support for the suggestion that children's beliefs about obligation are embedded within their understanding of social groups. But they do not disentangle children's own affective biases toward the groups they have encountered in the world from their broader understanding of how social groups function in general.

Across development, children build intuitive theories – abstract, domain-specific, causal-explanatory frameworks – of how the world works (Wellman & Gelman 1992). In the social world, children's intuitive theories involve both (a) what social group members are like, and (b) how social group members act toward one another (Rhodes 2013). Importantly, these theories are *normative* in nature – they involve not just descriptions of how group members act, but also prescriptions of how group members are *supposed to act* (Haward et al. 2018; Prasada & Dillingham 2009). In other words, children's intuitive theories establish the obligations by which group members are believed to be bound. To that end, by the preschool years, children view moral obligation as shaped by group membership: They think that people are more likely to harm out-group members than in-group members (Chalik & Rhodes 2014; 2018; Chalik et al. 2014; Rhodes 2012), and that people will protect in-group members over out-group members from harm (Chalik & Rhodes 2018). Furthermore, as soon as children receive input to suggest that a given behavior is morally relevant, they spontaneously expect that behavior to play out according to group boundaries, even if they have had no experience with the particular behavior or groups in question (Chalik & Dunham 2020). Importantly, in contrast to most of the in-group preferences documented in the target article, these findings have all come from tests of children's *third-person* reasoning. In these third-party paradigms, children's judgments cannot be based on any personal biases that they hold


in favor of their own social groups, since children are not members of the groups they are reasoning about. Thus, these findings do not seem to rely on a sense of "we" – rather, they depend on a sense of "they." This sense is abstract in nature, involving children's beliefs about how groups are supposed to function in the world, rather than whatever children have actually experienced with the specific groups in their environment.

This possibility need not be in direct opposition to the one presented by Tomasello. It is certainly possible that children could build their understanding of obligation from *both* their experiences in collaborative partnerships and their intuitive theories of how social groups function. Yet, if both of these proposals are true, then an open question remains: When and how do children incorporate their specific experiences into their more abstract expectations of the world? The relation between children's personal biases, as documented in first-person work, and their abstract expectations of the world, as documented in third-person work, remains largely unexplored, and will be an important area for future research.

An additional issue raised by the intuitive theories account regards the time-course of children's developing understanding of obligation. Tomasello reviews a great deal of work suggesting that it is only after age 3 that children have a true sense of "we." Yet, the strong conclusion that children do not incorporate social groups into their beliefs about moral obligation until this age is premature. If children hold an intuitive theory by which social groups mark moral obligation, and notions of obligation are thus embedded in representations of social groups, then children may begin to develop these beliefs as soon as they start to recognize that social distinctions exist in the world. A great deal of work now shows that infants are sensitive to social groupings within the first year of life (Bar-Haim et al. 2006; Kinzler et al. 2007; Quinn et al. 2002), and that toddlers can represent novel social groups, given the right input (Diesendruck & Deblinger-Tangi 2014; Rhodes et al. 2018). Furthermore, infants and toddlers do appear to have different expectations about how people will interact with one another, depending on group membership (Bian et al. 2018; Jin & Baillargeon 2017; Ting et al. 2019). This evidence is somewhat limited, and there is undoubtedly much about children's understanding of obligation that continues to develop beyond age 3; still, to some extent, it seems likely that children begin to hold these concepts within the first three years of life.

Thus, the argument that children's beliefs about obligation arise from their experiences in collaborative partnerships and groups is a strong one. Yet, it is incomplete without also considering the intuitive theories of social groups that children hold regardless of their direct experiences, as well as the social group-based judgments made by infants and toddlers.

Who are "we" and why are we cooperating? Insights from social psychology

Margaret S. Clark, Brian D. Earp 
and Molly J. Crockett 

Department of Psychology, Yale University, New Haven, CT 06520-8306.
margaret.clark@yale.edu

<https://clarkrelationshiplab.yale.edu/people/margaret-clark-brian.earp@yale.edu> <https://oxford.academia.edu/BrianDEarp>
molly.crockett@yale.edu <http://www.crockettlab.org>

doi:10.1017/S0140525X19002528, e66

Abstract

Tomasello argues in the target article that a sense of moral obligation emerges from the creation of a collaborative “we” motivating us to fulfill our cooperative duties. We suggest that “we” takes many forms, entailing different obligations, depending on the type (and underlying functions) of the relationship(s) in question. We sketch a framework of such types, functions, and obligations to guide future research in our commentary.

Drawing on developmental, comparative, and philosophical perspectives, Tomasello gives an account of moral obligation, which he argues has received “almost total neglect” from psychologists. We agree with Tomasello that a sense of obligation arises from cooperative agreements between humans, often functioning as a powerful “stick” to prevent behavior that would lead to one’s own guilt or partner resentment. In fact, psychologists focusing on dyadic and group behavior have long discussed such means of prioritizing “we” over “me” in human societies: Kelley’s (1982) discussion of transformations-of-motivation from individual to shared concerns is a key example. The importance of cooperative norms and the risk of resentment and guilt implied by breaking them also have been emphasized by researchers such as Walster et al. (1978) who traced shifts from “me” to “we” and identified resultant obligations as part of their equity theory. Others have noted that moral judgments of the self and other are caused by meeting or failing to meet cooperative expectations, the nature of which often vary by relational context (e.g., Bloom 2011; Clark & Boothby 2013; Rai & Fiske 2011).

Different relationship types have been distinguished in terms of the adaptive functions they serve (Bugental 2000; Clark & Mills 1979, 2012; Fiske 1992); the demands people place on cooperative partners depend on the origins and nature of their interdependence (Bugental 2000; Clark & Mills 1979). We suggest that taking into account who “we” are (in Tomasello’s sense), as well as “why” and “how” we are cooperating based on the functions normatively served within our relationship(s), will be crucial for making progress in the social psychology of moral obligation.

Consider Wendy, who could easily provide a free hot meal to a hungry young child named Peter, but neglects to do so. Has she breached a moral obligation? It depends on the nature of the relationship between them. If Wendy is Peter’s mother, the answer is probably yes (barring unusual circumstances). If Wendy is the unrelated owner of a local restaurant, the answer is probably no. Or consider John, who fails to pay his driver, Susan, for a ride to the airport. If Susan is his taxi driver, he likely has breached an obligation. If Susan is his sister or spouse, however, he likely has not.

These examples highlight a difference between two relationship types described in the literature: communal and exchange (Clark & Mills 1979, 2012; see also Fiske 1992). Communal relationships are often exemplified by friends, family, and romantic partners. In these relationships, people (normatively) assume a special responsibility for one another’s welfare. They track each other’s needs and desires (Clark et al. 1989), note the responsiveness of the other to their own needs (Clark et al. 1998), and offer

non-contingent support as necessary to promote partner welfare (Clark et al. 1987). Failure to offer such support reduces liking, elicits hurt feelings (Clark & Mills 1979; Lemay et al. 2010), and, we expect, triggers negative moral judgments (e.g., resentment). By contrast, failure to directly compensate the other for needs-responsive support in a communal relationship typically does not cause such negative responses (as with John/Susan above).

Compare this to exchange relationships, often exemplified by casual acquaintances or customers/sellers. In these relationships, needs typically are not tracked (unless for purposes of selling or exchanging), but the other’s contributions to joint tasks are tracked (Clark et al. 1989; Clark 1984). Goods and services are (normatively) provided on a contingent basis, and costly help is generally not offered in non-emergency situations (Clark et al. 1987). Here, repaying debts and willingness to accept payments leads to enhanced liking (Clark & Mills 1979) and does not elicit negative judgments (McGraw & Tetlock 2005), whereas failure to pay likely would be judged as morally objectionable (and refusal to accept payment would seem strange and uncomfortable, if not immoral). The *degree* of perceived wrongfulness of an action likewise depends on relational context (Simpson & Laham 2015; Simpson et al. 2016; Tepe & Adymi-Karakulak 2018).

Communal and exchange relationships serve different functions. Bugental (2000) proposed some overlapping functions and added others as shown in Table 1. She ties each one to a specific adaptive goal or recurrent coordination problem faced by our species.

Everyday relationships serve one or more of these functions to varying degrees in different contexts. For example, a parent-infant relationship normatively serves the attachment and hierarchy functions across almost every context, but *not* the mating, reciprocity, or coalition functions. Teammates are expected to serve the coalition function, with a captain also serving the hierarchy function, but in most societies the captain normatively does *not* serve the mating, attachment, or reciprocity functions with other team members.

Feelings of moral obligation (and judgments of blame for failing to uphold such obligations) often will be specific to the functions that are central to the relationship at hand. So, a parent would be heavily blamed (and would likely feel guilt) for failing to serve the attachment function with their infant, but a captain would not be so blamed (nor would likely feel guilt) for failing to serve this function with a teammate, and so forth.

Two final points regarding the importance of asking who “we” are. Tomasello focuses on “we” relations that are voluntarily entered into by individuals who regard each other as relevantly equal in terms of obligations, rights, and power; but the parent-infant example highlights that at least some relationships are unequal in these respects and may be non-voluntarily entered into. Indeed, some “we” relations are imposed on people by the situations in which they find themselves, some of which may be functional for one party to the relationship but not the other (see Kelley et al. 2003).

Second, Tomasello focuses on negative judgments resulting from obligation-failures (the sticks). It will be fruitful to consider the personal and interpersonal rewards to be gained by obligation-fulfillment (the carrots) as well. Previous work suggests that serving relationship functions in ways that are desired and which exceed normative expectations will disrupt smooth, habitual interdependent routines and will likely elicit positive emotions (Berscheid & Ammazalorso 2001). Might moral praiseworthiness judgments be similarly elicited?

Table 1. (Clark et al.) Relationship functions, adapted from Bugental (2000).

Function	Adaptive Goal/Coordination Problem to be Solved
Mating	Finding and maintaining sexual partners; ultimately, producing and ensuring the survival of offspring
Attachment	Ensuring that a person's well-being is secure, without strings attached to the giving or receiving of support; maintaining safety; encouraging learning
Reciprocity	Coordinating cooperative behavior between people with similar (or equal) status, power, or responsibility
Hierarchy	Coordinating cooperative behavior between people with different (unequal) status, power, or responsibility
Coalition	Forming and maintaining a group identity with (potentially unrelated) others working toward common goals

To summarize, the varied nature of relational contexts and functions shape moral judgments (Bloom 2011; Clark & Boothby 2013; Haidt & Baron 1996; Rai & Fiske 2011), as well as many other psychological phenomena (Reis 2008; Clark et al. 2017). Future work should consider who “we” are, taking into account (1) the different functional and normative bases of different relationship types, (2) both voluntary and non-voluntary forms of interdependence, and (3) both positive and negative moral judgments.

Integrating perspectives: How the development of second-personal competence lays the foundation for a second-personal morality

John Corbit  and Chris Moore

Department of Psychology and Neuroscience, Dalhousie University Life Sciences Centre, Dalhousie University, Halifax, Nova Scotia B3H 4R2.

john.corbit@dal.ca; chris.moore@dal.ca
<https://esldalhousie.wixsite.com/esdl>

doi:10.1017/S0140525X1900236X, e67

Abstract

The integration of first-, second-, and third-personal information within joint intentional collaboration provides the foundation for broad-based second-personal morality. We offer two additions to this framework: a description of the developmental process through which second-personal competence emerges from early triadic interactions, and empirical evidence that collaboration with a concrete goal may provide an essential focal point for this integrative process.

As Tomasello highlights in the target article, a key distinguishing feature of human collaboration is the ability to form a second-personal agent “we,” which regulates “I” and “you” in our commitments towards achieving a joint goal. Collaboration characterized by second-personal agency is the foundation for the development of a broad-based second-personal morality of

fairness and mutual respect. Further, through engaging in joint intentional collaborative interactions, children come to understand the self-other equivalence of “I” and “you,” thereby providing the foundation for individuals to relate to each other as equals. Collaboration structured by second-personal agency appears to be unique to humans, and the apprehension of perspectival equivalence involved in this form of social understanding may well explain why humans also uniquely show a commitment to treating collaborators fairly (Tomasello et al. 2005).

We have two goals in this commentary on the target article. First, we intend to apply a developmental account of second-personal agency focused on the role of triadic interactions (Moore & Barresi 2017) to complement Tomasello’s developmental account of second-personal morality. Second, we highlight recent findings that uncover the social contextual factors within joint intentional collaboration that activate second-personal moral concerns.

Whereas Tomasello argues for a fundamental role of joint intentional collaboration, his account is relatively silent on how a joint intentional schema comes about developmentally. Following Moore and Barresi (2017; see also Barresi & Moore 1996), we propose that it is through the triadic interactions arising in the second half of the first year that children gain the social experience necessary to form the joint intentional schemas that Tomasello argues are essential for second-personal morality. Triadic interactions contain second-person information of various forms (see Moore & Barresi 2017), but we suggest that it is through experiencing triadic interactions whereby infants and interactive partners share intentional orientations to objects and states of affairs that infants are able to coordinate their own first-person intentional orientation with their third-person perspective on the intentional orientation of their partner (Moore & Barresi 2017). It is this experience of shared intentionality within triadic interactions that allows children to form a joint intentional schema. The ability to form a joint intentional schema allows children to engage in the type of dual-level collaboration wherein, as Tomasello argues, children come to recognize the self-other equivalence from which a second-person morality develops (Tomasello 2019).

A central argument in Tomasello’s target article account is that a sense of fairness develops through joint intentional collaboration. This preference for fairness emerges through the formation of a joint-agentive “we,” which is embodied by mutual action towards a common goal. Within this form of dual-level collaboration, children come to recognize self-other equivalence with their collaborative partners. To date, much of the evidence that collaboration increases children’s concern for fairness has come in the context of collaborating to earn resources (Corbit et al. 2017; Hamann et al. 2011; Warneken et al. 2011). However, when children collaborate to earn resources, there are several social contextual factors that may activate fairness concerns; working to earn resources, working to achieve a common goal, and enjoying a positive social exchange. Thus, although Tomasello convincingly argues for the importance of joint intentional collaboration in the target article, parsing the influence of these factors will provide a better understanding of the psychological processes through which joint intentional collaboration fosters fairness concerns.

In order to systematically investigate the influence of each of these factors, Corbit (2019) assessed children’s (3–5 years old) sharing before and after engaging in three collaborative contexts: achieving a shared goal of earning resources, achieving a shared goal independent of resources, and playing a social game. Children increased sharing to collaborators with whom they had accomplished a shared concrete goal, independently of

whether the resources they shared were earned together or obtained outside of the collaborative context. Importantly, sharing increased only when children collaborated towards a concrete goal, and not when the goal of their collaborative interaction was to play a social game. An important question that emerges from these findings is why might a concrete goal increase sharing, whereas a primarily social goal does not?

Joint intentional collaboration towards a concrete goal may serve a unique role in fostering a second-personal moral obligation. When a shared goal is to play a social game there is less need for mutual (second-personal) regulation of the role for “I” and “you.” This is because the goal of playing a social game can be accomplished in a myriad of ways; so as long as partners remain engaged, their social goal will be achieved. When there is a concrete goal, and “we” are dependent on one another to achieve that goal, then “I” must regulate “you” (and vice versa), guided by the shared intention “we” have. Indeed, when children collaborate towards a concrete goal, over repeated iterations they are able to form role ideals that structure the way each collaborator must fulfill their role towards achieving the joint goal. If the goal that “we” have is to engage socially, then the role that “you” and “I” play is secondary to that of the second-personal agentive “we.”

An obligation to treat collaborators fairly necessitates that “you” and “I” are recognized as equivalent, yet distinct, agents currently engaged in shared intentionality. Fairness is intrinsically a relative process where I compare my stake to yours, and “we” as equivalent agents can agree on a mutually satisfactory outcome. We propose that in collaboration where the role “I” and “you” must be mutually regulated by “we” in order to achieve a shared concrete goal, the resulting commitment to mutually satisfactory (“fair”) outcomes will be greater than when “we” is the focal point of an interaction and the individual roles “you” and “I” play are secondary. Thus, a shared concrete goal may provide the focal point within joint intentional collaboration to facilitate the integration of first-, second-, and third-personal information providing a foundation for a broad-based second-personal morality.

Psychological consequences of the normativity of moral obligation

Stephen Darwall 

Department of Philosophy, Yale University, New Haven, CT 06520-8306.
stephen.darwall@yale.edu
<https://campuspress.yale.edu/stephendarwall/>

doi:10.1017/S0140525X19002437, e68

Abstract

An adequate moral psychology of obligation must bear in mind that although the “sense of obligation” is psychological, what it is a sense of, moral obligation itself, is not. It is irreducibly normative. I argue, therefore, that the “we” whose demands the sense of obligation presupposes must be an ideal rather than an actual “we.”

Perhaps unsurprisingly, I broadly agree with Tomasello’s target article. I am a moral philosopher, however, not a psychologist. Still, even if moral philosophy is an a priori discipline at the

foundational level where I tend to work, it cannot avoid giving some hostages to psychological fortune. And mine does so in spades, since I argue that central aspects of deontic morality – obligation, right, wrong, and so forth – are conceptually related to the psychological states P. F. Strawson called “reactive attitudes” through which we hold people morally accountable (Darwall 2006; 2013b; 2013c; Strawson 1968). So, it has been enormously exciting, and something of a relief, to find confirmation in Tomasello’s work for the second-personal psychology of mutual accountability.

Psychologists tend to identify “prosocial” motivation with sympathetic concern and the desire to benefit. As Tomasello shows, however, what is essential to distinctively human forms of cooperation is not a desire for others’ goods, but a deontic motivation: a “sense of obligation,” including the desire to treat others fairly and with respect. This is a motive of *right* rather than good. It is conceptually related to what can be warrantedly demanded of us. The very idea of moral obligation, I argue, is the concept of what we can justifiably be held accountable for doing, what it would be *blameworthy* to fail to do without excuse (Darwall 2006; 2013b; 2013c).

Although the “sense of obligation” is psychological, what it is a sense of is decidedly not, any more than mathematic intuition, clearly a psychological phenomenon, concerns something psychological. The object of the latter is an a priori mathematical structure, and that of the former is an essentially *normative* one. This fundamental fact has significant consequences that need to be borne in mind.

Tomasello distinguishes two different stages at which the sense of obligation enters, both ontogenetically and phylogenetically. Whether in collaborative play in early childhood or the obligate collaborative foraging of our hunter-gatherer ancestors, an initial stage involves a sense of obligation *to* individual partners as having equal claims to the collaborative benefits and to respect as an equal partner. A second stage involves wider cooperation within a cultural group and a sense of obligation, less to particular individuals, than “objectively” or “period.”

The psychological, as well as philosophical, engine driving the sense of obligation in both cases, for Tomasello, is that joint undertakings create a “we” with the standing to make demands of collaborating parties. When individuals cooperate voluntarily, their “we” can make demands of each that ratifies claims the other has on them as an equal partner, underwriting the parties’ sense of obligations *to* one another. When the cooperative scheme is more diffuse and less voluntary, as with a cultural group, the “we”s demands are not on behalf of individuals. The resulting sense of obligation is not *to* anyone individually. They are on behalf of shared cultural values that inform a sense of “objective” obligation.

Now, as we noted, obligation is a normative, rather than a psychological (or sociological) phenomenon. It follows that any “we”s expectations or demands can generate obligations only if they are *justified* demands that it would be *blameworthy* to ignore or flout. It is important to see, moreover, that being culpable cannot be understood in terms of what any actual “we,” whether voluntary or implicit, *actually* blames on pain of robbing obligation of its normativity. What is blameworthy is what there is *justification* to blame. I argue, therefore, that blame involves a standpoint of a presupposed moral community of second-personally competent persons – that is, moral agents or subjects, as such (Darwall 2006; 2013b; 2013c). It presupposes an ideal “we.”

No de facto group can create obligations by default. Individuals become obligated to one another through agreements only because

of a background requirement on all moral agents to keep agreements (other things being equal). We can become obligated to one another through agreements only if it would be blameworthy to violate them without excuse, where this consists in there being justification for demands (and blame for violators) from a presupposed “we” of any (second-personally competent) person. The inescapable point is that both “second-personal morality” and “objective morality,” as Tomasello calls them, presuppose morality in the intrinsically normative sense. They presuppose what we might call “objective morality without the scare quotes.”

Here are two reflections of this fact. First, voluntarily agreeing to something can obligate one only if a would-be partner’s forcing you to do as they wish without your agreement would be wrong and would wrong you. This background moral fact is presupposed. The wrongness of forced “cooperation” cannot completely derive from the demands of any de facto group, obviously not from a not-yet-constituted group, and not from any more encompassing de facto group, either.

Second, Tomasello notes experimental evidence that young children tend to equalize the benefits of collaboration. It seems implausible that they generally agree to this in advance or “write it into” the contract constituting their “we.” Much more likely is that there is a moral presumption of equal treatment that is presupposed in the phenomenon of cooperation itself (Darwall 2006). Tomasello hypothesizes “self-other equivalence” and that the children are “equal causal forces.” But why suppose that this is generally true de facto? The relevant equivalence is irreducibly normative and moral, not empirical. In the relevant sense, it holds even if (say, because one is ambidextrous and the other is not), one is actually capable of playing both roles and the other is not, or if one contributes more than the other in a strictly causal sense.

Obligation at zero acquaintance

David Dunning^a , Detlef Fetchenhauer^b
and Thomas Schlösser^b

^aDepartment of Psychology, University of Michigan, Ann Arbor, MI 48109 and

^bInstitute for Sociology and Social Psychology, University at Cologne, 50931 Köln, Germany.

ddunning@umich.edu

<https://lsa.umich.edu/psych/people/faculty/ddunning.html>

detlef-fetchenhauer@uni-koeln.de

<https://www.iss-wiso.uni-koeln.de/en/institute/staff/f/prof-dr-detlef-fetchenhauer/>

t.schloesser@uni-koeln.de

<https://www.iss-wiso.uni-koeln.de/de/institut/personen/s/pd-dr-thomas-schloesser/>

doi:10.1017/S0140525X19002498, e69

Abstract

Social obligation begins far before people establish explicit cooperative relationships. Research on trust suggests that people feel obligated to trust other people even at zero acquaintance, thus trusting complete strangers even though they privately expect to be exploited. Such obligations promote mutually beneficial behavior among strangers and likely help people build goodwill needed for more long-lasting relationships.

We resonate with Tomasello’s observation that much of human cooperation is driven by something akin to obligation. Pro-social behavior is not so much an act of “giving” as it is of “giving in” – to social rules, roles, and duties (see Cain et al. 2014).

We not only find much agreement, we suggest that the scope of his analysis applies far wider than Tomasello asserts. He proposes that other-directed obligations arise when people enter explicit collaborative activities with shared goals and joint intentions.

We, however, would argue that obligation governs human behavior much sooner than that. For communities to thrive and develop, their members must honor certain rules and obligations no matter how anonymous, transient, or minimal their supposed relationships might be (Henrich et al. 2010).

Work in our labs on interpersonal trust directly demonstrates the obligatory nature of pro-social behavior even at zero acquaintance (for reviews, see Dunning et al. 2019; Fetchenhauer et al. 2017). In laboratory trust games, we give participants \$5 and tell them they can either keep the money or hand it over to a complete stranger whom they will never meet and who will never meet them. Importantly, if they hand over the money, we will inflate it to \$20 and give the stranger their own choice. They can keep the \$20 or give \$10 back to the participant. Such a choice represents the core decision at the heart of many definitions of trust: People must decide whether to open themselves to potential exploitation by another person in order to produce a chance of personal benefit (Rousseau et al. 1998).

The choices that participants make contradict a strict account of neoclassical economics. Although participants on average think there is only a 40% to 50% chance the stranger will hand money back, clear majorities decide to trust that stranger (Fetchenhauer & Dunning 2009). Pointedly, total strangers trust one another in one-off anonymous exchanges even though many expect their trust to be exploited and their welfare to suffer (see also Berg et al. 1995). They gamble on the sociality of the stranger even though they would rarely make the same bet if gambling instead on a lottery wheel (Fetchenhauer & Dunning 2012). To an economist, the source of risk should not matter. However, to our participants, it very much does.

Our research converges toward an explanation of this high rate of trust that, if not the same as Tomasello’s, certainly rhymes with it. Many indicators point to social obligation as the source of this decision. Participants on average report that trusting a stranger is more something they should do rather than something they want to do (Dunning et al. 2014).

More important, the emotions that people attach to trust follow the logic of social rules and obligations. The prospect of breaking a social obligation produces high levels of anxiety and guilt (Higgins 1987). In our games, those who feel more anxious, guilty, and remorseful about distrusting the other person than they do about risking their money are the ones who trust the most – and these emotions predict trust beyond any economic factor (Dunning et al. 2014; Schlösser et al. 2016).

Further work suggests the specific obligation that participants feel they must honor, again one that is compatible with Tomasello’s account. Participants trust because they feel they must respect the character of the other person (see Dunning et al. 2016). To distrust the other person would be to call that person’s integrity into question. People express intense anxiety at the prospect of doing so (Schlösser et al. 2015).

In addition, anecdotally, participants express that they choose to trust not out of expectation but rather out of “hope” and “faith”

(Dunning et al. 2012). When asked, our participants explain their trustfulness by noting, “I just want to live in a world where people trust each other.” Thus, they appear to take a “we” perspective, as Tomasello would put it, even at zero acquaintance. And as “we > me,” they feel obligated to trust even if from a “me” perspective this might be irrational.


In retrospect, the adaptive value of this obligation to trust at zero acquaintance is clear. First, it allows strangers, not only those already in collaborative relationships, to benefit from the behavior of others (Henrich et al. 2010). We hand over our credit cards to complete strangers in restaurants. We purchase that trinket on eBay and assume it will show up in the mail. We procure our daily calories by going to stores run by people we never meet. Perhaps another way to put it is that we are embedded in crucial webs of implicit and fleeting collaborative relations with many of the strangers we encounter every day.

In addition, obligations to trust and respect each other from the start helps people to build the mutual goodwill needed for more intense and complex relationships involving valuable shared goals, joined intentions, and smooth cooperation (Lawler 2001).

Yet, what is largely missing in Tomasello’s article is any reference to the delicate nature of a “we > me” perspective. In our evolutionary past, there have always been temptations to act selfishly. It seems plausible that humans have developed all kinds of monitoring systems to protect them from exploitation by other group members (Cosmides et al. 2010). Thus, the observation that our participants so cynically underestimate others’ trustworthiness might be one such device to avoid being naïve and gullible.

Finally, being skeptical about others’ trustworthiness, but still feeling an obligation to trust behaviorally, turns out to be a beneficial strategy, at least in our laboratories. The strangers who receive the \$20 seem to respond to an obligation as well. Although they are expected to return money only about half of the time, they do so at a rate of roughly 80% (Fetchenhauer & Dunning 2009). As such, trust at zero acquaintance may be more an obligation than a preference, but it is one rewarded handsomely.

The divided we and multiple obligations

Bradley Franks  and Andrew Stewart

Department of Psychological and Behavioural Science, London School of Economics and Political Science, London WC2A 2AE United Kingdom.
b.franks@lse.ac.uk <http://www.lse.ac.uk/PBS/People/Professor-Bradley-Franks>
a.stewart1@lse.ac.uk <http://www.lse.ac.uk/PBS/People/Andy-Stewart>

doi:10.1017/S0140525X19002553, e70

Abstract

Tomasello’s account of the origins and nature of moral obligation rightly emphasises the key roles of social relations and a cooperative sense of “we.” However, we suggest that it overlooks the complexity of those social relations and the resulting prevalence of a divided “we” in moral social groups. We argue that the social identity dynamics that arise can lead to competing obligations in a single group, and this has implications for the evolution of obligation.

We welcome Tomasello’s persuasive account of the evolutionary and developmental origins of moral obligation as a “bipolar attitude” towards moral decisions. Tomasello suggests that the binding quality of obligation generates a sense of a moral “we” as “a kind of cooperative identity” (sect. 2.1.1, para. 1). We argue that, while Tomasello’s account of the origins of obligation to cultural group norms is correct in its outline, it overlooks significant aspects of the internal dynamics of how obligation and identity functions in cultural groups. We argue that the “we” is frequently divided so that the resulting moral obligations are multiple and sometimes contradictory. As such, moral obligation appears less bipolar and more multipolar, with its judgments looking less unidirectional than Tomasello assumes.

This moral “we” is an important claim that links Tomasello’s account with work on social identity and morality (e.g., Aquino & Reed 2002). For us, Tomasello’s picture (and the work on social identity of morality) offers only a partial account, since they overlook the possibility of a divided “we.” Namely, if a cultural group holds multiple moral values, those different values can drive different obligations for a single decision. For example, Kim and Jo are both Christians. Sam, a fellow Christian in their community, was caught stealing from both Kim and Jo. In response, Kim advocates punishment, while Jo advocates forgiveness – both supported by community norms. Sam breaks the wider community’s obligations for selfish reasons, but each of Kim’s and Jo’s responses seems to the other to fail to meet significant obligations within the Christian community’s norms. However, each is emphasising different but equally important moral motivations. This kind of phenomenon is echoed in moral ambivalence in other religious, political, and ideological groups. It can be explained straightforwardly by a social identity approach, in which the content and context of specific identities affects individuals’ thoughts and actions. Research by Van Tongeren et al. (2012; 2016) supports this notion as they have shown how religious individuals’ judgments can be shaped by differing virtues valued within the same identity group. Our ongoing research indicates that differing values within the same social identity can lead to diverging and ostensibly contradictory moral decisions (Stewart 2016; Stewart et al. forthcoming 2019). Using variations on the trolley dilemma (Foot 1967; Thomson 1985), we show how the content of a single identity can interact with different contexts to lead to contrasting moral obligations and equally contrasting moral decisions. This understanding adds nuances that are not accounted for in Tomasello’s explanation of “excuses, justifications, or apologies,” since multiple obligations can create multiple reasonable responses within a social group.

This quality is not rare in morality. The ability to support apparent internal contradictions seems to be a cornerstone of successful moral, religious, and ideological systems (i.e., ones that have been transmitted across many generations of cumulative cultural evolution, and are implicated in cultural group selection: e.g., Richerson et al. 2016). Curry et al. (2019) and others have demonstrated differing values of morality across and within cultural groups, which can lead to differing decisions over the same issue (for similar cases, see Graham et al. 2013; Greene 2013; Legare & Gelman 2008; Hall et al. 2019). Extending Tomasello’s concept of the “special social structure” for bipolar obligations, a complex moral system supports group cohesion. In complex cultural systems, failing to meet one obligation in fact may be following a competing obligation. Competing obligations may arise from different values within the group’s moral system that apply to the decision in question but which are equally or more

justifiable in the context. As such, the concept of obligation, when placed in concert with the social identity approach and its emphasis on context, offers a clearer understanding of how morality functions within and between cultural groups.

Our approach also suggests that the evolutionary picture of the origins of cultural-group obligation requires a role for a divided “we” with its multiple obligations. Tomasello’s account does not provide this: He offers a view of separate moral hunter-gatherer tribes who do not interact to any significant degree, akin to tribes in Greene’s *Tragedy of Commonsense Morality* (2013). Distinctive moral systems bind together groups and generate social selection pressures for good collaborators who can fulfil obligations. The prevalent moral norms for each group are “univalent” in directing behaviour in specific areas towards a single moral judgment and behaviour. This description is a plausible evolutionary starting point of cultural-group-based obligation. However, other accounts of hunter-gatherer sociality suggest a long evolutionary history of cross-tribe integration (e.g., Chapais 2008; 2011; Hill et al. 2011). In these accounts, groups combined via pair bonding and mating, providing descendants with both the genes of the integrated tribes as well as their cultural traditions. This inter-tribe integration and interaction lays the foundations for a divided “we” with multiple and possibly contradictory obligations for members of the combined group. We do not claim that such changes necessarily led to seamless and harmonious integration of those different moral values. Rather, our suggestion is that the recurrence of such situations helped create social selection pressures not only for good collaborators, but also for people who are skilled navigators through complex morality and its resulting multipolar obligations. In a world with a divided “we,” these are the collaborators who could most skilfully cooperate within a group holding divergent values from component tribes.

Moral systems are complex. Contradictory moral values are held between and within groups, where they can prompt different responses. The notion of obligation, when treated as monolithic, underplays the complexity and context-dependence of such responses. By treating obligation as a dynamic function, the vacillation between values provides a clearer understanding of how morality and group identity are related both today and in our evolutionary past.

Shared Intentionality, joint commitment, and directed obligation

Margaret Gilbert 

Department of Philosophy, University of California, Irvine, CA 92697-4555.
mgilber@uci.edu http://www.faculty.uci.edu/profile.cfm?faculty_id=5365

doi:10.1017/S0140525X19002619, e71

Abstract

Tomasello frequently refers to joint commitment, but does not fully characterize it. In earlier publications, I have offered a detailed account of joint commitment, tying it to a sense that the parties form a “we,” and arguing that it grounds directed obligations and rights. Here I outline my understanding of joint commitment and its normative impact.

In discussing his important empirical findings, Tomasello often invokes joint commitment, as when he says that “children at around three years of age begin to appreciate the normative force of joint commitments to collaborate” (sect. 2.1.2.1, para. 3). He ties it to a sense that one is part of a “we” and associates it with obligations to the other parties and with the parties’ protesting actions contrary to the commitment. While linking it to the making of agreements and the like, he does not offer an account of joint commitment itself. I have developed a detailed account of joint commitment through a long list of publications (e.g., Gilbert 1989; 1996; 2006; 2014), associating it with a sense that one is part of a “we,” with obligations to the other parties, and with the parties’ rebuking one another for contrary action and demanding conformity. Of the theorists of shared intentionality to whom Tomasello briefly alludes in his preliminary discussion, I stand apart in having invoked joint commitment. Indeed, I see it as the core of human sociality (Gilbert 2003). Given his citations of my work in his earlier writings (e.g., Tomasello 2014a; 2016a), it is reasonable to assume that Tomasello means to invoke joint commitment in the sense in which I intended it. This commentary summarizes my understanding of joint commitment and its relation to directed obligation.

The general notion of commitment at issue is normative rather than psychological. As I understand it, one is normatively committed to act in some way if one has reason to act in that way, in the sense that should one abstain from so acting, then, all else being equal, one has acted in error. This is not necessarily a moral error. Rather, one has not done what one ought to do, in a central sense of “ought” that does not imply a specifically moral basis for doing the thing in question. One way of becoming normatively committed is by making a personal decision.

From a normative point of view, a decision does more than create the possibility of acting in error. It excludes a number of potentially countervailing factors from consideration. Among these are the personal inclinations, desires, and self-interest of the person in question that opposes the action decided upon. In short, it has exclusionary normative force.

Whereas, with a personal decision, the decision maker commits himself (or herself), in the basic case of a joint commitment, the two or more people in question commit all of them as one. This commitment creates a possibility of error in action for each of the parties and has similar exclusionary properties. It also introduces important relations between the parties, to be discussed below. (For discussion of basic versus non-basic cases, see, e.g., Gilbert 2006, pp. 140–41.) I focus on basic cases in what follows.

What is required of each party to a joint commitment is that each talk and act as would the representative of a single body or person with a particular goal, belief, or other such psychological attribute, depending on the case, aligning their respective utterances and actions with those of the others as needed. This behavior is required primarily in the context of interactions among the parties themselves.

A joint commitment is formed once the following conditions have been fulfilled: Each party has expressed to the other or others his readiness for the joint commitment of them all in the way in question, in conditions of common knowledge. That is, roughly, it is out in the open with respect to the people in question that these expressions have been made. At this point the joint commitment has been established. (See Gilbert 2006 for some elaboration.)

Just as it takes two or more people to make a joint commitment, it takes the same people to rescind it. Importantly, then,

no one party has a unilateral power of rescission. The conditions for rescinding a joint commitment parallel those for forming or creating one: Each must express his readiness to accede to the termination of the commitment. For a relatively nuanced discussion of how to be freed from a joint commitment, see Gilbert 2006, pp. 141–44.

One reason to suppose that in a variety of circumstances, including the context of acting together, people think that a joint commitment has been made is the fact that they take themselves to have the standing to rebuke one another for acting in particular ways and to demand that they not act in those ways, where neither morality nor the self-interest of either party is clearly at issue (Gilbert 1990). It can be argued that those who are jointly committed have the standing to issue demands for conformity and rebukes for non-conformity to the commitment by virtue of their co-imposition of the commitment on the parties. Further, it is hard to explain the standing to issue demands and rebukes in any other way. For an extended discussion of these two points, see Gilbert (2018b).

I have tightly connected the standing to demand and rebuke to directed obligation: For one person to be obligated to another to perform a certain action is for the latter to have the standing to demand that action of the former. I distinguish directed obligation from another kind of so-called obligation – moral requirement – and note that I can be obligated to you to perform an action that I am morally required not to do, as when two people agree together to perform a dastardly deed. See, for instance, Gilbert (2018b).

Tomasello raises the question of the relation of joint commitment thinking to moral thought, and makes several points that accord with my position. There can be joint commitments on a large scale to uphold as a body a particular moral code. In that case, the parties will have obligations to one another to act according to that code. Again, situations of joint commitment are likely to provoke a variety of moral concerns. See Gilbert (2018a, pp. 761–64), for some discussion of the relation of joint commitment to the moral concept of fairness.

Conflicting obligations in human social life

Jacob B. Hirsh^a , Gariy Shteynberg^b
and Michele J. Gelfand^c

^aInstitute for Management & Innovation and Rotman School of Management, University of Toronto, Toronto ON M5S3E6, Canada; ^bDepartment of Psychology, University of Tennessee, Knoxville TN 37917; and ^cDepartment of Psychology, University of Maryland, College Park, MD 20742.
jacob.hirsh@utoronto.ca <http://www.jacobhirsh.com>
gshteynb@utk.edu <http://www.gariyshteynberg.com>
mgelfand@umd.edu <http://www.michelegelfand.com>

doi:10.1017/S0140525X19002425, e72

Abstract

Tomasello describes how the sense of moral obligation emerges from a shared perspective with collaborative partners and in-group members. Our commentary expands this framework to accommodate multiple social identities, where the normative

standards associated with diverse group memberships can often conflict with one another. Reconciling these conflicting obligations is argued to be a central part of human morality.

Tomasello's target article describes how the sense of moral obligation emerges from a shared perspective with collaborative partners and in-group members. Once adopted, this shared perspective motivates compliance with the normative expectations of one's group, thereby facilitating social coordination and behavioral regulation. Although we agree that adopting shared perspectives with significant others plays a central role in human motivation and cognition (Shteynberg 2015; 2018), there is a critical aspect of human morality that is not discussed in the target article: the conflict of obligations. We argue that facing obligational conflicts is a key process by which humans become moral agents, pursuing a moral code that transcends the demands of any particular social role.

Tomasello's article focuses on relatively simple situations in which a single group perspective (or "we") serves to regulate one's actions. In these contexts, there is a relatively straightforward motivation to comply with the normative expectations of one's group. Violating these norms can threaten one's group membership, while conforming to them helps to preserve relational ties (Cialdini & Goldstein 2004). As identification with the group increases, so too is there a stronger motivational force driving the desire for normative compliance (Terry et al. 1999). In single-group contexts, human morality can thus be evaluated with a single question: Is an individual conforming to or deviating from the group norms?

Although social psychologists have traditionally studied identity dynamics in the context of a single social group, there is a growing recognition that each person can belong to many partially overlapping groups at the same time (Bodenhausen 2010; Kang & Bodenhausen 2015; Roccas & Brewer 2002). Critically, each one of these group-bounded identities carries with it a unique set of normative expectations for guiding behavior. While these norms are occasionally in alignment with one another, they are also prone to conflict. A lawyer may, for example, be expected to act in a highly assertive and competitive manner at work based on the norms of her office. When interacting with her family, however, she may feel obliged to adopt a softer and more nurturing persona in accordance with traditional gender norms. If these incompatible normative standards become activated at the same time (e.g., at the office holiday party), the result is an aversive state of behavioral conflict and uncertainty (Hirsh & Kang 2016). In these cases, the obligations that derive from one's various group-based identities cannot all be enacted at once, as satisfying a moral duty in one domain can undermine it in another.

Recognizing the complexity of human identity and social life thus provides an important extension of Tomasello's framework for the moral psychology of obligation. In particular, a given individual will have not only one, but many distinct shared perspectives that can be brought to bear on a given situation. The result is a constant flux of felt obligations that wax and wane with the salience of one's various group identities. In light of this normative diversity, human morality goes beyond the simple question of whether or not to comply with the felt obligations of one's group, expanding to include the more complex question of which obligations to prioritize in a given situation. Although the easiest solution is often to ignore any moral obligations that

are not directly relevant to the current social context (Lerner & Tetlock 1999), we are nonetheless often forced to reconcile incompatible social demands. Indeed, the moral uncertainty that accompanies ethical dilemmas, wherein two or more ethical norms are in apparent conflict, has been argued to play a key role in the engagement of moral reasoning processes (Hirsh et al. 2018; Rest 1986). Navigating the tensions between diverse group obligations is thus a central aspect of human social life and a core component of moral agency and personal identity (Shteynberg 2012).

Different social environments also vary substantially in the extent to which a group's normative expectations are clearly defined and enforced (Gelfand et al. 2017). Cultural contexts that have been subject to a broader range of ecological and historical threats tend to adopt stricter normative standards for governing behavior, with stronger sanctions for those who deviate from social expectations (Gelfand et al. 2011). The result is a "tight" culture, wherein individual freedom is sacrificed for the sake of well-regulated group coordination. In contrast, cultural contexts that have faced fewer threats tend to be more permissive or "loose", allowing a broader range of socially acceptable behavioral options at the expense of having less-predictable social encounters. Given that tight cultures tend to be more socially cohesive, people within them are likely to experience the feeling of personal obligation to normative standards in a relatively unambiguous manner, as described by Tomasello. Cultural environments with looser norms and more flexible identity structures, however, are more likely to afford conflicting perspectives and inconsistent normative standards. Moral agency in such contexts is even more clearly rooted in an individual's attempts to harmonize competing expectations (e.g., through moral reasoning).

In summary, modern human life is far more complex than the single-group contexts that defined our early human ancestors. Advances in globalization, communication technology, and economic development have all contributed to a social landscape that features a diverse multiplicity of group-based perspectives and normative expectations. In attempting to live up to one's duty, a central challenge faced by any moral agent is to reconcile the conflicting obligations that derive from adopting multiple roles and identities. Attempts to integrate these conflicting demands are fundamental to the emergence of personal moral agency, and hence should not be overlooked in any account of human morality. Accordingly, a key frontier of future research is to understand the strategies that individuals use to negotiate these conflicting obligations; how the strategies vary across individuals, situations, and cultures; and the consequences they have for social coordination across multiple actors. Situating Tomasello's framework within a broader social-ecological context thus allows a more detailed description of moral obligation in human life.

Personalizing the demands of reason

Charles Kalish 

Society for Research in Child Development, Washington, DC 20006.
ckalish@srcd.org <http://corundum.education.wisc.edu>

doi:10.1017/S0140525X19002541, e73

Abstract

Children come to joint action with a generalized sense of "reason," which carries normative implications, before personalizing reasons. A general sense of ought precedes specific notions of individual perspective.

Tomasello's target article addresses the developmental origins of obligation in children's participation in shared goals and joint plans of action. However, shared intentionality and joint action depend on the ability to construe behaviors as actions, as rational and reasonable. Rationality is also at the heart of the origin story of obligation. As noted in the target article, an obligated individual must be a rational agent with "the capacities to make and regulate oneself by normative judgments about what attitudes are warranted" (Darwall 2013c, p.47). This idea of warranted attitudes provides important conceptual resources for the construction of obligation. Being warranted is already a kind of normative relation.

From a very early age, infants make sense of behaviors using something like a rational frame (Gergley & Csibra 2003). Agents have reasons to do things. If the goal is to reach an object, an agent has a reason to move toward that object. To do otherwise is a mistake, a violation of some kind. By the time children can talk, they justify and explain in terms of rational action. Of course the person seeking their dog will look where it is: it would be silly to do otherwise (Bartsch & Wellman 1995). Indeed, children's appreciation of humor is a key marker of this perspective. Nothing is funnier than to see an adult (intentionally) do something wrong. Talking into a shoe like it is a phone is nonsensical, but only in relation to a rational, correct way to use a shoe.

Children's early conceptions of rational action are more objective, depersonalized, non-representational than are adults'. Reasons are not qualities of individuals as much as they are features of situations. Perner and Roessler (2010) described this as a "teleological" perspective (see also Gergley & Csibra 2003; Kalish 2005 on "natural status"). The dog-seeker has a reason to look under the bed, because that is where the dog is. A more sophisticated conception of action personalizes reasons. Actors have reasons "in their own lights," as they represent the world to be. The person who does not know the dog is under the bed has no reason to look there. The philosophical literature distinguishes "justifying reasons" used to evaluate actions from "explanatory reasons" used to causally account for actions. Young children seem prone to take justifying reasons as explanatory. This is often characterized as a form of egocentrism. If an act makes sense from the child's point of view, it must make sense for the actor as well.

At first blush, the hypothesis that children's conception of rational action moves from general ("the reason") to second-party ("their reason") seems exactly contrary to Tomasello's account of the dyadic then later generalization of obligation. However, the kind of normative commitment involved in rational action is not obligation. Indeed, it may be the personalization of reason that establishes the dyadic context of obligation (and vice versa).

The big problem with children's early egocentric construction of rational action is conflict. Classically, the focus has been on conflict between representation and reality (e.g., false belief). Conflict between perspectives is equally challenging, although only certain kinds of incompatibility actually conflict. People have different goals, reasons to do different things. Even in a clear zero-sum

situation – say, two people in a boat desiring different destinations (Rakoczy et al. 2007) – there is no necessary incompatibility of perspective. One can think, “I have a reason to go to X, and you have a reason to go to Y.” Only one goal will be satisfied, but that is commonplace: goals frequently go unfulfilled, plans are thwarted. That you have a reason to go to Y is accidentally, empirically incompatible with my reason to go to X. A deeper conceptual conflict arises only after we have established a common goal, a joint intention. If we have agreed, “We will go to X together” and you steer the boat to Y, now there is a problem.

The problem of coordination underlies Tomasello’s account of the origin of obligation. Obligation is a cognitive mechanism to establish, maintain, and enforce shared goals and action plans. In this role, obligation provides a kind of answer to the question of “What went wrong?” when coordination breaks down. Especially in efforts to restore coordination, obligation focuses attention on the mental processes involved in an other’s behavior. Recognizing that someone is doing something wrong or silly entails understanding “warranted attitudes,” but does not necessarily involve understanding what causes wrong or silly behavior. This is just an evaluation. Obligation is part of a causal story. To understand someone as obligated is to understand them as responsive to reasons in a certain way. Reasons have to cause (or at least motivate) their behavior. That is why citing the obligation is an effective way of restoring coordination. I remind you of the reason, and expect you to adjust your behavior in light of it.

It may be in contexts of shared intentionality that children start to think about the causal processes involved in acting on reasons. When coordinating with another, a child needs to track and repair specific individual attributions of goals, and reasons. It is where the observation, “Sometimes people don’t do what they should” turns to “Why don’t they?” Participating in joint action requires concern with maintaining shared perspective. The depersonalized teleological or justifying understanding of reasoned action (people do what makes sense) becomes enriched with a personal causal/explanatory understanding.

Tomasello’s target article has identified a critical nexus in social cognitive development. The construction of a situational and dyadic notion of obligation draws on an earlier general and depersonalized notion of “acting for a reason.” In coming to understand obligations, children are revising their understanding of action and moving to a more fully representational theory of mind. Understanding people as acting for reasons always involves a tension between a general, justifying “what makes sense” form, and an individual, explanatory “what they think” form. Obligation in joint action turns out to be a critical meeting point of these two forms both conceptually and developmentally.

Is that all there is? Or is chimpanzees group hunt “fair” enough?

Angelica Kaufmann 

Department of Philosophy, Faculty of Humanities, Bar-Ilan University, Ramat-Gan, 5290002 Israel.

angelica.kaufmann@gmail.com <https://biu.academia.edu/AngelicaKaufmann>

doi:10.1017/S0140525X19002309, e74

Abstract

Tomasello claims that we lack convincing evidence that nonhuman animals manifest a sense of moral obligation (i.e., the concept of fairness) in their group activities. The philosophical analysis of distinctive evidence from ethology, namely group hunting practices among chimpanzees, can help the author appreciate the distinctive character of this behaviour as a display of fairness put into practice.

Tomasello argues that “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” (sect. 2.1, para. 1). According to the author, a chimpanzees group hunting does not amount to a display of moral behaviour in the form of fairness. Other primatologists have argued differently (Brosnan & de Waal 2003; de Waal 2006a; Høgh-Olesen 2009). Most notably, Boesch (2005, p. 629) described the complexity of this practice: “Like a team of soccer players, individuals react opportunistically to the present situation while taking into account the shared goal of the team. Some players will rarely make a goal, like defenders and goalies, but the success of the team will critically depend upon their contribution. This is very reminiscent to group hunting in chimpanzees where synchronisation of different coordinated roles, role reversal, and performance of less successful roles favour the realisation of the joint goal.”

The philosophical analysis of group hunting practices can help Tomasello appreciate the distinctive character of this behaviour, and to reconsider the claim that we lack convincing evidence that nonhuman animals manifest a sense of moral obligation, like fairness, in their group activities.

We know that these group hunting strategies require a coalition of up to four individuals: the driver, who forces the target prey (normally a red colobus monkey) into a specific direction into the tree canopy; the blocker, who makes sure that the target prey does not de-route from the direction imposed by the driver; the chaser, who attempts to chase the target prey from above; and the ambusher, who is positioned at the end of the induced trajectory and is ready to trap the target prey. Crucial to the success of the hunt is that chimpanzees shall effectively coordinate a single hunt as long as each of the participants to the hunt remains *loyal* to his role. Subsequently, the spoils are distributed taking into account the role covered during the hunting. As a result, individuals who worked harder and exposed themselves to higher risks get more than the others when it comes time to be rewarded.

As Boesch and Boesch-Achermann (2000) explained, the hunting behaviour is a learning process that starts around 8–10 years of age and that takes about 20 years of practice in order to be mastered. Different roles in the hunt require different levels of expertise and can be performed by more or less experienced individuals. This practice is very demanding because it requires (1) the capacity to understand the behaviour of another species, that is, the red colobus monkeys, and (2) the capacity to coordinate actions among individuals towards a common goal across time. Along with the authors we can establish that the mechanisms in group hunting is kept stable by (a) a mechanism for

individual recognition (that is, the mutual assignment of roles), (b) temporary memory of actions in the recent past (that is, time-delayed coordination of the actions), (c) attribution of value to those actions (that is, acknowledgment of the roles), and (d) social enforcement of those values (that is, fair distribution of the spoils).

I shall explain the philosophical implications of the analysis of this behaviour. Along with Tomasello, we can appeal to Bratman's planning theory (2014) to explain that what is crucial for a sense of obligation psychologically is indeed the capacity for shared agency which Bratman has outlined in a number of core qualities required for complex multi-agent and time-delayed action coordination. Differently from Tomasello, I argue that these correspond to the features that, according to Boesch and Boesch-Acheremann (2000), keep the chimpanzees' hunting stable. Bratman argued that the capacity for temporally extended intentional agency depends on the ability to (b) retain memory of actions in the recent past. This supports the capacity for self-governance that depends on the ability to (c) attribute values to the actions in which the agents are engaged. The latter triggers (d) mutual responsiveness to those actions in virtue of the values that are attributed to them. Then, this results in (a) a mechanism for mutual recognition that enables shared intentional activity (Kaufmann 2015; 2017).

I elaborate on this and explain that, in addition, the mechanism of group hunting might be underpinned by a relatively stable normative or, we can even say, proto-institutional component that may be applied to explain how far fairness can extend among chimpanzees. I shall call this phenomenon *action-free perdurance*, and I will suggest how the display of fairness among chimpanzees, and presumably among other animal species, is made possible.

If we think about what distinguishes a hunting team from, say, a football team, arguably this is what we shall call "*action-free perdurance*" (of a state of affairs). Social entities, such as football teams, with an institutional component exist prior and subsequently to the online coordination of shared actions. Their existence is independent of the actual or online performance of a shared action. Instead, for group hunting, it seems that the normative aspect exists only in an online mode – that is, during the actual performance of a shared action. This is because the cognitive capacities that allow for conceptual thought and language may be requested for creating institutional reality, and for maintaining the normative power of institutions. Since language seems to facilitate the establishment and the perdurance of a state of affairs in the absence of online coordination of shared actions, it follows that creatures lacking conceptual capacities, presumably, lack those for the creation of institutional facts as well. Nevertheless, this different cognitive equipment allows for *action-bound* acts of fairness as a consequence of hunting practices.

Last, it is interesting to notice that, hunter-gatherer populations of humans in Paraguay and Venezuela carry on hunting practices with analogous developmental paths to those of chimpanzees, both in terms of the time necessary for learning and of the age range during which individuals hunt more frequently and more efficiently (Kaplan et al. 1985; Walker et al. 2002). Further comparative analysis with chimpanzees group hunting may even tell us about the continuity of shared agency in species other than us.

The moral obligations of conflict and resistance

Melanie Killen^a  and Audun Dahl^b 

^aDepartment of Human Development and Quantitative Methodology, University of Maryland, College Park, MD 20742 and ^bPsychology Department, University of California, Santa Cruz, CA 95064.

mkillen@umd.edu www.killenlab.umd.edu

dahl@ucsc.edu <https://esil.ucsc.edu/people/audun-dahl/>

doi:10.1017/S0140525X19002401, e75

Abstract

Morality has two key features: (1) moral judgments are not solely determined by what your group thinks, and (2) moral judgments are often applied to members of other groups as well as your own group. Cooperative motives do not explain how young children reject unfairness, and assert moral obligations, both inside and outside their groups. Resistance and experience with conflicts, alongside cooperation, is key to the emergence and development of moral obligation.

How do humans acquire moral obligation, a unique species-specific ability that enables individuals to live in large groups peacefully, create rules of justice, fairness, and rights, and protect the interests of the minority? Tomasello's thought-provoking answer, motivated by an evolutionary perspective, starts with cooperation, through which identification with groups, and the internalization of cultural norms, provides the basis for an objective moral obligation. Tomasello proposes that individuals live up to their moral obligations because they identify with their social groups: "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" (sect. 2.2.3, para. 3).

By rooting moral obligations in cooperation and group identification, however, Tomasello's account invites two challenges: (1) How to explain that individuals, including young children, separate group norms from moral norms, often sparking conflicts; and (2) How to explain that individuals, including young children, extend moral norms to members of other groups? Answers to these questions will explain how individuals seek to rectify failures of moral obligation to others that permeate human existence.

Concerns with others' welfare, rights, and justice often conflict with group norms or authority commands. These conflicts give rise to civil rights movements, corrections to gendered and racial discrimination, and intrapersonal dilemmas about whether to obey an authority (Killen & Smetana 2015; Turiel & Dahl 2019). The abundance of such conflicts have led philosophers and developmental psychologists to separate moral concerns from norms imposed by authorities and groups (Sen 2009; Turiel 2002). This contrast is evident, for instance, when authority commands conflict with moral obligations to protect others: "a good soldier obeys orders, but a good human being doesn't massacre the innocent" (Korsgaard 1996, p. 102; see Turiel & Dahl 2019).

On this view, morality has two key features: (1) moral judgments are not solely determined by what your group thinks, and (2) moral judgments are often applied to members of other groups as well as your own group. Initially, some researchers

proposed that young children confuse group norms and moral concerns with rights, welfare, and justice (Kohlberg 1971; Piaget 1932). Over the past 40 years, however, researchers have shown that children separate moral concerns with welfare, rights, and justice from authority commands and group consensus by around three years of age (Dahl & Killen 2018; Killen & Smetana 2015; Turiel 2002). Based on this research, we contend that resistance to and reflecting on social conflicts, alongside cooperation, is key to the emergence and development of moral obligations.

Thus, while we grant that cooperation plays a fundamental role in moral development, young children also disagree with parents, peers, and siblings about which clothes they should wear, who owns which toy, and whether it is okay to hit back (Dahl 2014; Nucci & Weber 1995; Ross et al. 2014; Smetana et al. 2012; Wainryb et al. 2005). The dynamics of young children's resolutions to peer conflict reveal that prosocial behaviors during conflict are related to peaceful post-conflict interactions, indicating that cooperation often stems from interpersonal conflict (Spivak 2016).

A further question that arises is how do children determine which cultural norms are legitimate, and which norms perpetrate negative social interactions with others? Cultures are notoriously complex, emanating and disseminating messages that are contradictory and inconsistent (Gelfand et al. 2016). Children evaluate social norms, often rejecting normative expectations that could harm others or deny deserved resources (Rizzo et al. 2018). In fact, the existence of conflict between different types of obligation demonstrates that moral norms are often distinct from group norms (or there would be no conflict).

Tomasello proposes that children (and early humans) adhere to an obligation to apply moral codes to the in-group prior to the out-group, until at some point, when a "universalization" based on the identification with all humans (not Martians) takes place. In fact, young children endorse moral obligations toward both in-group and out-group members. Conflicts with in-group members reveal that interactions are not always cooperative: Siblings have extensive conflicts over sharing toys and refraining from the infliction of harm (Goulding & Friedman 2018), and same-gender peers have conflicts about whose turn it is to play a game.

Conflict within the in-group may even appear prior to conflict with the out-group, given the lack of exposure to other groups early in development (exceptions include children from biracial and bicultural families; see Gaither et al. 2014). Although some moral transgressions toward in-group members are rated as more serious, moral transgressions toward out-group members are still judged as wrong (Rhodes & Chalik 2013; Mulvey 2016a). With age, children recognize that in-group preferences are unfair (Rutland & Killen 2015).

No doubt, in-group preferences and outgroup distrust create challenges for children (and adults) when considering inter-individual treatment. Preference for the in-group occurs when there is outgroup threat (Tajfel & Turner 1986). Without an obvious threat, though, children have a propensity to seek out other children whether they are from the same group or a different group (Nesdale 2004). Children from various racial and ethnic backgrounds also reason that moral obligations apply to those from different gender and ethnic groups (Killen 2007).

The developing ability to separate moral obligations from group norms, and to apply moral obligations to both in- and out-group members, provide the basis for addressing pressing societal questions. From an early age, children recognize that non-

cooperative behavior is often necessary to achieve moral aims. This leads them to rectify inequalities, resist unfair practices, and challenge stereotypic expectations in situations involving inter-individual treatment (Elenbaas 2019; Rizzo et al. 2018; Rochat et al. 2014). Children understand the cost of resistance to group norms, however, and are concerned about such consequences as social exclusion (Mulvey 2016b; Rutland et al. 2015) and ostracism (Song et al. 2015).

Balancing competing moral and group obligations begins at an early age, sometimes collaborating and sometimes challenging group norms. Yet these balancing acts continue to develop through intra- and interpersonal as well as intra- and intergroup conflicts across the lifespan, shaping the trajectory of human societies.

The sense of obligation in children's testimonial learning

Pearl Han Li , Annelise Pesch and Melissa A. Koenig

Institute of Child Development, University of Minnesota, Minneapolis, MN 55436.
lix5592@umn.edu pesch047@umn.edu mkoenig@umn.edu
<http://www.cehd.umn.edu/icd/people/faculty/cpsy/koenig.html>

doi:10.1017/S0140525X19002486, e76

Abstract

We extend Tomasello's discussion of children's developing sense of obligation to testimonial learning. First, we review a battery of behaviors in testimonial exchanges that parallel those described by Tomasello. Second, we explore the variable ways in which children hold others accountable, suggestive that children's evaluations of moral and epistemic responsibilities in joint collaborative activities are distinct.

Tomasello argues that children's developing sense of obligation towards collaborative partners and in-group members develops within joint collaborative activities. Although we agree that the human sense of obligation is evident in collaborative activities, we believe that certain aspects of Tomasello's discussion can be extended to the case of testimonial learning. However, recent evidence that children treat epistemic and moral responsibilities distinctly suggests that the behaviors children use to hold others accountable may differ depending on the type of obligation or commitment that has been broken.

In addition to Tomasello's argument that there is a sense of obligation in moral cooperation, *testimonial exchanges* can also allow speakers and listeners to make commitments to each other, showcasing *epistemic cooperation*. In other words, learning from testimony depends not only upon monitoring speakers for signs of knowledge or competence (for a review, see Harris et al. 2017), but also upon recognizing those with willingness to cooperate and fulfill their commitments to tell the truth. In fact, much philosophical research on the role of trust in testimonial learning has argued that when addressees believe what they are told, take a speaker's word for it and trust her for the truth (Ross 1986), they recognize the speaker's commitment to the truth of what she says (Moran 2005), receive a unique kind of epistemic warrant to rely on that speaker (Holton 1994) and can hold

the speaker epistemically responsible as a result (McMyler 2011). In other words, when addressees believe something based on a speaker's authority, their trust is grounded in a mutual regard between the speaker and the addressee, which is similar to Tomasello's "second-personal responsibility" in joint collaboration. Similar to Tomasello's view that breaches of obligations can result in reactive attitudes (e.g., blame, protests), contemporary epistemologists also argue that listeners who accept a speaker's claim acquire a suite of rights, including rights to hold speakers accountable, rights of complaint, rights to demand an apology, and rights to forgive (Gilbert 1992; Hinchman 2005; McMyler 2011). Moreover, if addressees take someone "at her word," they are entitled to pass responsibility back to the speaker when their belief is challenged, if they so choose (i.e., "pass the epistemic buck," Brandom 1994; Goldberg 2006).

Consistent with work in philosophy, a growing body of developmental research suggests that in testimonial exchanges that involve forms of commitment, children engage in a battery of behaviors that parallel those described by Tomasello in the domain of action. Tomasello describes that the human sense of obligation to collaborative partners manifest in (1) reactive attitudes when the partner fails to play her role in a joint commitment; (2) leave-taking behavior when breaking a commitment; and (3) fairness and reciprocity among second-personal agents. Similarly, when a speaker intentionally offers incorrect information in direct testimonial exchanges, even infants detect, interrupt, and actively correct speakers who make mistakes either by misnaming common objects (Begus & Southgate 2012; Grassmann & Tomasello 2010; Jaswal 2004; Koenig & Echols 2003; Pea 1982) or by presenting difficult or anomalous requests (e.g., "Can you bring me the refrigerator?"; Grosse et al. 2010; Shwe & Markman 1997; Wellman et al. 2019). Not only can children direct objections at deviant speakers, they also actively protest against addressees who fail to fulfill the commands of a speaker (Rakoczy & Tomasello 2009), demonstrating an early understanding of the cooperative roles of both parties in testimonial exchanges. As for the analogy of children's "leave taking" behavior, recent evidence reveals that children trust a speaker who acknowledges her ignorance and a speaker who is accurate at similar rates, but reject information from an inaccurate speaker who acts as though listeners should believe her false claims (Kushnir & Koenig 2017). It is possible that after hearing a speaker acknowledge her ignorant state of mind, children credit the speaker for clarifying her epistemic limits. Thus, compared to an inaccurate agent who fails to do so, children are more likely to trust the previously ignorant speaker for new information. Further research suggests that children's reciprocity towards speakers is also similar to that of cooperative activities. Indeed, young children prefer to share more resources with accurate informants over inaccurate ones (Ronfard et al. 2019), and choose to reciprocate with informative communicative partners over speakers that intentionally withhold information (Dunfield et al. 2013). Together, these lines of research suggest that children hold others not only morally responsible, but also epistemically responsible.

Tomasello argues that children display various behaviors indicative of their understanding of obligation in joint intentional collaboration. Do these behaviors differ depending on the type of commitment that has been established? Recent work suggests that children may evaluate moral and epistemic commitments differently. For example, an agent's failure to uphold a spoken agreement reduced children's decisions to wait for future rewards and their willingness to share with that agent, but did not affect

their willingness to learn new information from her (Pesch & Koenig 2018). Likewise, although an in-group member's antisocial behavior reduced children's liking of and willingness to share with her (which is consistent with Tomasello's review of children's understanding of moral obligations within their in-group), children's decisions to seek new information from (Hetherington et al. 2014) or imitate (Wilks et al. 2018) the in-group member were not attenuated. Work like this suggests that a breach in *moral obligations* may not affect children's evaluations of whether the agent can fulfill their *epistemic obligations* (and vice versa).

In conclusion, we think that various aspects of Tomasello's obligation account can be extended to the case of testimonial learning, and that children likely treat testimonial transactions as implicating distinctively interpersonal commitments and responsibilities. Appreciating the shared nature of epistemic responsibility between speakers and listeners opens new opportunities for future empirical research to identify the various devices used by speakers to commit themselves to the truth of what they say, and the ways in which children detect epistemic commitments and violations in their decisions to trust.

A lifelong preoccupation with the sociality of moral obligation

Zoe Liberman^a and John W. Du Bois^b

^aDepartment of Psychological & Brain Sciences, University of California, Santa Barbara, CA 93106-3100 and ^bDepartment of Linguistics, University of California, Santa Barbara, CA 93106-3100.

zoeliberman@ucsb.edu labs.psych.ucsb.edu/liberman/zoe
dubois@ucsb.edu dubois.faculty.linguistics.ucsb.edu

doi:10.1017/S0140525X1900253X, e77

Abstract

Tomasello provides compelling evidence that children understand that people are morally obligated toward members of their social group. We call for expanding the scope of inquiry to encompass the full developmental trajectory of humans' understanding of the relation between moral obligation, sociality, and stancetaking in interaction. We suggest that humans display a lifelong preoccupation with the sociality of moral obligation.

Tomasello provides a compelling account in the target article that moral obligations play a foundational role in the evolution of human ultra-sociality, unique among primates (Tomasello & Vaish 2013; Tomasello 2014b). In particular, Tomasello acknowledges the importance of "we-intentions" (Searle 1995b; 2010), through which social actors performatively constitute human institutions and the facts they make possible, such as promises, debts, and other obligations. Tomasello clearly highlights the ability of children as young as three to engage in "we-intentions", and to think of themselves as members of sociocultural groups, with important consequences for moral obligation. But by focusing on young children, Tomasello offers little to clarify how the capacity for attunement to shared intentionality might develop

in its earliest stages, and how it might continue to play a pivotal role in social interactions later in life. Older children, adolescents, and adults continue to participate in dynamic exchanges of stances (Du Bois 2007) that invoke, evoke, or simply presuppose moral obligations. We believe that the sociality of moral obligation represents a lifelong preoccupation of humans. Its foundations are evident early in infancy, and the dynamics of its constitution and the scope of its effects become both more nuanced and more comprehensive with development.

How early in development do humans see the world as structured by social groups, and see themselves as morally obligated to follow their group's norms? Tomasello argues that a shift happens around 3 years of age. Specifically, he writes, "Ontogenetically, it is only children after three years of age who identify with their culture in this manner" (sect. 2.2, para. 1). But as Tomasello's own recent work shows, even toddlers (18-month-olds) protest when people do not conform to a partner's action, particularly when the action is situated within a normative game (Schmidt et al. 2019). That is, when toddlers see someone do a novel action, they are more likely to protest a second person who performs a different action, in cases in which the first person clearly marked that the action corresponded to a norm ("How we do it"). Although these actions are taking place within a partnership, they seem to indicate a broader understanding of obligations incurred via a group's social norms.


In fact, a rapidly growing body of research suggests that the origins of reasoning about people as members of social groups is already in place in infancy, even before children would be able to verbally protest (Lieberman et al. 2017a). For example, in addition to preferring to approach and learn from people from their own group (e.g., Begus et al. 2016; Buttelmann et al. 2013; Kinzler et al. 2007), infants in the first year of life expect people from a group to act alike (Powell & Spelke 2013), to affiliate (Lieberman et al. 2017b), and to be similar to one another (Lieberman et al. 2016). Thus, even preverbal infants appear to structure the social world into social groups, and to make inferences about people's likely traits and behaviors based on group membership. Can these abilities be reconciled with an understanding of the "we-intentions" that would provide evidence of cultural group membership? If so, are they indicative of potential developmental precursors which can grow with the experience of participating in the norms of a cultural group? If not, is there any type of non-verbal evidence that could suggest attention to "we-intentions" in infancy? In general, some clarification is needed about where children's clear understanding of moral obligation to members of their own cultural group comes from.

This points to further questions of mechanism for the learning of *we-intentions*. Moral norms must be experienced, and learned, through the scaffolding of interactional practices of stancetaking. Moral evaluation is a stance act that is enacted by an individual, but through social interaction it becomes contextualized as a *we* act. According to the stance triangle model (Du Bois 2007), conversational partners calibrate the relationship between their respective stances, undertaking a simultaneous commitment to moral evaluation of an object; ethical positioning of themselves; and affiliative alignment (convergent or divergent) with their interlocutors (Du Bois 2007; 2014; Du Bois et al. 2014). When a shared stance object commands our joint attention (Tomasello 1988), it invites not only my expressed evaluation but yours, plus a third aspect which may be explicit or implicit: an assessment of the alignment between my stance and yours.

Children are attuned to stance alignment early, as suggested by toddlers' stances framed in dialogic resonance (Köymen & Kyratzis 2014). Stance alignment continues to play a pivotal role in the interactional construction of moral and social identity through childhood, adolescence, and adulthood (de León 2019; Goodwin 2006; Stivers et al. 2011), in both typically developing children and children with autism (Du Bois et al. 2014; Sterponi 2004).

Although humans surely feel themselves subject to the compelling force of moral obligations, they are also exquisitely attuned to conflicting norms, which may compete for their allegiance within and across groups. The possibility of negotiating which obligation currently carries the greater moral imperative, given the complexities and ambiguities of everyday social life, may lead people to invoke, compare, reject, or adopt one moral principle over another, ultimately resolving (or not) the vexed question of what is the right thing to do. In the target article, Tomasello provides striking evidence that moral obligations emerge by early childhood and develop across the school years. But we suggest it is important to consider a broader developmental time course and to come to terms with the cognitive and affective affordances, as well as interactional practices, that scaffold the emergence of an increasingly nuanced attunement to moral obligation within a social framing. Humans are an incredibly social species: Our drive to cooperate and take stances that convey alignment with our own social group may be evidenced across the lifespan. Future research will benefit by bringing together converging evidence from across the lifespan, drawing from naturally occurring interactions and experiments, to clarify why, how, and with what consequences humans manage the sociality of moral obligation.

The sense of moral obligation facilitates information agency and culture

Heather M. Maranges^{a,b} , Roy F. Baumeister^c
and Kathleen D. Vohs^d

^aDepartment of Psychology, Florida State University, Tallahassee, FL 32306; School of Psychology; ^bConcordia University, Montreal, QC H4B 1R6, Canada; ^cUniversity of Queensland, St Lucia QLD 407, Australia; and ^dCarlson School of Management, University of Minnesota, Minneapolis, MN 55455.

maranges@psy.fsu.edu www.heathermaranges.com
r.baumeister@psy.uq.edu.au <http://www.roybaumeister.com>
vohsx005@umn.edu <https://carlsonschool.umn.edu/faculty/kathleen-vohs>

doi:10.1017/S0140525X19002334, e78

Abstract

Tomasello argues that humans' sense of moral obligation emerges early in development, relies on a shared "we," and serves as the foundation of cooperation. This perspective complements our theoretical view of the human self as *information agent*. The shared "we" promotes not only proximal cooperative goals but also distal ones via the construction of shared understanding – it promotes culture.

Tomasello provides an elegant account of humans' unique sense of moral obligation. He argues that this sense of moral obligation begins early in human development, features a shared "we," and facilitates cooperation. He cites work demonstrating that human children, but not other apes, feel and act on a special sense of obligation to their cooperative partners, in-group, and cultural group – they share resources fairly and regulate their partners and themselves according to the implicit or explicit agreements of the group. This analysis provides a potentially powerful explanation to some vital phenomena about human selfhood and sociality. This comment elaborates that connection.

Although the beginnings of culture may be found in several dozen species (de Waal 2008), it is only among humankind that culture is extensively developed so as to permeate almost every aspect of life (Baumeister 2005). A major part of culture is shared knowledge and understandings. Bourdieu (1977), a French sociologist, coined the term *doxa* to refer to "that which goes without saying" – the shared assumptions that enable human beings, even strangers, to interact cooperatively for mutual benefit. Two people who share an extensive *doxa* can begin interacting efficiently without having to explain everything from the ground up. We have argued that theories of the human self need to include a recognition that human selves are *information agents*, or builders of the *doxa* (Baumeister et al. 2018).

The early-emerging sense of "we" and moral obligation are necessary for information agency. Humans have special motivations and abilities to build the *doxa*: They not only collect information (as other curious animals do) but proactively share it with group members, critique it, pass along information from others, emphasize shared understanding even over possibly more accurate private knowledge, and more. For example, people, but not other animals, sacrifice valuable resources to share information that may be useful to others (e.g., Feinberg et al. 2012; Tamir & Mitchell 2012), add or subtract details to gossip to emphasize what is of most importance to the group (e.g., DiFonzo & Bordia 2007), and as individuals or institutions, systematically pass on culturally relevant information (i.e., they teach; Csibra & Gergely 2009).

Based on Tomasello's analysis, we speculate that moral obligation was among the earliest sources of motivation to identify other minds that might require the information one already has. In particular, obligation as collaborative self-regulation, which entails group-mindedness and appreciation of social norms, encourages people to share information broadly. Broadly shared knowledge enables group members to access information that they may not have personally gathered, to act on the same information and beliefs, and, in turn, to cooperate more effectively. The extensive benefits from ongoing cooperation would be a powerful selection factor in favor of minds that could manage information agency. Tomasello's work showed that this obligation is sharply reduced with out-group members and even with in-group members who have not collaborated in the current project. The latter understand why they are not invited to share even despite being part of the in-group.

Crucial to our analysis is not only the complementary understandings that one person owes another a reciprocal favor, but also the mutuality of that understanding. The sense of obligation is thus not just a first step toward improved cooperation on specific tasks. It helps lay a foundation for culture. As Tomasello says, it is not just that one person feels an obligation and the other feels a right: they both know and accept that the other has those feelings. Indeed, this soon moves beyond dyadic interaction, so that the *doxa* belongs to the group. In other words, not only does

the individual feel an obligation – the individual knows that everyone in the group understands the obligation and will react negatively if the individual refuses to honor it. Indeed, someone who shirks such an obligation will be known in the group as selfish and uncooperative. This will be discussed first, as gossip relates the nonreciprocation to the shared understanding of the norm, and then the person's selfishness will become part of the *doxa*. Everyone would know the person is selfish and uncooperative. And this could prove fatal.

In many animals, first interactions with strangers are occasions for distrust, hostility, and even aggression. Humans are different. Studies with the trust game have shown that many people will make cooperative overtures to strangers (although preferably those from the same culture), even at risk to themselves. The reason appears not to be a strong expectation that one's trust will be rewarded but rather a feeling that one is obligated to show respect for the other's integrity, even if one privately doubts it (Dunning et al. 2014) or judges the stranger as having low self-control and therefore poor moral character (Maranges & Ainsworth 2020). If the other does not cooperate too, then the interaction can quickly turn antagonistic; but starting off with a brief, tentative willingness to make oneself vulnerable so as to offer a collaborative partnership is a huge step toward human culture. This step likely rests on a deep sense of moral obligation to other group members.

Tomasello's view helps explain the psychological foundation of information agency – the shared "we" serves as a superordinate schema that directs the self toward information beneficial to the group and subsequent feelings of obligation beget motivation to glean, edit, and share information with other group members. Indeed, this builds on the analysis of the human self as an adaptation to enable the animal body to participate in a cultural group (Baumeister et al. 2016). In other words, building and maintaining the *doxa* is a basic purpose of the human self. It is itself a cooperative enterprise. The collective understanding of the moral psychology of obligation was probably a key step in this.

Obligations without cooperation

Julia Marshall 

Department of Psychology, Yale University, New Haven, CT 06511.
julia.marshall@yale.edu juliaannemarshall.com

doi:10.1017/S0140525X19002565, e79

Abstract

Our sense of obligation is evident outside of joint collaborative activities. Most notably, children and adults recognize that parents are obligated to care for and love their children. This is presumably not because we think parents view their children as worthy cooperative partners, but because special obligations and duties are inherent in certain relational dynamics, namely the parent-child relationship.

Tomasello contends that our sense of obligation is instantiated through "agreement-like" social interactions. For example, if I promise a co-worker that I will attend her birthday party, I commit myself to do so; and I would need to apologize or justify myself if I failed to attend the birthday festivities. Tomasello

further argues that, from a developmental perspective, this understanding of obligation is first evident in children “inside, but not outside, of collaborative activities structured by joint agency with a partner.”

I agree that obligations are often established through mutual cooperative agreements. But, I disagree that (1) obligations arise exclusively out of such arrangements and that (2) children first feel the demand of an obligation only in the context of activities structured by joint agency. Instead, it seems that some obligations are inherent in certain types of *social relationships* – namely, our kinship relationships – and, importantly, that children recognize this from a young age.

Consider a father with a sick baby girl. The way he feels toward his daughter fits into Tomasello’s conception of obligation in some ways but not in others. He likely feels a demanding and coercive force to care for his sick daughter, and he would be judged harshly if he decided to leave his daughter and go on a vacation (Bloom 2011; 2013; Gopnik 2009). Importantly though, a baby is not a cooperative partner with her father in any meaningful sense; they are not attempting to accomplish a joint goal. Instead, the parent is obligated to their child because the parent-child relationship inherently requires care and love. And, this holds true even when the baby grows up. Consider the case where a child grows up severely brain damaged, which renders them unable to presume the role of collaborative partner; a parent is still obligated to care for their child. Furthermore, extreme cases aside: Most parent-child relationships also purposefully lack a joint collaborative element and are instead fundamentally grounded in unconditional care and love (Clark & Mills 1993).

Social psychological and developmental research reveals that children and adults understand that obligations exist not only between collaborative partners but also between kin. For one, adults evaluate unhelpful parents more harshly than unhelpful friends followed by unhelpful strangers (Haidt & Baron 1996), as do older children and adults (Marshall et al., *in press*). These effects emerge presumably because parents are more obligated to their children and are, in turn, evaluated more negatively for failing to help them. In line with this interpretation, cross-cultural work reveals that children at 7 years of age and older in both the United States and India as well as adults in both countries consider parents as obligated to their children (Miller et al. 1990) – parents have to help their children even if they do not want to and even if no law existed requiring parents to take care of their children.

I along with my collaborators investigated how children even younger than 7 reason about parent-child obligations. In these studies (Marshall et al., *in preparation*), we were interested in whether younger children – like adults and older children – consider parents to be obligated to their children. We tested children not only in the United States but also in India, Germany, Japan, and Uganda to assess whether children vary in their sense of obligation across different cultures. We presented children as young as 5 and as old as 10 with scenarios wherein a child needs help, either because they were very hungry and did not have food or because they fell and hurt themselves at a park. Importantly, these scenarios did not involve any joint collaboration or commitment; neither the parent nor the child were working toward a shared goal and had not agreed to work together to accomplish a collaborative aim.

Consequently, we asked participants to determine whether the parent of the child in need *has* to help the needy child. We measured obligations in this way because researchers have established that asking children about what people have to do captures children’s sense of deontological obligation as distinct from children’s

sense of what usually occurs (i.e., frequency) or of what people want to occur (i.e., preference; Kalish & Lawson 2008). The results of these studies rendered a clear pattern: Children and adults alike, regardless of age and culture, recognize that parents are obligated to their kin. That is, all participants indicated that a parent has to help their child. These studies, to us, demonstrate that children as young as 5 do not think obligations are restricted to contexts involving joint collaboration, but rather that obligations are present within the parent-child relationship even absent a mutual goal.

Nonetheless, these studies only involved children as young as 5; Tomasello notes that even toddlers – children around the age of 3 – recognize the normative force of joint commitments. Whether children younger than 5 also recognize the normative force of kinship relationships is unknown. We suspect that toddlers would, but it remains an open question for future research. Regardless, the research discussed here, to us, provides compelling evidence that children at least around the early elementary school years appreciate the demanding quality of certain types of relationships. This is not because children consider parents and children as cooperative partners but because obligations are instantiated within certain relational dynamics.

The nature of obligation’s special force

David Olbrich 

Department of Philosophy, University College London, London, WC1E 6BT, United Kingdom.
davidolbrich@gmail.com

doi:10.1017/S0140525X19002450, e80

Abstract

Tomasello’s characterization of obligation as demanding and coercive is not an implication of the centrality of collaborative commitment. Not only is this characterization contentious, it appears to be falsified in some cases of personal conviction. The theory would be strengthened if the nature of obligation’s force and collaborative commitment were directly linked, possibly through Tomasello’s notions of identity and identification.

It is a consensus among philosophers that moral obligation possesses some kind of special force in deliberation. Tomasello’s interpretation of obligation ascribes three additional characteristics: He suggests that it is *peremptory*, that it is *demanding*, and that it is *coercive* or at least has a “kind of coercive quality” (target article’s introduction, para. 1, item 1 [“Special Force”]). The idea of peremptoriness is not mentioned again in this connection and could have been explored further. But the rest of the interpretation will be contentious because there may be ways of apprehending obligation that do not represent it as demanding, peremptory or coercive. The suggestion that the conjunction of these traits forms a key explanandum within the theory of obligation will therefore attract the objection that Tomasello’s theory of obligation is too restrictive, since it is designed to deliver the result that obligation is always all three.

Room for objection here is partly due to the possibility of alternative theories of obligation's special force. That obligation is always felt to be demanding or coercive is not a direct implication, for instance, of either the following theories: that a) obligation (perhaps in just one sense of "obligation") is what one has decisive moral reason to do (Parfit 2011), or that b) obligation is just what one can be intelligibly blamed for violating or what one might intelligibly feel guilt for violating (Adams 2002).

On the other hand, some cases of personal conviction provide apparent counterexamples to the claim that felt obligation is always demanding and quasi-coercive. Not everything that it feels necessary to do for moral reasons is felt to be coerced or demanded from one. Williams (1981; 1993) illustrated how the performance of some obligations is felt to be expressive of character-constituting dispositions, so that someone refusing to betray their spouse may discover that they are literally unable to betray them and that that is a fact about their character. The fact that they *would* feel the obligation as a demanding one were they to be reluctant to discharge it does not show that, absent that reluctance, the obligation is felt to be demanding. Similarly, Tomasello's suggestion that a violation of obligation would disrupt one's identity focuses too much on the agent's relations to violation rather than successful performance; identity has affirmative aspects too, and it may be that the performance of obligation is felt necessary because that is who the agent feels they are. This is not an instance of feeling coerced.

There are two elements of Tomasello's theory on which this issue will generate pressure. The first is Tomasello's idea, drawn from Darwall (2006), that apprehension of obligation involves the agent in an awareness of the demands placed on them by others who have the authority to make those demands. If our obligations are demanded of us, then the conclusion is inescapable that they will be felt to be demanding; conversely, a more expansive conception of how obligation may be felt will require an abandonment of the insistence that obligations are authoritative demands. Second, there is the proposal that the coerciveness of obligation originates in the existence of an implicit "threat – to my identity" (sect. 3, para. 6) from the partners who may reject me, breaking up the "we" which is part of my identity, if I do not play my part in the collaborative commitment that defines us. But even if such a threat is inescapable and omnipresent, it does not follow that it should exhaust every aspect of the special force of obligation. A noncoercive, non-demanding special force will require an independent explanation.

Thus this issue regarding a slight expansion in one explanandum touches on deep aspects of the theory. To the extent that Tomasello's theory is offered as a supplement to Darwall's theory of obligation's demandingness, explaining what the authority to make demands consists in, and how demands come to be seen as legitimate, these problems will loom large since they require a revision or extension of that very theory which is being supplemented. But the idea of the collaborative commitment that Tomasello contributes might alternatively bear development outside that heavy theoretical machinery.

The prospects for such a development are promising. The existence of a collaborative commitment does not directly entail the involvement of demands, directed by agents in that joint commitment on other agents within it, that they carry out that commitment. Rather, the issuing of demands, as it appears in the experiments Tomasello cites, evidently serves a regulatory purpose: agents who stray from what the collaborative commitment requires of them are brought back to the required behaviour. Even if the implicit issuing of demands is involved in all actual collaborative

commitment behaviour, the connection here is contingent and not constitutive. The securing of compliance through the coercion of an implied threat of deprivation of group membership, too, acts as a second-order compliance mechanism. On some occasions these mechanisms may be neutralized or weakened; although such situations may disrupt the ability to ensure compliance with the commitment, the inoperativeness of the compliance mechanisms would not render the shared commitment itself inert. When this potential complexity in the relation between commitment, demand, and threat is registered, room is made for a more flexible treatment of the various, possibly heterogenous, ways in which obligations may figure within deliberation and be felt emotionally.

The basic phenomenon, shared commitment, may make its own contribution to the way in which the obligations of that commitment are felt. Distinctive in Tomasello's posited we > me motivation is the notion of a commitment that agents themselves identify with insofar as they share in the collaborative project that constitutes the relevant aspect of in-group membership. If an agent identifies with this commitment deeply enough, they will *ipso facto* be motivated to discharge it, but the fact that the commitment originated externally will make a difference to the kind of force the obligation has for them. That it is possible to develop the idea of commitment without recourse to the notion of demand shows that Tomasello's theory is not quite as close to Darwall's theory as he thinks it is.

A hard choice for Tomasello

Philip Pettit^{a,b} 

^aUniversity Center for Human Values, Princeton University, Princeton, NJ 08544 and ^bSchool of Philosophy, Australian National University, Canberra, ACT 0200, Australia.

ppettit@princeton.edu <http://www.princeton.edu/~ppettit>

doi:10.1017/S0140525X19002346, e81

Abstract

Michael Tomasello explains the human sense of obligation by the role it plays in negotiating practices of acting jointly and the commitments they underwrite. He draws in his work on two models of joint action, one from Michael Bratman, the other from Margaret Gilbert. But Bratman's makes the explanation too difficult to succeed, and Gilbert's makes it too easy.

Michael Tomasello begins his engaging target article by marking two features of the human sense of obligation. First, it has peremptory force: "Unlike the most basic human motivations, which are carrots, obligation is a stick" (Introduction, para. 1, item 1 [Special Forces]) And second, it is socially embedded: 'Breaches of obligations often prompt normative protest, from the offended party, and apologies, excuses, and justifications, from the offender' (Introduction, para. 1, item 2 [Special Social Structure]).

I agree that obligations, at least those we associate with morality, have these features. Together, they mean that when we face a conflict between inclination and obligation, going with the inclination – giving in to it – counts as a sort of failure, leading to "a sense of guilt"; in particular, a failure for which others will hold us responsible and which we will be expected to explain away or apologize for.

Tomasello postulates that selectional pressures have elicited a capacity among humans to pursue goals jointly: a “species-unique ... preference ... for pursuing goals by collaborating with others” (sect. 2.1.2.1, para. 1). And he thinks that this generates a feeling among them “of obligation or responsibility to treat their partner respectfully and fairly (sect. 2.1.3, para. 1). He documents this idea in evolutionary and developmental terms, showing that even toddlers feel “the normative force of joint commitments to collaborate” (sect.2.1.2.1, para. 3).

The account sketched is not only evolutionary and developmental, it is meant to operate at the level of both interpersonal and society-wide morality. My remarks here bear most directly on the evolutionary story at the interpersonal level. They rehearse a difficulty I have elaborated elsewhere (Pettit 2018b), on the basis of a distinct, language-dependent story about how morality could have emerged among creatures like our forebears in Erewhon (an anagram of “nowhere”) (Pettit 2018a).

In expanding on the fact that we are a collaborative, interdependent species, Tomasello draws on the literature of joint action, in particular the work of Michael Bratman (2014) and Margaret Gilbert (2014). But these philosophers have presented two very different models of joint action, as indeed Tomasello (2016a) has noted elsewhere, and I see a problem for his derivation of the human sense of obligation, if he starts from either. Hence my claim that he faces a hard choice.

Bratman offered a normatively austere account of how joint action may materialize. According to him, all that may be needed, roughly, is that it is a matter of common assumption among two or more agents that they each desire a certain result – say, saving a child in the water; that there is a certain plan whereby they achieve that result together rather than on their own – this may involve forming a chain into the water; and that if they play their part in that plan, then others will play theirs.

Gilbert offered an alternative under which this is not sufficient, and may not even be necessary. According to the alternative, what is also needed, roughly, is that each at a crucial point recognizes that there is a “joint commitment” among the parties to carry out the plan, where this shows up in the fact that they will complain to and about any defector.

Gilbert’s account is normatively richer than Bratman’s, because it presupposes that the people involved in the joint action have mastered a normative concept: that of a joint commitment. This concept is normative insofar as it is already tied up with the idea of obligation and responsibility; as she presents it, people can understand what a joint commitment is only insofar as they have an idea of what it is to hold someone responsible.

If Tomasello goes with Bratman’s model of joint action, then it is not clear to me how he can get a sense of obligation out of it (Pettit 2018b). He argues plausibly that the parties to joint activity, even understood on this model, will each try to choose and control partners so as to maximize the chance of success, recognizing the traits that make a partner attractive; and that their reliance on joint action will lead, across a variety of contexts, to a pattern of partner assessment and selection, reform and rejection. But why should this introduce a sense of obligation among participants?

As I view things, such joint activity might have evolved, at least in a basic form, under familiar mechanisms of reciprocity illustrated by the tit-for-tat discipline (Axelrod 1984; Nowak 2006). Thus, I see no convincing reason why “early humans,” in view of having developed “a unique form of partner control” (Sect. 2.1.1, para. 2) – and even a unique sense of a joint “we” – should have begun to hold one another responsible for failures,

protesting normatively “against non-cooperative behavior” and giving “the non-cooperator a chance to mend her ways voluntarily of her own accord” (Sect. 2.1.1, para. 2).

Would Gilbert’s account serve Tomasello better? Yes, but without offering an explanation of how early humans could have progressed from a stage where they lacked normative concepts, in particular that of obligation, to one at which they developed them. Her model puts the normative in place at the beginning so that it leaves us unsurprised, but also unenlightened, by the fact that it also appears at the end.

I am persuaded by Tomasello that we have inherited distinctively collaborative dispositions from early humans, and that even children articulate them normatively. But I think that the dispositions might have operated in early humans without normative conceptualization and that something else is needed to explain how we came to think in a normative fashion and to regulate our interactions normatively.

How is the moral stance related to the intentional stance and group thinking?

Hannes Rakoczy 

Department of Developmental Psychology, Institute of Psychology, D- 37073 Göttingen, Germany.

hannes.rakoczy@psych.uni-goettingen.de

https://www.psych.uni-goettingen.de/en/development?set_language=en

doi:10.1017/S0140525X19002413, e82

Abstract

The natural history of our moral stance told here in this commentary reveals the close nexus of morality and basic social-cognitive capacities. Big mysteries about morality thus transform into smaller and more manageable ones. Here, I raise questions regarding the conceptual, ontogenetic, and evolutionary relations of the moral stance to the intentional and group stances and to shared intentionality.

The target article tells a natural history of our moral stance. Although philosophers have long been analyzing the conceptual peculiarities of our specifically moral engagements with each other, our reactive attitudes, second-person feelings, and the like, Tomasello’s project aims at tracing their natural foundations. Building on decades of research in developmental and comparative psychology, Tomasello closely ties the phylogeny and ontogeny of the moral stance to more basic social-cognitive capacities – in particular the intentional stance, shared intentionality, and group thinking. This project certainly presents massive progress in naturalizing morality.

Here, I simply ask how, exactly, the moral stance relates to the intentional stance, to shared intentionality and to group thinking. Clearly, the four stances (intentional stance, shared intentionality, group stance, moral stance) are very closely related, and one of the main merits of the target article is to highlight this very nexus. At places, however, the target article reads as if the links were so close that the stances almost coincide: “This creates the most basic [...] 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” (sect. 2.2.1, para. 3). This passage thus sounds like there may really be only one unitary distinction in our social cognition: between rational agents (subject to the intentional stance) with whom one can enter into shared intentionality, that are members of one’s tribe (group stance) here and subject to the moral stance; and the rest (“barbarians”) there, who are neither rational nor members of one’s group, partners for shared intentionality nor morally relevant agents.

But the real relations between the four stances are probably much more complex: That some basic form of the intentional stance and shared intentionality can come apart is exemplified by the Machiavellian intelligence of nonhuman great apes (who, according to Tomasello, understand each other as rational agents in some basic sense without ever entering into *shared* intentionality) and (perhaps) human psychopaths. The relation between shared intentionality and the group stance is complicated by the fact that there is no such thing as the one and only way to categorize groups. Rather, social group concepts work in complex, crisscrossing, and partly hierarchical ways. In a soccer game, for example, “we” (the back four) play our part so that “we” (the 11 players of our team) succeed against “them,” but a different “we” (the 22 of us on the pitch) are playing this match according to conventional rules that yet another “we” (the community of football players) follow. Each of these hierarchically structured group categorizations go along with some forms of shared intentionality, but arguably very different ones ranging from very concrete and action-based to abstract and far-reaching in temporal and spatial scope. Tomasello’s bi-partition into dyadic and group-based shared intentionality is an important step towards recognizing this plurality of shared intentionalities, but it may not be fine-grained enough.

This becomes even clearer when considering the relations between group thinking, shared intentionality, and the moral stance, or normative stances more generally. Some normative stances may be intimately linked, conceptually, ontogenetically and evolutionarily, with more or less small-scale, tribe-like group conceptions. This pertains, roughly speaking, to local, conventional customs that solve coordination problems. Such customs are arbitrary: they are such and so (driving on the right; putting the knife right of the plate), but could have been different (driving and knives left). Correspondingly, they are context-relative in a strong sense: Since they derive their existence from the shared intentional practices of the more or less small-scale group, they are valid only in the context of the group’s activities; and if the group changes its intentions, agreements, or decisions, the conventions change.

The moral stance as we know it, in contrast, seems not to be contingent in the same ways on the concrete shared intentionality of local, small-scale group practices. If it depends on group categorizations and corresponding shared intentionality at all, these are located at much higher levels of abstractions, pertaining to the class of rational agents, potential interlocutors, or the like. The moral stances derive their status from our sharing a nature of rational, sentient beings rather than from sharing an intention to perform this or that concrete collective activity (Nagel 1970a).


Quite young children seem to be aware of this fundamental difference between rational and moral versus conventional normative stances to some degree: They understand that moral norms pertaining to human well-being and justice, for example, lack the relativity to local, group-based customs and shared intentionality typical of arbitrary conventions (Turiel 1983). And they enforce norms of rationality and moral norms vis-à-vis other

agents irrespective of their group status, whereas they hold only in-group members responsible with respect to conventional norms (Schmidt et al. 2012).

In general, group categorization and the corresponding forms of shared intentionality range from the concrete, action-based, transient, and very local matters of, say, a pair of dancers to the maximally abstract category of rational sentient beings (which, depending on technology and ideology may include [some] animals and robots besides humans). The normative stances related to these different levels and layers of groups and shared intentionalities differ in fundamental ways. It will be crucial for future research to better understand how, exactly, group stances, the intentional stance, shared intentionality and the moral stance, relate to each other developmentally. This would be enlightening theoretically, but may also be important politically: In light of threats by new moral tribalisms around the globe, such a better understanding may help to secure moral enlightenment and progress with its ever-widening scope. As Sellars emphasized long ago:

To think of a featherless biped as a person is to construe its behaviour in terms of actual or potential membership in an embracing group each member of which thinks of itself as a member of the group. Let us call such a group a ‘community’. Once the primitive tribe, it is currently (almost) the ‘brotherhood’ of man, and is potentially the ‘republic’ of rational beings (cf. Kant’s ‘Kingdom of Ends’).” (Sellars 1962, p. 76f.)

Caregiving relationships as evolutionary and developmental bases of obligation

Rachna B. Reddy^a and Henry M. Wellman^b 

^aDepartment of Anthropology, University of Michigan, Ann Arbor, MI 48109 and

^bDepartment of Psychology, University of Michigan, Ann Arbor, MI 48109.

rachnare@umich.edu <https://sites.lsa.umich.edu/rachnareddy/>

hmw@umich.edu

doi:10.1017/S0140525X19002504, e83

Abstract

Obligation as defined by Tomasello requires mutually capable parties, but one-sided caregiver relationships reveal its developmental and evolutionary precursors. Specifically, “coercive” emotions may prompt protective action by caregivers toward infant primates, and infants show distress toward caregivers when they appear to violate expectations in their relationships. We argue that these early social-relational expectations and emotions may form the base of obligation.

In this target article, Tomasello tackles the internalized motivations that he believes drive the morality of a collective “we” that is foundational to our distinctly cooperative and moral species. He argues for a human-specialized feeling and cognitive understanding of obligation that is “coercive.” It drives us like a “stick” rather than a “carrot.” The emotions that surround it, such as guilt and shame, are reflective ones.

We agree that obligations and the emotions surrounding them are (1) understudied yet importantly linked to the prosocial,

group-serving proclivities of our species, (2) early developing in children, and (3) from current evidence seemingly absent in our closest primate relatives. We also agree that child developmental and primate-comparative findings can be used to support evolutionary hypotheses. But we think Tomasello has overlooked a crucial place to examine evidence for the evolution and development of obligation. Specifically, he writes that obligative forces arise only between individuals who can hold one another mutually accountable: thus, human infants are excluded from obligatory “contracts.” No doubt mutual accountability is foundational to obligatory understandings. But, *one-sided caregiver* relationships may be precisely where the emotional and cognitive precursors to obligation have their evolutionary and ontogenetic starts.

In species that care for their young, there are few forces more emotionally “coercive” than an infant in danger. Even non-parents may see to their needs. For example, female ring-tailed lemurs permit infants who are not theirs to leap onto their backs when a raptor flies overhead (Gould 1992), colobus monkey males clutch infants to their chests for safety when chimpanzees hunt them (Reddy, personal observations, 2019), and teenaged chimpanzees may whimper and search for their orphaned younger siblings if they fall behind the group, risking becoming lost themselves (Reddy & Mitani 2019).

The preceding behaviors do not necessarily reveal the sort of action and feelings that Tomasello emphasizes as “mature” signs of obligation. No current evidence exists, for example, that non-human primates of any age ruminate upon their guilt if their vulnerable charges (or their “mutually accountable” friends) are killed when they could have acted (despite exhibiting grief-like symptoms upon the deaths of close companions, e.g., Engh et al. 2006). Yet, the swiftness with which they act protectively, and the visible distress they show when others are in danger indicate the probability of coercive emotions that prompt these actions – emotions that could provide evolutionary and developmental foundations to a more reflective emotion like guilt and the actions it prompts (e.g., apology).

Moreover, although young human infants (and nonhuman primates) may not understand “terms” of obligation in full force, they do appear to have expectations about their social relationships and show distinct responses in situations when those expectations are violated. For example, when adults do not smile at them in the classic “still face” paradigm, human infants cry. From 8–10 months of age they also show distress and act proactively if they witness moral violations or unfairness or experience it themselves (Hamlin 2013).

When their mothers attempt to wean them, young chimpanzees throw “tantrums,” screaming, wailing, and occasionally hitting and biting their mothers. When young humans don’t get their way, they throw tantrums too, as characteristic of the “terrible twos.” The distress shown by infants of both species toward their caregivers may reflect their displeasure at the experience alone (e.g., hunger) and their knowledge that their caregiver will tolerate such an outburst. However, tantrums may also reflect a proto-form of emotional “outrage” that infants have when their caregivers behave differently from “usual” within their relationship. How caregivers respond in these situations is further interesting. Chimpanzee mothers attempt to ignore their infants’ tantrums, but occasionally appear to become distressed and frustrated themselves: when their infants hit them repeatedly on the back, for example, mothers may turn around, grab infants by the arms and shake them. Critically, however, mothers often then swiftly pull infants into tight embraces (Van Lawick-Goodall 1968;

Reddy, unpublished data, 2019). The human parallels are obvious: parents often report guilt about how angry they become in the face of child tantrums. We do not suggest that anything like “guilt” and “apology” motivate these “reconciliatory” actions in chimpanzees (cf. de Waal & van Roosmalen 1979), but chimpanzee mothers do follow their aggressive actions with a comforting one that ameliorates their infants’ distress.

Furthermore, the emotions and expectations human infants and non-human primates have about their own caregiving relationships suggest that it is some quality of social bonds themselves, not collaborative capability or willingness of the individuals alone, that provide the basis for proto-obligatory “terms.” Consider one of Tomasello’s own examples: Chimpanzees do not share meat with all others who attempted to hunt, but they do share with and beg from friends who are their allies over others (Mitani et al. 2000). They continue to share meat when one of the friends has grown old and weak and is no longer a competent collaborative partner. Furthermore, at older ages (at least) both nonhuman primates and human children show emotion and form social expectations not only about their own relationships, but also about the relationships of third parties (e.g., Cheney & Seyfarth 1980; Rhodes & Wellman 2017).

Essentially, we propose that proto-obligatory motivations and expectations may begin with, “X helps me because X is my mother/I must help X because X is my offspring.” As individuals in group-living species grow up, this understanding may develop into “because X is my friend” and “because X is Y’s daughter.” In our species, children (from at least age 3 onward) may develop obligatory feelings and understanding as Tomasello describes of a more self-reflective nature: “I *must* help X.” It is nearly impossible to disentangle the preceding motivations in natural contexts, suggesting the need for work that will integrate “natural” observations of relationships with experimental studies. For children, how and when do actions and feelings of obligation arise when the other person is a caregiver (e.g., mother)? For chimpanzees and children, students of Tomasello have already begun related work, finding that friendships influence helping propensities in both chimpanzees and human children (Engelmann et al. 2019).

Does the concept of obligation develop from the inside-out or outside-in?

Marjorie Rhodes 

Department of Psychology, New York University, New York, NY 10003.
marjorie.rhodes@nyu.edu <http://kidconcepts.org>

doi:10.1017/S0140525X19002590, e84

Abstract

Tomasello proposes that the concept of obligation develops “from the inside-out”: emerging first in experiences of shared agency and generalizing outward to shape children’s broader understanding. Here I consider that *obligation* may also develop “from the outside-in,” emerging as a domain-specific instantiation of a more general conceptual bias to expect categories to prescribe how their members are supposed to behave.

Tomasello proposes that the concept of *obligation* emerges in a manner that I will call, “from the inside-out.” A sense of obligation emerges first in children’s experiences of shared agency – where they feel a sense of obligation to their collaborative partners – and is then generalized outward to develop a more abstract sense of obligation among group members. On this account, children begin to feel obligated to their group members and to see their group members as obligated to them – simply because they are in the same pack – by around their third birthday.

Here I consider that the concept of *obligation* might also (or alternately) develop “from the outside-in.” That is, that children have abstract intuitions about social structure that develop separately from their own interpersonal interactions.

Already by age 3, and perhaps earlier in infancy (Ting et al. 2019), children have an abstract understanding that social groups specify who is obligated to one another and who is not (Chalik & Rhodes, *forthcoming*). We know that this understanding is already abstract – not tied to children’s own perspective – because children rely on it to guide their social understanding even when they are not members of the relevant groups.

To illustrate, in Rhodes (2012; also described in the target article, sect. 2.2.2.3, para. 2) children were introduced to two made-up social groups (called “flurps” and “zazes”); children had never heard of these groups before and were not assigned to be members of either group. In these studies, by age 3, children predicted that a flurp, for example, would harm a zaz, rather than another flurp. In Rhodes (2012), I speculated that these findings had to do with a sense of obligation – that children expected the flurp to avoid harming another flurp because, as group members, they were obligated to protect one another. Indeed, in Rhodes and Chalik (2013), children judged flurps to hold intrinsic obligations to one another but not to zazes.

If the concept of *obligation* that drove children’s responses in these studies developed from the inside-out, then we might expect children to solve these problems by placing themselves in the shoes of a flurp. That is, they might identify with the agent, and think of the other flurps as their in-group members. We might expect this to be particularly so around age 3, when children are just beginning to generalize from their own experiences outward. But this is not how children thought about these problems. When children are themselves put into made-up groups during experiments (e.g., Dunham et al. 2011), the first (and sometimes only) thing that happens is that they feel and respond more positively toward members of their own group (for review, see Dunham 2018). But children did not show any evidence of in-group positivity in these studies – when asked who a flurp would do something nice for, they responded at chance. Children’s responses in these third party scenarios – where they reliably expected flurps to harm zazes but to be equally nice to everyone (Chalik & Rhodes 2018; Rhodes 2012) – is exactly opposite to how children respond when they themselves are placed in made-up groups (where they are nicer to in-group members but not particularly mean to out-group members (Buttelmann & Boehm 2014; Dunham et al. 2011)). Thus, when children predicted that flurps would harm zazes, they relied on abstract intuitions about how group memberships specify social obligations, rather than their own first-person perspective.


From where would children get such an early developing abstract understanding of *obligation* if not by generalizing their own social experiences? One possibility is that children’s concept of obligation develops as a specific instantiation of a more general conceptual bias to treat categories as constraining what their members are supposed to do. This is a domain-general feature

of early concepts. For instance, children do not just think that cheetahs usually *do* run fast, they think they are *supposed to* (Haward et al. 2018; Foster-Hanson & Rhodes 2019). Children think there is something *wrong* with a category member who does not follow the norms of their group (Roberts et al. 2017) – and they hold this intuition just as strongly for categories of animals as for categories of people (Foster-Hanson et al. 2018). Children also think that the clearest and most informative example of an animal category is the one that best illustrates these prescribed properties, even if such an instance is rare (e.g., the very fastest cheetah in the world; Foster-Hanson & Rhodes 2019; Foster-Hanson et al. 2019). As another example, children also think that artifacts are supposed to fulfill their intended functions, and again, that there is something defective with one that does not (Diesendruck et al. 2003).

Thus, by early childhood, children have abstract, domain-general intuitions that categories prescribe how their members are supposed to be. In the social domain, this is instantiated as an expectation that categories constrain how people are supposed to treat one another, whereas in other domains it is instantiated in other ways (e.g., regarding how animals are supposed to get food, avoid predators, and so on). From this perspective, the general processes that drive conceptual development – in which children actively build hierarchical representations to make sense of various domains of experience (Gopnik & Wellman 2012) – can lead children to develop an abstract understanding that members of a group hold special obligations to one another, separately from their own experiences with shared or collective agency.

Of course, both types of developmental processes could simultaneously be at play – children could both develop a sense of their own obligations via their experiences with shared agency and begin to generalize them out, while at the same time, the mechanisms that underlie conceptual development lead them to construct an abstract understanding of the normative implications of group membership. A full developmental theory of how this critical concept arises would need to examine all of these processes and how they might relate to one another across early childhood development.

Feelings of obligation are valuations of signaling-mediated social payoffs

Amanda Rotella^a , Adam Maxwell Sparks^b
and Pat Barclay^a 

^aDepartment of Psychology, University of Guelph, Guelph, ON, N1G 2W1, Canada and ^bDepartment of Anthropology and Center for Behavior, Evolution and Culture, University of California, Los Angeles, CA 90095-1553.
arotella@uoguelph.ca <http://amandarotella.ca/research/>
adspar@fastmail.com <http://adammaxwellsparks.com/>
barclayp@uoguelph.ca <http://www.patbarclay.com/>

doi:10.1017/S0140525X19002322, e85

Abstract

We extend Tomasello’s framework by addressing the functional challenge of obligation. If the long-run social consequences of a decision are sufficiently costly, obligation motivates the actor to forgo potential immediate benefits in favor of long-term social

interests. Thus, obligation psychology balances the downstream socially-mediated payoffs from a decision. This perspective can predict when and why obligation will be experienced.

Tomasello advances a novel framework regarding the phylogeny and development of obligation, describing obligation as a coercive motivation. However, he did not explicitly define this motivation. To extend Tomasello's argument, we advance the following definition of obligation: *An obligation is a motivational sentiment in response to social expectations (e.g., requirements, commitments, taboos, rules) and potential social costs (e.g., loss of relationships, reputational consequences) which serves to motivate an individual to perform an action in a particular manner or to a level beyond what would maximize one's inclusive fitness if there were no social costs for acting otherwise.*

This definition adds to Tomasello's characterization of obligation by specifying that (1) there is a baseline willingness to perform said actions – for example, the material outcome is useful to me or to those I value based on kinship or interdependence or stake (see Aktipis et al. 2018, Balliet et al. 2017); (2) social obligations can make one willing to perform these actions differently or to a higher level than one might otherwise “want” (e.g., allocating resources differently than the level predicted from kinship, independence, and stake); (3) these obligations derive from perceived reciprocal or reputational consequences of (not) doing so (i.e., future costs and benefits); and (4) that the strategy of this system is at the functional level, and need not be consciously accessible.

For obligation to be adaptive, the mechanism must resolve a fitness problem. Here, it accounts for negative social consequences for not performing an expected action, or not performing it in the expected way. Performing an obligated action sends information – a positive social signal to an individual or group, whereas failing to perform the action sends the opposite signals (e.g., that one (de)values the relationship or membership, that one is (un)likely to default on existing relationships, or that one is (un)trustworthy). Given that reputation-based signals inform social decisions of others, such as partner choice, approach, and avoidance (Barclay 2013; 2016; Barclay & Willer 2007; Sylwester & Roberts 2010; 2013; Wu et al. 2016), (in)action could result in broad social consequences with long-term costs, such that someone (not) performing an action may be chosen less often as a social partner and receive less help from others. Consistent with this interpretation, multiple lines of evidence suggest that people calibrate their behaviors according to the perceived reputational costs and benefits of their actions (Barclay 2013; 2016; Barclay & Willer 2007; Feinberg et al. 2014; Rotella et al., in preparation; Wu et al. 2016). Notably, when one fails to complete the expected action, they may experience social emotions (e.g., guilt, shame) to motivate them to repair the relationship or diminish the reputational costs (Ketelaar & Au 2003; Schniter et al. 2012; Sznycer et al. 2016; 2018). This characterization of obligation posits that the intensity of the perceived obligation will correspond to the perceived social consequences from failing to meet the expectations of others.

We can formalize this at the functional level. There are multiple pathways to fitness outcomes (reviewed by Barclay & Van Vugt 2015), so there are multiple proximate motivation systems that cause our willingness to help. For example, one can have a kin-based interest and a reciprocal exchange interest in the same partner (e.g., “I love you brother, but this is a huge favor –

you'd better repay it”). Thus, one's total fitness interest and corresponding proximate willingness to help (h) is some cumulative function of fitness consequences derived based on kinship (k), interdependence/stake (i), demands from reciprocal partners (d), signaling value (s), and payoffs from various other social expectations (e), such that $h = k + i + d + s + e$. These various social outcomes (d , s , e) constitute one's obligation (o). If these motivations are additive, then one's obligation $o = d + s + e$. To generalize this to non-additive functions, one's willingness to help is some function (f) of these factors, such that $h = f(k, i, d, s, e)$. As such, one's obligation is the component of one's willingness which goes beyond the level directly predicted by kinship and interdependence/stake alone – that is, $o = f(k, i, d, s, e) - f(k, i)$. This model is consistent with the idea that as social demands are greater, stronger feelings of obligation will be experienced. Further, obligation will be perceived as a motivating force in situations where there is a conflict of interest, resulting in a proximate ambivalence when you “have to do” X but “want” to do Y .

Although joint intentionality often precedes obligation, our theorizing suggests that joint intentionality is not required to experience obligation; the feeling of obligation arises anytime that failure to complete the obligated action would result in negative social consequences. For example, dieting or raising pets and houseplants do not require joint intentionality. However, once these commitments are public knowledge (joint knowledge or expectations), one may feel obligated to persist because desisting would convey negative information about oneself.

In summary, we posit that obligation will be experienced when (1) there are learned social expectations (towards an individual or group) that, if failed, can result in far-reaching social consequences by impacting one's reputation; (2) these expectations implicitly or explicitly motivate people to act in a different manner or to a different degree than they would otherwise act; and (3) the function of moral obligation is to forgo short-term benefits likely to be associated with long-term social consequences. In the absence of reputational concerns (real or perceived), it is unlikely that feelings of obligation will be elicited. Our theorizing is consistent with the characteristics of obligation described by Tomasello, such that obligation is a special motivational force with a coercive quality which has a special social structure in human society, given the complexity of human social interactions. Thus, we extend his model by emphasizing the role of social repercussions – especially via signaling – in driving the experience of obligation, and that obligation is particularly salient when there is a conflict between one's immediate inclusive fitness interests and the potential downstream social consequences of deviating from others' expectations.

Who are “we”? Dealing with conflicting moral obligations

Alex Shaw^a and Shoham Choshen-Hillel^b

^aDepartment of Psychology, University of Chicago, Chicago, IL 60637 and
^bHebrew University of Jerusalem, School of Business Administration and the Center for the Study of Rationality, Mt. Scopus, Jerusalem 9190501.
 ashaw1@uchicago.edu <https://www.dibslab.uchicago.edu/>
 Shoham@huji.ac.il <https://shohamchoshen.wixsite.com/psych>

doi:10.1017/S0140525X19002577, e86

Abstract

Satisfying one's obligations is an important part of being human. However, people's obligations can often prescribe contradictory behaviors. Moral obligations conflict (loyalty vs. fairness), and so do obligations to different groups (country vs. family when one is called to war). We propose that a broader framework is needed to account for how people balance different social and moral obligations.

So many vows. They make you swear and swear. Defend the King, obey the King, obey your father, protect the innocent, defend the weak. But what if your father despises the King? What if the King massacres the innocent? It's too much. No matter what you do, you're forsaking one vow or another.

– Jamie Lannister, *Game of Thrones*. (Benioff & Weiss 2012).

Tomasello provides a detailed account suggesting that the sense of obligation is a central human motivation. According to this account, the sense of obligation arises out of the need to cooperate in order to survive and thrive in the social world. We agree that obligations are indeed an important part of being human. However, we highlight the notion that mutualistic cooperation is just one of many potential triggers of social obligations – people have obligations to myriad individuals, groups, and moral principles even in situations where they do not directly or profitably cooperate with those individuals. A key question left unanswered by Tomasello's account, we argue, is how people navigate such obligations when they conflict.

Our social world often confronts us with obligations that contradict one another. One prominent case is allegiances to different groups. Tomasello theorizes that obligation is based on a “we” experience with one's group. However, most individuals form “we” relationships with several different groups and subgroups: people are parents, siblings, friends, bosses, subordinates, citizens, and members of the human race (Turner et al. 1987). The obligations to all these groups often clash. For example, a man called to war may be torn between his duty to serve his country, and his direct responsibility for taking care of his family. People further feel a sense of obligation to their moral principles that can also conflict. People believe that it is important to avoid harm, but also that it is important to prevent others from dying – but what should one do when harming someone can lead to saving more lives (Greene 2008)? Be a loyal friend, and also treat everyone equally and fairly – but what if helping your friends means showing favoritism (Shaw et al., *in press*; Waytz et al. 2013)? Be honest, and be kind – but what do you do when a friend who just got a terrible haircut asks: “Does my hair look good?” (Levine & Schweitzer 2014). Research on adults suggests that people are able to balance these competing obligations in complex ways.

There is now a growing body of research demonstrating that such conflicts of obligations emerge early on in human development. This research suggests that the ability to resolve these conflicts may be an important part of children's early normative understanding. By the age of 6, children feel tensions between reciprocity and fairness; feeling conflicts about whether they should like someone who gave them more than others or someone who was fair and gave everyone the same (Shaw et al. 2012). Relatedly, children understand both the obligation of efficiency (i.e., maximizing social welfare), and the obligation to be fair and maintain

equity, and attempt to strike a balance between them (Choshen-Hillel et al., *in press*). Children are also able to navigate multiple concerns about avoiding harm and saving others when introduced with Trolley-like dilemmas (Levine et al. 2018). Further, children can balance concerns with honesty and benevolence, judging lies that are meant to benefit others as less immoral than selfish lies (Fu et al. 2015). Indeed, in some circumstances the cold truth might be seen as less kind and desirable than a polite lie. Taken together, these results reveal that navigating conflicting moral and social obligations is an important and basic part of existing and thriving in the social world.

Research reveals that there are many factors that determine which obligations people choose to follow. For example, one's goals will influence what obligations seem important: Fairness might be more important for being identified as a good leader whereas being loyal might be more important for being identified as a good friend (Everett et al. 2018). The social domain in question will also shape the obligations that seem important (Fiske 1992). When it comes to obligations toward notions of equality and merit, people may think that one should give rewards based on merit in domains such as pay for work, but prefer equality in domains such as voting – one vote per person. Cultures will also dramatically vary in the extent to which they emphasize moral values like honesty, fairness, kindness, and loyalty, leading to different balances (e.g., Fiske 1992; Miller & Bersoff 1992). Indeed, factors such as culture or socioeconomic status may also influence children's decisions (Choshen-Hillel et al., *in press*; Rochat et al. 2009). Finally, what might make matters even more complicated is that people's intuitions about conflicting obligations may change throughout the course of their development, with older children believing it is good to overcome moral conflict and younger children believing it is better not even to have moral conflict in the first place (Starmans & Bloom 2016). Future work should explore how cultural inputs shape both the kinds of obligations that children have and how they balance conflicting obligations as well as how these obligations may shift throughout the course of development.

In summary, we concur that obligations are important, but this is just a starting point for understanding the rich tapestry of interlacing obligations that human beings must traverse. Thus, we applaud Tomasello for bringing these obligations to the forefront and for outlining the importance of mutualistic cooperation in triggering social obligations. However, we think it is also time to acknowledge the wide swath of obligations that all people must face and the difficulties caused by having to navigate between obligations that are as contradictory as they are compelling. We thus propose that a wider framework is needed to account for how children and adults balance their different alliances and social obligations.

How does inequality affect our sense of moral obligation?

Walter Sinnott-Armstrong

Philosophy Department, Duke University, Durham, NC 27708.
ws66@duke.edu <https://sites.duke.edu/wsa/>

doi:10.1017/S0140525X19002310, e87

Abstract

Tomasello's novel and insightful theory of obligation explains why we sometimes sense an obligation to treat each other equally, but he has not yet explained why human morality also allows and enables much inequality in wealth and power. Ullman-Margalit's (1977) account of norms of partiality suggested a different source and kind of norms that might help to fill out Tomasello's picture.

Tomasello's new and illuminating theory accounts for many otherwise mysterious aspects of our human sense of obligation. In particular, humans (including children, but not apes) are motivated to share rewards equally with (and only with) partners in joint intentional collaborations, and they object to violations of this norm. He writes, "Most basically, as children participate in joint intentional collaboration, they see that: (1) both participants are equal causal forces in producing the mutually intended outcome; (2) both partners could switch roles as needed; and, most crucially, (3) 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" (sect. 2.1.3, para. 4). Later steps are then presented as "scaled-up versions of the joint commitments characteristic of joint intentional collaboration at the first step of our account" (sect. 2.2.2.2, para. 4).

This insightful picture does fit many interactions, including those in cited experiments that Tomasello designed, but it is still far from complete (no surprise!). As Tomasello notes (note 3), children also interact with adults whom they are expected to obey. This obligation to obey superiors does not seem to spring from collaboration among equals. Moreover, collaborators are not (1) "equal causal forces" when one works harder and produces more. We might sense an obligation to show equal respect of some kind to such collaborators, but they are usually not seen as deserving equal parts of the products of the collaboration. Furthermore, partners often cannot (2) "switch roles," such as when a quarterback (or a coach) cannot play lineman on an American football team. And even when two people can play the same role, the same standards do not always (3) "apply to anyone in that role." When teenagers volunteer to cook dinner, they are often not expected to cook as well as their parents. My point is not only that adults do not treat each other as equals when they misbehave as in bullying, domination, and racism. In addition, children – and especially adults – accept lots of inequality as morally permitted, desirable, or even obligatory, as in unequal power relations between bosses and employees, generals and privates, judges and defendants, police and citizens, parents and children, and even wealthy and poor citizens, in some cases. Equality is seen as obligatory at some levels but not others.

Tomasello might try to explain these inequalities in moral obligations in terms of collective agreements and commitments, which can lead to social norms and roles that are supposed to create special rights and obligations for some people that others in the same group do not possess equally. However, not all of the parties to these collective agreements are equal, and these roles often do not seem to arise from collaborations between equals. What collaboration between children and parents gives rise to the obligation of children to obey their parents? And are underprivileged groups really equal partners in the collective agreements that give police power over them? And are employee and employer equal when one accepts a job from the other or makes laws governing such contracts? Tomasello is right to

bring up social norms and roles, but it is not clear how they fit within his simple model of collaboration among equals or his "scaled-up" account of "objective" obligation.

This problem does not refute Tomasello's central claims. It only shows that he needs to add some explanation of inequalities in obligations. One promising way to supplement Tomasello's theory would draw upon Edna Ullman-Margalit's unfortunately neglected model of "norms of partiality" (Ullman-Margalit 1977). Very roughly, the basic idea is that some (not all) norms arise from negotiations among unequals in which privileged leaders use their greater power and wealth to solidify and extend their personal advantages, while underprivileged subordinates hold out for as much as they can get under the circumstances, even if that is far less than equality. This origin in inequality and cultural negotiation can explain certain features of those norms of partiality that are hard to explain in terms of collaboration among equals.

One example (not endorsed by Ullman-Margalit but suggested by others) tries to explain why obligations to help the needy (whether or not they are not partners in collaborations) are generally seen as weaker than obligations not to cause harm. Although some philosophers disagree, studies have shown that most people judge killing a beggar to be worse (and to violate a stronger obligation) than failing to help a beggar even when that means letting the beggar starve. This feature of our sense of obligation could be explained if it arises from negotiations between unequals. The rich and powerful discourage a strong sense of obligation to help the needy, because they have much to lose and little to gain from that obligation; but they encourage a strong sense of obligation not to harm, because they are vulnerable to being harmed. They can be killed just like a beggar, even if they have guards. And the preferences of the rich and powerful shape our intuitions about the relative strength of these obligations because the rich and powerful control news sources, churches, governments, and other institutions that shape the sense of obligation in most cultures.

I am not at all sure that this story is accurate or succeeds in accounting for this feature of our sense of obligation. My main point is only that Tomasello's theory is limited and incomplete until he adds some new element to account for unequal obligations and also for peculiar features of our sense of obligation that might have arisen from inequalities in power and wealth. Some story about negotiation among unequals (as Ullman-Margalit suggested) could fit that bill, although Tomasello might have a very different explanation. In any case, I look forward to hearing how he explains our sense of obligation in situations of inequality.

Cooperation and obligation in early parent-child relationships

Ross A. Thompson 

Department of Psychology, University of California, Davis, CA 95616-8686.
rathompson@ucdavis.edu <https://psychology.ucdavis.edu/people/rathom>

doi:10.1017/S0140525X19002474, e88

Abstract

Tomasello's moral psychology of obligation would be developmentally deepened by greater attention to early experiences of

cooperation and shared social agency between parents and infants, evolved to promote infant survival. They provide a foundation for developing understanding of the mutual obligations of close relationships that contribute (alongside peer experiences) to growing collaborative skills, fairness expectations, and fidelity to social norms.

Moral development theorists like Piaget (1932) and Kohlberg (1969) argued for the uniquely generative qualities of egalitarian peer relationships for the growth of moral understanding. Influenced, perhaps, by their view, Tomasello has very little to say about the relevance of parent-child interaction to the developing moral psychology of obligation. But whereas Piaget, Kohlberg, and other traditional moral development theorists contrasted peer cooperation with authority-oriented parent-child relationships, contemporary researchers have a very different view of these relationships (Thompson 2012). Infants learn about cooperative social interaction in the contexts of social play, feeding, and other shared activities of joint intentionality within the family. They expect parents to act consistently with their roles, and protest when parents fail to do so (e.g., Adamson & Frick 2003). A developing morality of obligation derives, in part, from the mutual responsiveness of parents and children in relationships of reciprocal cooperation.

It would be reasonable to expect that humans' ultra-cooperative nature would be manifested in early parent-child interaction, especially because human infants are born more dependent on the cooperative solicitude of their caregivers than are other great apes. The species-typical preference of humans for pursuing goals collaboratively cannot be viewed independently of the conditions required for the survival of human young. Infants' helplessness contributes to the creation of the "social cradle" (Carpendale et al. 2013) of cooperative social interactions that contribute to their survival, especially in the context of alloparenting that is likely to have characterized infant care (Hrdy 2009). But the social cradle is also likely to have made infants exquisitely attuned to the cooperative responses they receive and observe in those around them because cooperative encounters sensitize a helpless infant to whom they could trust to provide reliable support. Understanding from whom cooperative solicitude would be received and, by extension, who could be observed cooperating with others is likely to have enabled infants to direct their bids for support to appropriate adults (Chisholm, *in press*).

This early developing attunement to cooperative relations likely underlies the differential responses of young infants to observed "helpers" and "hinderers" argued by Hamlin (2013) to be a predicate to a moral sense. It is also a foundation to secure or insecure attachment. Infants' sensitivity to cooperative relations changes, of course, with psychological development. We have argued, for example, that as toddlers acquire an awareness of shared intentionality during the second year, they begin to understand parental sensitivity as the adult sharing the toddler's goals in the situation, and they may extend this orientation to third parties (Newton et al. 2016). This is, in Tomasello's terms, the shared "we" in parent-child interaction extended to another.

Another developmental transition in parent-child relationships occurs when young children become capable of – and are motivated to – reciprocate the cooperation they receive. Maccoby

(1984) and more recently Kochanska (2002) have argued that parental socialization during this period is a process of inducting the child into a system of reciprocity, manifested in the formation of a jointly binding, reciprocal, and mutually responsive relationship. In this relationship, young children are motivated to cooperate with parental requests and incentives not because of threat or coercion but owing to the reciprocal obligations that close relationships entail and to maintain relational harmony. Considerable research in the study of conscience development has tested this view and shown, in several longitudinal samples, that when early parent-child relationships were characterized by this "mutually responsive orientation" children subsequently showed greater cooperativeness in parent-child interaction, exhibited greater compliance in the parent's absence, offered greater prosocial completions to incomplete stories, and showed a stronger sense of self as a moral actor (see Kochanska et al. 2019). In Kochanska's view, early parent-child relational experience characterized by shared reciprocity, including felt mutual obligation, launches young children on a developmental pathway that includes greater cooperative conduct and responsiveness to shared responsibilities outside the family.

Early attunement to cooperative relations manifests also in young children's development of secure attachment in the context of sensitive care. Just as a mutually responsive orientation is not guaranteed in early parent-child relationships, so also parental sensitivity is not necessarily ensured in any infant's experience. Thus distinguishing which parental (and other alloparental) partners could be counted on for cooperative support is important for the security felt within those relationships and also to the developing readiness to seek and establish similar relationships of shared agency with others. This may be one reason why the security of attachment is longitudinally associated with many elements of cooperative obligation that Tomasello discusses, such as cooperation with adults and peers, social problem-solving and conflict-avoidance with peers, as well as fidelity to social norms and fairness sensitivity in conscience development (Thompson 2019). Although individual differences in the readiness to established shared agency with another are not a focus of Tomasello's analysis, individual differences in the human sense of obligation warrants systematic study. Interestingly, in an essay written at the conclusion of World War II and thus two decades before the publication of *Attachment*, Bowlby (1946) discussed research documenting the feeling of "We'ness" evoked by children's participation in democratically structured groups, by contrast with the feeling of "I'ness" evoked by participation in autocratically directed groups. His essay urged governments to devote attention to the developmental conditions giving rise to human cooperation and conflict management.

Taken together, the young children studied in the evocative studies described by Tomasello do not arrive at cooperative interactions of joint intentionality with peers and adults *de novo*. Their capacities for establishing the "we'ness" of collaborative activity, mutual obligation, and in-group solidarity are also established in the attunement to cooperative relations and the mutual obligations of relationships in the family. Tomasello's moral psychology of obligation would be developmentally deepened by consideration of a young child's long experience of parent-child interactions that provide foundations to a developing sense of social obligation that precede and are concurrent with the developmental processes he so evocatively discusses. The unique achievements arising from the egalitarian context of peer relationships build on those occurring in the family.

Author's Response

The many faces of obligation

Michael Tomasello^{a,b} 

^aMax Planck Institute for Evolutionary Anthropology, 04103 Leipzig, Germany and

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

michael.tomasello@duke.edu

doi:10.1017/S0140525X19002620, e89

Abstract

My response to the commentaries focuses on four issues: (1) the diversity both within and between cultures of the many different faces of obligation; (2) the possible evolutionary roots of the sense of obligation, including possible sources that I did not consider; (3) the possible ontogenetic roots of the sense of obligation, including especially children's understanding of groups from a third-party perspective (rather than through participation, as in my account); and (4) the relation between philosophical accounts of normative phenomena in general – which are pitched as not totally empirical – and empirical accounts such as my own. I have tried to distinguish comments that argue for extensions of the theory from those that represent genuine disagreement.

As many before me, let me begin by thanking the commentators for their thoughtful comments and by apologizing to them for not being capable of responding equally as thoughtfully to all 32 commentaries in one coherent piece in return. I group my responses, insufficient though they be, around four main topic areas. The first three topics revolve around the three diachronic time frames involved: evolution/phylogeny, development/ontogeny, and cultural evolution. I deal with these each in turn, although in a different order. I then address, fourthly, some philosophical issues raised by commentators who have their doubts that a fully naturalistic explanation of obligation (or any other type of normativity) is possible at all.

R1. Variation within and between cultures

A number of commentators pointed out the myriad complexities of the human sense of obligation in its mature form as it operates in a modern culture, and this is indeed a dimension of things that I mostly neglected in my target article. I suggest that there are two aspects to the issue: (1) the participation of individuals in a collective commitment to the cultural group, and (2) the participation of individuals in various roles to which the culture assigns various normative obligations. I focused mostly on the first, whereas the commentators have noted my neglect of the second.

Maranges, Baumeister, & Vohs (Maranges et al.) emphasize that individuals become members of cultures by sharing in its particular *doxa*, its shared and implicit presuppositions and collective understandings about the way things are. **Kish Bar-On** stresses that individual members of cultures share commitments and obligations not just to one another but to the culture's shared

ideals, values, and moral principles. **Dunning** stresses the glue that is provided among group members, even in-group strangers, by a sense of mutual trust. And this was my central theoretical point at the second step of my story: At some point in human evolution and human ontogeny, individuals begin to feel a part of, even to identify with, a social structure larger than themselves, and without this feeling of solidarity/trust and shared identity there would be no feeling of obligation. As I asked rhetorically: would we feel a sense of obligation to invading Martians?

Nevertheless, many commentators pointed out that things are much more complicated than this. **Clark, Earp, & Crockett (Clark et al.)** point out that, within a modern human culture, individuals feel solidarity with and identify with other individuals and groups based on many different types of social relations. Thus, in communal relations of interdependence we feel obligations to all who contribute in all aspects of life, whereas in exchange relations the obligations to a buyer or seller, for example, are quite contained and transitory. Their commentary is a beautiful elaboration of the different forms that obligation can take in complex societies in which we relate to one another in many complex ways. **Franks & Stewart** echo this theme but emphasize that modern humans participate in many different types of groups – for example, religious, political, and ideological – and these often have conflicting obligations with which individuals must deal. They suggest that this complexity may have been there from the beginning, contributing to the evolution of the human sense of obligation, especially as different hunter-gatherer groups interacted with one another across space and time. **Anderson, Ruisch, & Pizarro (Anderson et al.)** emphasize that complexity also ensues from the distinctions that humans make between prescriptive (what one should do) and proscriptive (what one should not do) obligations, as they arise in the variegated social relationships and groupings characteristic of a modern culture. **Sinnott-Armstrong** points out that many of our obligations are to people with whom we actually have competitive relationships and/or asymmetrical power relationships: for example, employer and employee. He proposes paying attention to “norms of partiality” in which the stronger partner uses her strength to forge a favorable social norm with a weaker partner who has little choice. But, following **Gilbert**, I would argue that the parties creating a social norm still have some obligation to uphold it, even if other considerations (e.g., the more powerful partner's selfish behavior) lead one of them, in the end, to break it. If I promise you to do something, and later find out that it would blow up the world, I should not do it; but I still need to show respect by informing you that I am breaking my promise.

Killen & Dahl argue that the focus on cultural conformity not only misses interesting and important group diversity, but actually misrepresents human morality and its accompanying sense of obligation. Their point is that human morality is decidedly not about group conformity, but rather about the individual living in accordance with her own ideals even if those conflict with those of the group. And even when an individual does accept some cultural norm as legitimate, it is she who judges whether it is or is not legitimate from her perspective. **Hirsh, Shteynberg, & Gelfand (Hirsh et al.)** echo some of these same concerns, pointing out that individuals in modern cultural groups have many conflicting demands from their many different roles and participation in different types of groups (as argued above), and that the essence of morality is resolving these conflicts. **Shaw & Choshen-Hillel** also emphasize the many conflicts of obligation inherent in life within a cultural group, for example, to avoid

harm to individuals but to be fair to everyone. They opine that “obligations are important, but this is just a starting point for understanding the rich tapestry of interlacing obligations that human beings must traverse.”

This general point is one with which I agree wholeheartedly (even if I did not emphasize it in the target article). In a previous publication (Tomasello 2016a, p. 115), I argued that the process by which individuals construct their moral identities within a culture is precisely by attempting to resolve conflicting demands: “Moral beings can never escape the feeling of obligation to make principled decisions when the norms of their group do not apply in a straightforward manner or, more problematically, when they conflict with one another. Individuals must always in some sense freely assent to and identify with the moral decisions that they make.” The key is that the individual constructs a moral identity and thereby takes responsibility for her decisions in adjudicating among multiple obligations. **Olbrich** in fact believes that we can dispense altogether with **Darwall’s** notion of external demands coming from a partner or compatriot as an essential part of obligation. He believes that such demands are not of the essence of the matter, but are merely “a second-order compliance mechanism.” The essence of the matter is that the individual identifies directly with the joint commitment she has made. This focus on the positive side of obligation is also emphasized by **Beeler-Duden, Yucel, & Vaish (Beeler-Duden et al.)**, who point out that we have emotional support, such as the emotion of gratitude, for fulfilling our obligations because we want to (not just because we have to). This is an extremely interesting point about which I have speculated in the past but which is not represented in the target article: People operate from aspirational motives to live up to some very high standards (typically somehow “higher” than is normative in the group, e.g., because of some religious aspirations), as well as feeling an obligation to conform to the standards even if they do not want to at some personal level. The relation between obligational and aspirational relations to cultural norms and standards is an extremely interesting question for future research.

In addition to the fact that individuals must somehow navigate a social terrain full of various and possibly conflicting obligations, both **Bender** and **Buchtel** emphasize that, in different cultural groups, obligations can take much more varied forms than we Western scientists ever consider. **Bender** documents examples from non-Western cultures of social structures that generate obligations of the type we do not normally consider in Western discussions. **Buchtel** notes, in particular, that in some East Asian cultures – echoing the point of the previous paragraph – it is quite common for individuals to take a more aspirational attitude to their obligations, and feel joy in fulfilling them. In my target article, I had a brief section on the lack of cross-cultural research, and these commentaries serve to emphasize that lacuna even more strongly.

R2. Phylogenetic origins of obligation

Kaufmann claims that chimpanzees’ group hunting of monkeys is fully collaborative in a human-like sense. As evidence, she cites **Boesch’s** description of the process in terms of a shared goal with individual roles (like the players on a soccer team). But one can describe the plants, streams, soil, and microorganisms of an ecosystem as playing roles in that ecosystem as well; that does not mean, however, that they understand themselves to be playing such roles. The question is how the chimpanzees

understand what they are doing. Other researchers who have observed chimpanzees hunting monkeys do not describe it using shared goals and individual roles, and I myself have done much experimental research trying to pin down pieces of the process. Clearly, what chimpanzees are doing is collaborative in a general sense, and clearly it was the evolutionary precursor of humans’ more complex forms of collaboration. But it is missing key elements both in terms of coordination of the activity via joint goals and social/communicative coordination, and in terms of dividing the spoils via notions of fairness (Tomasello et al. 2012). **Kaufmann** goes on to compare the chimpanzee version favorably to modern-day hunter-gatherers. But when human hunter-gatherers are hunting, they do such things for their partners as give them weapons, clear trails for them, share information with them, carry their child for them, repair their weapon for them, instruct them in best techniques, and so forth (Hill 2002). There are no reports of chimpanzees helping one another in their group hunting in any way. This tells you that the cooperative structure of the human version is somehow different.

Reddy & Wellman believe that the ur-context for the evolution of a sense of obligation is the mother-child relationship. They point out that when a child screams, it is in effect “coercive” because the mother would much rather stay doing what she is doing; she is obliged to react. But such maternal behavior is widespread in virtually all mammals, and so the question arises why the full-fledged sense of obligation arose only in the primate lineage leading to humans. Humans’ closest living relatives, chimpanzees and bonobos, engage in group hunting of various small mammals (see above), and the proposal here is that this is the special context that took humans beyond general mammalian maternal relationships. In **Reddy & Wellman’s** formulation, “proto-obligatory motivations and expectations may begin with, ‘X helps me because X is my mother/I must help X because X is my offspring.’ As individuals in group-living species grow up, this understanding may develop into ‘because X is my friend’ and ‘because X is Y’s daughter.’” Perhaps. But I simply do not see how this general reciprocity relation contains the seeds of an individual who sees that others have legitimate claims on her by virtue of their mutual interdependence.

And this is in many ways the same problem I have with the analysis of **Rotella, Sparks, & Barclay (Rotella et al.)**. They basically propose an account of the origins of a sense of obligation in terms of reciprocity. I have argued at length (in Ch. 2 of **Tomasello 2016a**) that reciprocity cannot be the psychological basis of human morality, normativity, and obligation. Their argument is “the feeling of obligation arises anytime that failure to complete the obligated action would result in negative social consequences” (p. 42). But I would claim that the feeling of obligation can arise only if those negative social consequences are of a particular type. Yes, concern for reputation would be an important step along the way, and, as noted in the article, humans have a concern for having a cooperative reputation that other great apes do not seem to have. This is a key starting point. But to get to obligation, I do not believe that concern with what “they” think of me is sufficient because it does not confer the legitimacy needed for a sense of obligation: I agree with them that I should do X, so it is the evaluation of “we” that concerns me. The standards that we have agreed to are what concern me. In **Rotella et al.’s** analysis, it is hard to see why I should feel the need to apologize or make an excuse to someone when I do not fulfill my obligation, or else to feel guilty on my own. This

is the special force of obligation beyond “negative social consequences.”

R3. Ontogenetic origins of obligation

Some commentators with a background in developmental psychology had different proposals for the cradle of children’s sense of obligation. The two main ones were (1) parent-child interactions and (2) the understanding of groups from a third-party perspective (rather than from participation). These will each be treated in turn, followed by treatment of a few other proposals from a developmental perspective.

R3.1. Parent-child relations as basic

Thompson criticizes the view that children’s earliest sense of obligation emerges from their more-or-less egalitarian interactions with peers and, in some contexts, with adults. Rather, he believes that the key is infants’ earliest interactions with their caregivers, as affectively laden relationships cemented by obligations. Thompson argues that “early experiences of cooperation and shared social agency between parents and infants ... provide a foundation for developing understanding of the mutual obligations of close relationships,” which can then be extended later to others. **Carpendale & Lewis** have a similar view, elaborated in more detail as follows: “Moral obligation is not the result of realizing that others are equals and therefore should be treated with respect. Instead it is a natural outcome of mutual affection and understanding. This caring and concern for others is what later develops into a sense of obligation, first to those close and, later, extended to others.” Although this is clearly a reasonable view, and has the virtue of beginning with children’s first important social relations, in the end I do not believe it can get us to obligation.

The main issue is that the sense of obligation is mainly self-regulatory. It is “another voice” alongside of humans’ most basic and natural tendencies for social action and interaction. In my opinion, children’s earliest social relations with mothers and other caregivers are forming the foundation for their primary skills and motivations for interacting with others in all kinds of ways, both cooperative and otherwise. The sense of obligation arises from an executive level of consideration that evaluates the social options and puts its finger on the scale for certain ones of them (particularly in contexts of joint agency in which “we” are evaluating “me”). But this is potentially an empirical question. **Marshall** reports a study (Marshall et al. [in preparation](#)) in which 5-year-old and older children expressed the belief that parents have obligations to their children (they *have to do* things for their children that others do not). But as Marshall herself notes, these are older children, and they would already have a group-minded understanding of the role-obligations inherent in various social relations in their culture, including parent-child relations. My view would be that infants and young children understand parent-child social relations in terms of basic emotional and social processes of mutual affection and reciprocity, but without a sense of obligation. Many mammalian mothers care for their infants and young children as assiduously as do human caregivers, but neither they nor their offspring come to an understanding of obligation (as far as we know). So my prediction would be that if Marshall’s study were done with 2- to 3-year-old children, they would not think that parents have special obligations to their children, above and beyond their natural social affections.

R.3.2. The understanding of groups as basic

Recent studies have found that, in looking at time studies, human infants have certain expectations about how members of a group (even if they are represented by shapes on a screen) behave. Children 3 years of age and older have intuitive theories about social groups independent of their own participation in them, and they can even speak coherently about the normative relations of members: things they *should* or *must* do for one another. Several commentators, therefore, expressed the view that perhaps children’s early expectations and intuitive theories about social groups – in contrast to their direct participation in collaborative social interactions – might be the developmental cradle of children’s early understanding of how individuals, including themselves, are obligated to one another.

Rhodes characterizes my view involving children’s direct participation in social interactions as “inside-out.” That is, they first feel a sense of obligation to their collaborative partners, and then generalize outside that to their and other social groups more broadly. She contrasts this with her own “outside-in” view that children’s understanding of obligations develops from their understanding of social groups independent of their own participation in them. **Chalik** expresses a similar view, focusing on children’s intuitive theories of social groups. Both commentators note that it is possible that both children’s direct collaborative interactions and their intuitive understanding of social groups contribute to the process, although, if true, the relation of these two processes remains totally unexplored. The possibility that these two sources develop in parallel is an interesting and plausible one, but for the moment I remain skeptical that the infant version is truly group-minded. More likely, in my view, the kinds of “outside-in” judgments about the obligations that group members have to one another that 3-year-old children make are not continuous with infants’ non-normative expectations about the behavior of similarly shaped dots on a screen. Instead, my proposal would be that 3-year-olds’ judgments reflect a maturationally structured understanding that first emerges at around that age: a kind of group-mindedness that encompasses both themselves in their own groups as well as others. This is clearly an extremely interesting and provocative question for future research.

In a somewhat similar vein, **Lieberman & Du Bois** criticize my account for its almost exclusive focus on the developmental period comprising several years on either side of the third birthday. They cite a good bit of the infant literature on children’s expectations and preferences toward experimental stimuli based on their apparent group status. But, again, the continuity of these nonnormative expectations and preferences with children’s later sense of moral obligation is at this point unexamined. They go on to point out, in agreement with some of the culture theorists discussed above, that working out obligations in social interactions with others is a lifelong process that involves understanding not only other individuals but also the social norms of the group (at whatever level of analysis). **Bartsch** provides some very interesting developmental analyses of young children in conversation with adults and other children focused on moral issues and obligation. It is undoubtedly in such discourse interactions that young children begin the process of understanding how to interact with others cooperatively in more complex ways across the lifespan.

R.3.3. Other views

In contrast to accounts focused on children’s intuitive theories of social groups, **Corbit & Moore** focus, as I do, on interactions in

which the child herself participates. They argue that my account does not really specify the kinds of interactions through which young children gain the ability to form a joint agency with others in the first place. They argue that infants' earliest triadic interactions with others around objects provide the requisite experiences whereby the infant learns to coordinate first-person perspective, second-person perspective, and the object on which they both are focused. In my view, this is a reasonable characterization, but, based on the comparative work with great apes – even those raised in rich interactions around objects with humans – I believe we must recognize a role for the maturation of capacities for shared intentionality that structure these early learning processes. Only if one can form a joint goal with joint attention, in my view, is it possible to take different perspectives in the first place; otherwise we are just doing and seeing different things. So, I am not convinced that infants are learning to coordinate perspectives before their capacities for joint intentionality mature. Corbit & Moore also cite evidence supporting the view that acting together toward a joint goal – even if that goal is not obtaining resources – leads children to distribute resources more fairly with others.

Kalish introduces the notion of a reason for action and notes that children's earliest moral statements belie a kind of objective notion: the reason my sister is opening the box is because there is candy in there. But, he argues, this is a rationality of individual action. When the child must coordinate with a partner toward a shared goal, now the reasons must become more person-specific involving the beliefs and attitudes of the partners vis-à-vis one another. He notes that if a partner in the collaboration misbehaves, he is obligated to give the other a reason and "citing the obligation is an effective way of restoring coordination. I remind you of the reason, and expect you to adjust your behavior in light of it." This is an elegant way of construing the situation, because, in my view, while individuals' behavior is driven by beliefs and desires, reasons only come into existence as kind of self-regulatory mechanisms in shared intentionality interactions in which one individual is obligated to the other already. **Rakoczy** notes a number of ways in which my account conflates – or at least fails to keep separate – children's sense of an intentional stance, a group stance, a moral stance, and shared intentionality. He very articulately illustrates how even a single interaction may instantiate these different stances and capacities such that, over developmental time, they become more distinct aspects of human social psychological functioning.

Li, Pesch, & Koenig extend the analysis of obligation in extremely interesting ways to testimonial exchanges and note the various ways that children holds others accountable – evincing a sense of obligation – epistemically.

R4. Whence normativity?

Four of the philosophers who have provided commentaries are those upon whose work I have drawn most deeply and directly to construct concepts of collaboration that are adequate for describing developmental changes in young children's ability to cooperate with others across the first years of life. It is gratifying that they all seem to believe that although my account is misguided, it is at least worthy of criticism. The most basic issue concerns the very possibility of a naturalistic account of normative phenomena, including obligation.

Bratman's (2014) account of shared intentional activities is mainly instrumental, with no sense of obligation among partners:

An adult and child can play with the ball together, and "The child can reason instrumentally concerning this jointly intended end of playing together without thinking that she owes it to the adult to give him the ball." **Pettit** thus says that "Bratman offers a normatively austere account of how joint action may materialize." In contrast, again according to Pettit, "Gilbert's account is normatively richer than Bratman's, because it presupposes that the people involved in the joint action have mastered a normative concept: that of a joint commitment." And, according to **Gilbert** herself, "The general notion of commitment at issue is normative, rather than psychological." **Darwall** concurs: "obligation is a normative, rather than a psychological (or sociological) phenomenon." Most psychologists, myself included, find it difficult to comprehend what obligation and other normative phenomena could possibly be if not psychological (or sociological) phenomena. For sure, obligation and other normative phenomena go beyond individual psychology; but that is precisely why we need a shared intentionality account.

Perhaps because of this non-naturalistic conception of obligation and other normative phenomena, all four philosophers seem to agree that a naturalistic account going from Bratman-like, instrumentally based collaboration to Gilbert-like, normatively based collaboration is problematic. **Pettit** says I have a hard choice. If I define joint action à la Bratman, then I cannot get to a normatively richer version involving obligation without some kind of magical leap. If I define joint action à la Gilbert, then I build obligation into the definition. **Bratman** makes a similar argument. He claims that collaboration itself – in his "thin" version – does not lead to the sense of obligation to one's partner, but rather obligation arises from joint commitments in Gilbert's sense, comprising a "web of assurances" between individuals that is only contingently linked to collaboration: "So what is undergirding [obligation] involves a crucial element that is separable from intentional collaborative activity."

But my phylogenetic and ontogenetic account is intended precisely to get us from Bratman-like collaboration to Gilbert-like collaboration as cemented by a joint commitment involving obligations. The key is not to think of normative phenomena in absolutistic terms but rather as graded. Joint commitments – as collaboratively created forms of collaborative self-regulation – can be weaker or stronger based on several factors. Consider a recent experiment modelled on the stag hunt from game theory (Siposova et al. 2018). Children were in a situation in which they could only get a large reward if they collaborated with an adult partner. As they were busy collecting small rewards (hare), the opportunity to collaborate for the large reward (stag) suddenly appeared at a distance. In one condition, the adult excitedly looked to the child with wide eyes and then to the stag in the distance without explicitly proposing any kind of joint commitment. The idea was that the adult would just be making it clear that she had just seen the stag and knew that the child had seen the stag and knew that the child had seen her seeing the stag, et cetera – that is, they had common ground knowledge that the stag was now available. But no one explicitly suggested a joint commitment. Nevertheless, not only did the child commit to pursuing the stag, but if the adult slacked off, the child quite often castigated her normatively. Our interpretation was that children understood that it was in their common knowledge with the adult that they depended on one another for collaborative success and that they both knew that the stag was present: which, it turns out, is sufficient to create a mutual sense of obligation. If we both know together, in common ground, that you need help and that

I am in a position to provide it (at little cost), I feel the pressure of obligation.

But clearly this is “less” of a joint commitment than if the adult says to the child “Let’s work together,” and the child responds “Okay.” In that case, the joint commitment is proposed and accepted explicitly and publicly. In the analysis of Pettit (2018a), the explicitness and publicity provide stronger assurances because they foreclose the possibility of excuses like “I didn’t know you were counting on me” and/or “I changed my mind [without excuse].” The point is that joint commitments and their associated obligations can vary in strength based on such things as how certain the participants are about the intentions of the other and how important one’s contribution is to joint success. Thus, one can imagine a situation in which one makes a joint commitment of one type or another, but one’s contribution is only minimally helpful; for example, I commit to you to join you at a party, but as only one of 30 people, and as only an acquaintance, I know that you are not depending on me that strongly, so breaking the commitment is not that big of a deal; my sense of obligation is weaker as a result. If joint commitments and their associated obligations can vary in strength based on several factors like certainty of the commitment’s content and its social importance, then one can imagine that early humans and young children begin with Bratman-like instrumental collaboration, and then such things as mutuality looks to up the ante in terms of commitments and obligations, and then explicit proposals and acceptances make the web of assurances stronger and more certain. This kind of gradualistic account makes a naturalistic explanation of the emergence of normative phenomena, including the human sense of obligation, at least a bit more plausible. Supportive of this idea, Kachel and Tomasello (2019) reported an experimental demonstration that implicit joint commitments lead to a weaker sense of obligation than do explicit ones.

My naturalistic account has two other key concepts as well, and attention to these might make the account more plausible to these and other philosophers. The first is borrowed from Nagel (1970b), and that is the idea that seeing other people as in some sense equivalent to myself (I am only one among many other individuals) provides a kind of cognitive, not motivational, basis for a sense of equal deservingness. I might want more of the spoils myself, and I might even take them, but I certainly do not feel that I deserve them because we (self and other) are, in all important respects, the same in this context. I feel obliged to treat you as you deserve to be treated. The second concept is borrowed from Korsgaard (1996), and that is the idea that the notion of obligation is tied to my cooperative or moral identity. I, as a member of a “we” community, judge people normatively. And I can turn this on myself and judge myself from the perspective of that “we.” It is simply an empirical fact that it is important to human beings that they can view themselves, from the perspective of “we” in the community, positively; they want to be an integral and contributing member of the “we,” and indeed they feel obliged to do so. This also lends legitimacy to the social demands that one feels from others. I reiterate these two points to emphasize that my account is not simply that the sense of obligation arises from individuals solving a puzzle together instrumentally. It arises from a cooperative way of life, which shapes human understanding of their relation to others and their need to form plural agencies in their quest to survive and thrive in their social group.

One final point on this general topic. Darwall takes issue with the proposal that obligation is a fully psychological phenomenon. He argues that “it is irreducibly normative and moral, not

empirical.” He does not believe that my account in which children come to feel a sense of obligation solely from their interdependent collaborations with equally deserving partners with whom they feel a sense of “we” is sufficient. He thinks that the normative sense of obligation presupposes something less of this world: not an empirical “we,” but an ideal “we.” I do not disagree with this, on one level, and indeed I said in my target article that the larger cultural group (at the second step) was not just a finite collection of individuals but “everyone who would be one of us,” thus tying the normative force to a kind of idealized characterization of the group with which I identify. Nevertheless, I do not see what Darwall could possibly mean by the claim that “Although the ‘sense of obligation’ is psychological, what it is a sense of is decidedly not.” The sense of obligation is a sense of what I should do to comport with the mutual understanding that my partners and I have about how someone in my position should treat others. What is not psychological about that?

R5. Conclusion

Reading and responding to 32 commentaries on one’s best efforts at a coherent theoretical account of something, anything, is a humbling experience. There exist so many different perspectives and valuing in the intellectual community that reads and profits from this journal. “We” in this community will make progress only if each of us feels as legitimate a kind of we > me sense of obligation to take all of these different perspectives and valuing seriously – and, perhaps, to make improvements as a result.

I believe that to make further progress on the questions I have raised, my account needs to be modified and improved in the following ways:

We need to integrate humans’ various normative attitudes, including the sense of obligation, with their more basic social relations, especially those that precede them in evolution and development, such as parent-child relations and perhaps friendships as the initial and most basic forms of human sociality.

We need to identify the many faces of obligation as it manifests in the many different social contexts and social roles in a modern society. Moreover, we need to recognize that there may be different forms of obligation in different societies with different social structures and arrangements and seek to identify those.

In terms of ontogeny, we currently have no good account of how children’s understanding of social partnerships and groups that they derive through participation in them are related to their understanding of social partnerships and groups that they derive through observation of others (from a third-party perspective). This is a challenging, but ultimately empirically tractable, question.

Finally is the issue of how far we can go in our understanding of normative attitudes, including the sense of obligation, empirically. From its beginnings, philosophy has recognized a kind of ideal world that is – to some degree and in one way or another – independent of (scientific) observation and experience. Applied to the current case: “ought” cannot be explained in terms of “is.” But I would argue that *the human sense* of “ought” is a psychological phenomenon capable of scientific explanation, and so the question is whether there is still something left over unexplained. Of course, no explanation of obligation, scientific or otherwise, can tell us what we ought to do. But we

can describe and explain the processes by which people come to feel that they ought and ought not to do certain things, without leaving out of account any supposed non-empirical aspects.

In any case, one of the surprises I had in writing this target article was how little empirical work there was on obligation. I am hopeful that the current exchange will lead more scientists and philosophers to investigate this interesting and centrally important human phenomenon from as many different perspectives as possible.

References

[The letters “a” and “r” before author’s initials stand for target article and response references, respectively]

- Adams, R. (2002) *Finite and infinite goods: A framework for ethics*. Oxford University Press. [DO]
- Adamson, L. B. & Frick, J. E. (2003) The still face: A history of a shared experimental paradigm. *Infancy* 4(4):451–73. doi:10.1207/S15327078IN0404_01 [RAT]
- Aktipis, A., Cronk, L., Alcock, J., Ayers, J. D., Baciu, C., Balliet, D., Boody, A. M., Curry, O. S., Krems, J. A., Muñoz, A., Sullivan, D., Sznycer, D., Wilkinson, G. S. & Winfrey, P. (2018) Understanding cooperation through fitness interdependence. *Nature Human Behaviour* 2(7):429. [AR]
- Ames, R. T. & Rosemont, H. J. (2014) Family Reverence (xiao 孝) in the Analects: Confucian role ethics and the dynamics of intergenerational transmission. In: *Dao companion to the Analects*, ed. A. Olberding, pp. 117–36. Springer. [EEB]
- Angle, S. & Slote, M. (2013) *Virtue ethics and Confucianism*. Routledge. [EEB]
- Anscombe, G. E. M. (1958) Modern moral philosophy. *Philosophy* 33(124):1–19. [EEB]
- Aquino, K. & Reed, I. I. (2002) The self-importance of moral identity. *Journal of Personality and Social Psychology* 83(6):1423. [BF]
- Arsenio, W. F. & Ford, M. E. (1985) The role of affective information in social-cognitive development: children’s differentiation of moral and conventional events. *Merrill-Palmer Quarterly* 31(1):1–17. <https://doi.org/10.1088/0143-0807/28/3/011> [SB-D]
- Axelrod, R. (1984) *The evolution of cooperation*. Basic Books. [PP]
- Bakken, B. (2000) *The exemplary society: Human improvement, social control, and the dangers of modernity in China*. Oxford University Press. [EEB]
- Balliet, D., Tybur, J. M. & Van Lange, P. A. (2017) Functional interdependence theory: An evolutionary account of social situations. *Personality and Social Psychology Review* 21(4):361–88. [AR]
- Barclay, P. (2013) Strategies for cooperation in biological markets, especially for humans. *Evolution & Human Behavior* 34(3):164–75. [AR]
- Barclay, P. (2016) Biological markets and the effects of partner choice on cooperation and friendship. *Current Opinion in Psychology* 7:33–38. [AR]
- Barclay, P. & Lalumiere, M. L. (2006) Do people differentially remember cheaters? *Human Nature* 17(1):98–113. [RAA]
- Barclay, P. & Van Vugt, M. (2015) The evolutionary psychology of human prosociality: adaptations, mistakes, and byproducts. In: *Oxford handbook of prosocial behavior*, ed. D. Schroeder & W. Graziano, pp. 37–60. Oxford University Press. [AR]
- Barclay, P. & Willer, R. (2007) Partner choice creates competitive altruism in humans. *Proceedings of the Royal Society B: Biological Sciences* 274(1610):749–53. [AR]
- Bar-Haim, Y., Ziv, T., Lamy, D. & Hodes, R. M. (2006) Nature and nurture in own-race face processing. *Psychological Science* 17(2):159–63. <https://doi.org/10.1111/j.1467-9280.2006.01679.x> [LC]
- Barresi, J. & Moore, C. (1996) Intentional relations and social understanding. *Behavioral and Brain Sciences* 19(1):10722. doi:10.1017/S0140525X00041790. [JC]
- Bartsch, K. & Wellman, H. M. (1995) *Children talk about the mind*. Oxford University Press. [KB, CK]
- Baumard, N., André, J. B. & Sperber, D. (2013) A mutualistic approach to morality. *Behavioral and Brain Sciences* 36(1):59–122. [aMT]
- Baumeister, R. F. (2005) *The cultural animal: Human nature, meaning, and social life*. Oxford University Press. [HMM]
- Baumeister, R. F., Ainsworth, S. E. & Vohs, K. D. (2016) Are groups more or less than the sum of their members? The moderating role of individual identification. *Behavioral and Brain Sciences* 39:e137. doi:10.1017/S0140525X15000618. [HMM]
- Baumeister, R. F., Maranges, H. M. & Vohs, K. D. (2018) Human self as information agent: Functioning in a social environment based on shared meanings. *Review of General Psychology* 22(1):36–47. [HMM]
- Begus, K., Gliga, T. & Southgate, V. (2016) Infants’ preferences for native speakers are associated with an expectation of information. *Proceedings of the National Academy of Sciences* 113(44):12397–402. [ZL]
- Begus, K. & Southgate, V. (2012) Infant pointing serves an interrogative function. *Developmental Science* 15(5):611–17. [PHL]
- Beller, S. (2008) Deontic norms, deontic reasoning, and deontic conditionals. *Thinking and Reasoning* 14:305–41. [AB]
- Beller, S. (2010) Deontic reasoning reviewed: psychological questions, empirical findings, and current theories. *Cognitive Processing* 11:123–32. [AB]
- Beller, S. (2012) Human deontic reasoning in the deontic square. In: *The square of opposition: A general framework for cognition*, ed. J.-Y. Beziau & G. Payette, pp. 471–84. Peter Lang. [AB]
- Beller, S., Bender, A. & KuhnMünch, G. (2005) Understanding conditional promises and threats. *Thinking and Reasoning* 11:209–38. [AB]
- Beller, S., Bender, A. & Song, J. (2009) Conditional promises and threats in Germany, China, and Tonga: Cognition and emotion. *Journal of Cognition and Culture* 9:115–39. [AB]
- Bender, A. (2007) Changes in social orientation: Threats to a cultural institution in marine resource exploitation in Tonga. *Human Organization* 66:11–21. [AB]
- Bender, A. & Beller, S. (2019) The cultural fabric of human causal cognition. *Perspectives on Psychological Science* 14:922–40. [AB]
- Bender, A., Kägi, W. & Mohr, E. (2002) Informal insurance and sustainable management of common-pool marine resources in Ha’apai, Tonga. *Economic Development and Cultural Change* 50:427–39. [AB]
- Benioff, D. & Weiss, D. B. (Writer) (2012) A man without honor [Episode 17]. In: David Benioff & D. B. Weiss’s (Executive Producers), *Game of Thrones*. Home Box Office. [AS]
- Berg, J., Dickhaut, J. & McCabe, K. (1995) Trust, reciprocity, and social history. *Games and Economic Behavior* 10(1):122–42. [DD]
- Berscheid, E. & Ammazarlorso, H. (2001) Emotional experience in close relationships. In: *Blackwell handbook of social psychology: Interpersonal processes*, ed. G. J. O. Fletcher & M. S. Clark, pp. 308–30. Blackwell. [MSC]
- Bian, L., Sloane, S. & Baillargeon, R. (2018) Infants expect ingroup support to override fairness when resources are limited. *Proceedings of the National Academy of Sciences* 115(11):2705–10. <https://doi.org/10.1073/pnas.1719445115>. [LC]
- Blatz, C. W., Schumann, K. & Ross, M. (2009) Government apologies for historical injustices. *Political Psychology* 30(2):219–41. [RAA]
- Bloom, P. (2011) Family, community, trolley problems, and the crisis in moral psychology. *The Yale Review* 99(2):26–43. [MSC, JM]
- Bloom, P. (2013) *Just babies: The origins of good and evil*. Broadway Books. [JM]
- Bodenhausen, G. V. (2010) Diversity in the person, diversity in the group: Challenges of identity complexity for social perception and social interaction. *European Journal of Social Psychology* 40(1):1–16. [JBH]
- Boesch, C. (2005) Joint cooperative hunting among wild chimpanzees: Taking natural observations seriously. *Behavioural and Brain Sciences* 28(5):692–93. [AK]
- Boesch, C. & Boesch-Achermann, H. (2000) *The chimpanzees of the Tai forest*. Oxford University Press. [AK]
- Bohner, G., Bless, H., Schwarz, N. & Strack, F. (1998) What triggers causal attributions? The impact of valence and subjective probability. *European Journal of Social Psychology* 18(4):335–45. [RAA]
- Bostyn, D. H. & Roets, A. (2016) The morality of action: The asymmetry between judgments of praise and blame in the action–omission effect. *Journal of Experimental Social Psychology* 63:19–25. [RAA]
- Bourdieu, P. (1977) *Outline of a theory of practice, vol. 16*. Cambridge University Press. [HMM]
- Bowlby, J. (1946) Psychology and democracy. *The Political Quarterly* 17(1):61–75. doi:10.1111/j.1467-923X.1946.tb01028.x. [RAT]
- Brandom, R. (1994) *Making it explicit*. Harvard University Press. [PHL]
- Bratman, M. (1992) Shared co-operative activity. *Philosophical Review* 101(2):327–41. [aMT]
- Bratman, M. E. (2014) *Shared agency: A planning theory of acting together*. Oxford University Press. [MEB, AK, PP, arMT]
- Brosnan, S. F. & de Waal, F. B. M. (2003) Monkeys reject unequal pay. *Nature* 425:297–99. [AK, aMT]
- Buchtel, E. E., Guan, Y., Peng, Q., Su, Y., Sang, B., Chen, S. X. & Bond, M. H. (2015) Immorality East and West: Are immoral behaviors especially harmful, or especially uncivilized? *Personality and Social Psychology Bulletin* 41(10):1382–94. <https://doi.org/10.1177/0146167215595606>. [EEB]
- Buchtel, E. E., Ng, L. C. Y., Norenzayan, A., Heine, S. J., Biesanz, J. C., Chen, S. X., Bond, M. H., Peng, Q. & Su, Y. (2018) A sense of obligation: Cultural differences in the experience of obligation. *Personality and Social Psychology Bulletin* 44(11):1545–66. <https://doi.org/10.1177/0146167218769610>. [EEB]
- Bugental, D. B. (2000) Acquisition of the algorithms of social life: A domain-based approach. *Psychological Bulletin* 126(2):187–219. [MSC]
- Butler, L. & Markman, E. (2014) Preschoolers use pedagogical cues to guide radical reorganization of category knowledge. *Cognition* 130(1):116–27. [aMT]
- Butler, L. P. and Tomasello, 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. [aMT]

- Buttelmann, D. & Boehm, R. (2014) The ontogeny of the motivation that underlies in-group bias. *Psychological Science* 25(4):921–27. [MR]
- Buttelmann, D., Zmyj, N., Daum, M. & Carpenter, M. (2013) Selective imitation of in-group over outgroup members in 14-month-old infants. *Child Development* 84(2):422–28. [ZL]
- Cain, D. M., Dana, J. & Newman, G. E. (2014) Giving versus giving in. *The Academy of Management Annals* 8(1):505–33. [DD]
- Carpendale, J. I. M. (2018) Communication as the coordination of activity: The implications of philosophical preconceptions for theories of the development of communication. In: *Advancing developmental science: Philosophy, theory, and method*, ed. A. S. Dick & U. Müller, pp. 145–56. Routledge, Taylor & Francis group. [JIMC]
- Carpendale, J. I. M., Hammond, S. I. & Atwood, S. (2013) Chapter Six – A relational developmental systems approach to moral development. In: *Advances in child development and behavior*, vol. 44, ed. R. M. Lerner & J. B. Benson, pp. 125–53. Elsevier Science. doi:10.1016/B978-0-12-397946-9.00006-3. [JIMC, RAT]
- Carpendale, J. I. M. & Lewis, C. (2004) Constructing an understanding of mind: The development of children's social understanding within social interaction. *Behavioral and Brain Sciences* 27(1):79–96. [JIMC]
- Carpendale, J. I. M. & Lewis, C. (2006) *How children develop social understanding*. Blackwell. [JIMC]
- Carpendale, J. I. M. & Lewis, C. (2015a) Taking natural history seriously in studying the social formation of thinking: Critical analysis of *A natural history of human thinking* by Michael Tomasello. *Human Development* 58(1):55–66. [JIMC]
- Carpendale, J. I. M. & Lewis, C. (2015b) The development of social understanding. In: *Handbook of child psychology and developmental science*, vol. 2: *Cognitive processes*, ed. L. Liben & U. Müller, pp. 381–424. Wiley-Blackwell. [JIMC]
- Chalik, L. & Dunham, Y. (2020) Beliefs about moral obligation structure children's social category-based expectations. *Child Development* 91(1):e108–e119. <https://doi.org/10.1111/cdev.13165>. [LC]
- Chalik, L. & Rhodes, M. (2014) Preschoolers use social allegiances to predict behavior. *Journal of Cognition and Development* 15(1):136–60. <https://doi.org/10.1080/15248372.2012.728546>. [LC]
- Chalik, L. & Rhodes, M. (2015) The communication of naïve theories of the social world in parent–child conversation. *Journal of Cognition and Development* 16:719–41. [aMT]
- Chalik, L. & Rhodes, M. (2018) Learning about social category-based obligations. *Cognitive Development* 48:117–24. <https://doi.org/10.1016/j.cogdev.2018.06.010>. [LC, MR]
- Chalik, L. & Rhodes, M. (forthcoming) Groups as moral boundaries: A developmental perspective. In: *Advanced in child development and behavior*, vol. 58, ed. J. Benson. [MR]
- Chalik, L., Rivera, C. & Rhodes, M. (2014) Children's use of categories and mental states to predict social behavior. *Developmental Psychology* 50(10):2360–67. <https://doi.org/10.1037/a0037729>. [LC]
- Chapais, B. (2008) *Primeval Kinship: How pair-bonding gave birth to human society*. Harvard University Press. [BF]
- Chapais, B. (2011) The deep social structure of human kind. *Science* 331:1276–77. [BF]
- Cheney, D. L. & Seyfarth, R. M. (1980) Vocal recognition in free-ranging vervet monkeys. *Animal Behaviour* 28(2):362–67. [RBR]
- Chisholm, J. S. (in press) Attachment and the deep history of culture. In: *Attachment: The fundamental questions*, ed. R. A. Thompson, J. A. Simpson & L. Berlin. Guilford. [RAT]
- Choshen-Hillel, S., Lin, Z. & Shaw, A. (in press) Children weigh equity and efficiency in making allocation decisions: Evidence from the US, Israel, and China. *Journal of Economic Behavior & Organization*. [AS]
- Cialdini, R. B. & Goldstein, N. J. (2004) Social influence: Compliance and conformity. *Annual Review of Psychology* 55(1):591–621. [JBH]
- Clark, M. S. (1984) Record keeping in two types of relationships. *Journal of Personality and Social Psychology* 47(3):549–57. [MSC]
- Clark, M. S. & Boothby, E. (2013) A strange(r) analysis of morality: a consideration of relational context and the broader literature is needed. *Brain and Behavioral Sciences* 36(1):85–86. [MSC]
- Clark, M. S., Dubash, P. & Mills, J. (1998) Interest in another's consideration of one's needs. *Journal of Experimental Social Psychology* 34(3):246–64. [MSC]
- Clark, M. S., Lemay, E. P. & Reis, H. T. (2017) Other people as situations: Relational context shapes psychological phenomena. In: *Oxford handbook of situations*, ed. D. Funder & R. Sherman, pp. 1–40. Oxford University Press. [MSC]
- Clark, M. S. & Mills, J. (1979) Interpersonal attraction in exchange and communal relationships. *Journal of Personality and Social Psychology* 37(1):12–24. [MSC]
- Clark, M. S. & Mills, J. (1993) The difference between communal and exchange relationships: What it is and is not. *Personality and Social Psychology Bulletin* 19:684–91. [JM]
- Clark, M. S. & Mills, J. (2012) A theory of communal (and exchange) relationships. In: *Handbook of theories of social psychology*, vol. 2, ed. P. A. M. Van Lange, A. W. Kruglanski & E. T. Higgins, pp. 232–50. SAGE. [MSC]
- Clark, M. S., Mills, J. & Corcoran, D. (1989) Keeping track of needs and inputs of friends and strangers. *Personality and Social Psychology Bulletin* 15(4):533–42. [MSC]
- Clark, M. S., Ouellette, R., Powell, M. C. & Milberg, S. (1987) Recipient's mood, relationship type, and helping. *Journal of Personality and Social Psychology* 53(1):94–103. [MSC]
- Corbit, J. (2019) Increased sharing between collaborators extends beyond the spoils of collaboration. *Journal of Experimental Child Psychology* 186:159–70. doi:10.1016/j.jecp.2019.06.003. [JC]
- Corbit, J., McAuliffe, K., Callaghan, T. C., Blake, P. R. & Warneken, F. (2017) Children's collaboration induces fairness rather than generosity. *Cognition* 168:344–56. doi:10.1016/j.cognition.2017.07.006. [JC, aMT]
- Cosmides, L., Barrett, H. C. & Tooby, J. (2010) Adaptive specializations, social exchange, and the evolution of human intelligence. *Proceedings of the National Academy of Sciences* 107(Suppl. 2):9007–14. [DD]
- Cosmides, L. & Tooby, J. (2004) Social exchange: The evolutionary design of a neurocognitive system. In: *The cognitive neurosciences III*, ed. M. S. Gazzaniga, pp. 1295–1308. MIT Press. [aMT]
- Csibra, G. & Gergely, G. (2009) Natural pedagogy. *Trends in Cognitive Sciences* 13(4):148–53. [HMM, aMT]
- Curry, O. S., Mullins, D. A. & Whitehouse, H. (2019) Is it good to cooperate? Testing the theory of morality-as-cooperation in 60 societies. *Current Anthropology* 60(1):47. [BF]
- Dahl, A. (2014) Definitions and developmental processes in research on infant morality. *Human Development* 57(4):241–49. doi:10.1159/000364919. [MK]
- Dahl, A. & Killen, M. (2018) Moral reasoning: Theory and research in developmental science. In: *The Steven's handbook of experimental psychology and cognitive neuroscience*, Vol. 4: *Developmental and social psychology* (S. Ghetti, Vol. Ed.), 4th edition, ed. J. Wixted, pp. 1–31. Wiley. doi:10.1002/9781119170174.epcn410. [MK]
- Dahl, A., Sherlock, B. R., Campos, J. J. & Theunissen, F. E. (2014) Mothers' tone of voice depends on the nature of infants' transgressions. *Emotion* 14(4):651–65. <https://doi.org/10.1037/a0036608>. [SB-D]
- Dahl, A. & Tran, A. Q. (2016) Vocal tones influence young children's responses to prohibitions. *Journal of Experimental Child Psychology* 152:71–91. <https://doi.org/10.1016/j.jecp.2016.07.009>. [SB-D]
- Darwall, S. (2006) *The second-person standpoint: Morality, respect, and accountability*. Harvard University Press. [SD, DO, aMT]
- Darwall, S. (2013a) “Bipolar obligation.” In: *Morality, authority, and law: Essays in second-personal ethics I*, ed. S. Darwall, pp. 20–39. Oxford University Press. [SD, aMT]
- Darwall, S. (2013b) *Honor, history, and relationship: Essays in second-personal ethics II*. Oxford University Press. [SD]
- Darwall, S. (2013c) *Morality, authority, and law: Essays in second-personal ethics I*. Oxford University Press. [SD, CK, aMT]
- de León, L. (2019) Playing at being bilingual: Bilingual performances, stance, and language scaling in Mayan Tzotzil siblings' play. *Journal of Pragmatics* 144:92–108. <https://doi.org/10.1016/j.pragma.2018.02.006>. [ZL]
- de Waal, F. B. M. (2006a) Joint ventures require joint payoffs: Fairness among primates. *Social Research* 73(2):349–64. [AK]
- de Waal, F. B. M. (2006b) *Primates and philosophers: How morality evolved*. Princeton University Press. [aMT]
- de Waal, F. (2008) *The ape and the sushi master: Cultural reflections of a primatologist*. Basic Books. [HMM]
- de Waal, F. B. M. & van Roosmalen, A. (1979) Reconciliation and consolation among chimpanzees. *Behavioral Ecology and Sociobiology* 5(1):55–66. [RBR]
- Diesendruck, G. & Deblinger-Tangi, R. (2014) The linguistic construction of social categories in toddlers. *Child Development* 85(1):114–23. <https://doi.org/10.1111/cdev.12130>. [LC]
- Diesendruck, G., Markson, L. & Bloom, P. (2003) Children's reliance on creator's intent in extending names for artifacts. *Psychological Science* 14(2):164–68. <https://doi.org/10.1111/1467-9280.t01-1-01436>. [MR]
- DiFonzo, N. & Bordia, P. (2007) Rumor, gossip and urban legends. *Diogenes* 54(1):19–35. [HMM]
- Du Bois, J. W. (2007) The stance triangle. In: *Stancetaking in discourse: Subjectivity, evaluation, interaction*, ed. R. Englebretson, pp. 139–82. John Benjamins. [ZL]
- Du Bois, J. W. (2014) Towards a dialogic syntax. *Cognitive Linguistics* 25(3):359–410. doi:10.1515/cog-2014-0024. [ZL]
- Du Bois, J. W., Hobson, R. P. & Hobson, J. A. (2014) Dialogic resonance and intersubjective engagement in autism. *Cognitive Linguistics* 25(3):411–41. [ZL]
- Dunfield, K. A., Kuhlmeier, V. A. & Murphy, L. (2013) Children's use of communicative intent in the selection of cooperative partners. *PLoS one* 8:e61804. doi:10.1371/journal.pone.0061804. [PHL]
- Dungan, J. A., Young, L. & Waytz, A. (2019) The power of moral concerns in predicting whistleblowing decisions. *Journal of Experimental Social Psychology* 85:103848. [AS]
- Dunham, Y. (2018) Mere membership. *Trends in Cognitive Sciences* 22(9):780–93. <https://doi.org/10.1016/j.tics.2018.06.004>. [MR, aMT]
- Dunham, Y., Baron, A. S. & Banaji, M. R. (2008) The development of implicit intergroup cognition. *Trends in Cognitive Sciences* 12(7):248–53. [aMT]
- Dunham, Y., Baron, A. S. & Carey, S. (2011) Consequences of “minimal” group affiliations in children. *Child Development* 82(3):793–811. <https://doi.org/10.1111/j.1467-8624.2011.01577.x>. [MR]

- Dunn, J. (1987) The beginnings of moral understanding: Development in the second year. In: *The emergence of morality in young children*, ed. J. Kagan & S. Lamb, pp. 91–111. University of Chicago Press. [KB]
- Dunn, J. & Munn, P. (1986) Siblings and the development of prosocial behavior. *International Journal of Behavioral Development* 9:265–84. [KB]
- Dunning, D., Anderson, J. E., Schlösser, T., Ehlebracht, D. & Fetchenhauer, D. (2014) Trust at zero acquaintance: More a matter of respect than expectation of reward. *Journal of Personality and Social Psychology* 107(1):122–41. [DD, HMM]
- Dunning, D., Fetchenhauer, D. & Schlösser, T. (2012) Trust as a social and emotional act: Noneconomic considerations in trust behavior. *Journal of Economic Psychology* 33(3):686–94. [DD]
- Dunning, D., Fetchenhauer, D. & Schlösser, T. (2016) The psychology of respect: A case study of how behavioral norms regulate human action. In: *Advances in motivation science*, vol. 3, ed. A. Elliot, pp. 1–34. Elsevier. [DD]
- Dunning, D., Fetchenhauer, D. & Schlösser, T. (2019) Why people trust: Solved puzzles and open mysteries. *Current Directions in Psychological Science* 28(4):366–71. [DD]
- Elenbaas, L. (2019) Interwealth contact and young children's concern for equity. *Child Development* 90:108–16. doi:10.1111/cdev.13157. [MK]
- Engelmann, J. M., Haux, L. M. & Herrmann, E. (2019) Helping in young children and chimpanzees shows partiality towards friends. *Evolution and Human Behavior* 40(3):292–300. [RBR]
- Engelmann, J. M., Herrmann, E. & Tomasello, M. (2012) Five-year olds, but not chimpanzees, attempt to manage their reputations. *PLOS ONE* 7(10):e48433. [aMT]
- 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):952–58. [aMT]
- Engelmann, J. M., 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. [aMT]
- Engelmann, J. M., Rapp, D., Herrmann, E. & Tomasello, M. (2018) Concern for group reputation increases prosociality in young children. *Psychological Science* 29:181–90. [aMT]
- Engelmann, J. M. & Tomasello, M. (2019) Children's sense of fairness as a sense of equal respect. *Trends in the Cognitive Sciences* 23(6):454–63. [aMT]
- Engelmann, J. M. & Tomasello, M. (2018) The middle step: Joint intentionality as a human-unique form of second-personal engagement. In: *The Routledge handbook of collective intentionality*, ed. M. Jankovic & K. Ludwig, pp. 433–46. Routledge. [MEB]
- Engh, A. L., Beehner, J. C., Bergman, T. J., Whitten, P. L., Hoffman, R. R., Seyfarth, R. M. & Cheney, D. L. (2006) Behavioral and hormonal responses to predation in female chacma baboons (*Papio hamadryas ursinus*). *Proceedings of the Royal Society B* 273(1587):707–12. [RBR]
- Everett, J. A., Faber, N. S., Savulescu, J. & Crockett, M. J. (2018) The costs of being consequentialist: Social inference from instrumental harm and impartial beneficence. *Journal of Experimental Social Psychology* 79:200–16. [AS]
- Fehr, E., Bernhard, H. & Rockenbach, B. (2008) Egalitarianism in young children. *Nature* 454(7208):1079–83. <https://doi.org/10.1038/nature07155>. [LC, aMT]
- Feinberg, M., Willer, R. & Schultz, M. (2014) Gossip and ostracism promote cooperation in groups. *Psychological Science* 25(3):656–64. [AR]
- Feinberg, M., Willer, R., Stellar, J. & Keltner, D. (2012) The virtues of gossip: Reputational information sharing as prosocial behavior. *Journal of personality and social psychology* 102(5):1015–30. [HMM]
- Fetchenhauer, D. & Dunning, D. (2009) Do people trust too much or too little? *Journal of Economic Psychology* 30(3):263–76. [DD]
- Fetchenhauer, D. & Dunning, D. (2012) Betrayal aversion versus principled trustfulness: How to explain risk avoidance and risky choices in trust games. *Journal of Economic Behavior and Organization* 81(2):534–41. [DD]
- Fetchenhauer, D., Dunning, D. & Schlösser, T. (2017) The mystery of trust: Trusting too much while trusting too little at the same time. In: *Trust in social dilemmas*, ed. P. Van Lange, B. Rockenbach & T. Yamagishi, pp. 139–54. Oxford University Press. [DD]
- Fiske, A. P. (1992) The four elementary forms of sociality: Framework for a unified theory of social relations. *Psychological Review* 99(4):689–723. [MSC, AS]
- Foot, P. (1967) The problem of abortion and the doctrine of double effect. *Oxford Review* 5:1–7. [BF]
- Foster-Hanson, E., Moty, K., Cardarelli, A., Ocampo, J. D. & Rhodes, M. (2019) Developmental changes in strategies for gathering evidence about biological kinds. <https://doi.org/10.31234/osf.io/gm96d>. [MR]
- Foster-Hanson, E. & Rhodes, M. (2019) Is the most representative skunk the average or the stinkiest? Developmental changes in representations of biological categories. *Cognitive Psychology* 110:1–15. <https://doi.org/10.1016/j.cogpsych.2018.12.004>. [MR]
- Foster-Hanson, E., Roberts, S. O., Gelman, S. A. & Rhodes, M. (2018) Categories convey normative information across domains. <https://doi.org/10.31234/osf.io/b37av>. [MR]
- Frankfurt, H. G. (1982) The importance of what we care about. *Synthese* 53:257–90 (reprinted in Frankfurt 1988, pp. 80–94). [KKB-O]
- Frankfurt, H. G. (1988) *The importance of what we care about*. Cambridge University Press. [KKB-O]
- Frankfurt, H. G. (2004) *The reasons of love*. Princeton University Press. [KKB-O]
- Frankfurt, H. G. (2006) *Taking ourselves seriously and getting it right* (ed. D. Satz, with comments by C. M. Korsgaard, M. E. Bratman, and M. Dan-Cohen). Stanford University Press. [KKB-O]
- Fu, G., Heyman, G. D., Chen, G., Liu, P. & Lee, K. (2015) Children trust people who lie to benefit others. *Journal of Experimental Child Psychology* 129 :127–139. [AS]
- Gaither, S., Chen, E., Corriveau, K., Harris, P., Ambady, N. & Sommers, S. (2014) Monoracial and biracial children: Effects of racial identity saliency on social learning and social preferences. *Child Development* 85(6):2299–316. doi:10.1111/cdev.12266. [MK]
- Gelfand, M. J., Chiu, C. & Hong, Y. (2016) *Handbook of advances in culture and psychology*, vol. 6. Oxford University Press. [MK]
- Gelfand, M. J., Harrington, J. R. & Jackson, J. C. (2017) The strength of social norms across human groups. *Perspectives on Psychological Science* 12(5):800–809. [JBH]
- Gelfand, M. J., Raver, J. L., Nishii, L., Leslie, L. M., Lun, J., Lim, B. C., Duan, L., Almaliah, A., Ang, S., Arndt, J., Aycan, Z., Boehnke, K., Boski, P., Cabecinas, R., Chan, D., Chhokar, J., D'Amato, A., Ferrer, M., Fischlmayr, I. C., Fischer, R., Fulup, M., Georgas, J., Kashima, E. S., Kashima, Y., Kim, K., Lempereur, A., Marquez, P., Othman, R., Overlaet, B., Panagiotopoulou, P., Peltzer, K., Perez-Florizno, L. R., Ponomarenko, L., Realo, A., Schei, V., Schmitt, M., Smith, P. B., Soomro, N., Szabo, E., Taveasin, N., Toyama, M., Van de Vliert, E., Vohra, N., Ward, C. & Yamaguchi, S. (2011) Differences between tight and loose cultures: A 33-nation study. *Science* 332(6603):1100–104. [JBH]
- Gergely, G. & Csibra, G. (2003) Teleological reasoning in infancy: The naive theory of rational action. *Trends in Cognitive Science* 7(7):287–92. [CK]
- Gilbert, M. (1989) *On social facts*. Routledge and Kegan Paul. [MG]
- Gilbert, M. (1990) Walking together: A paradigmatic social phenomenon. *Midwest Studies in Philosophy* 15(1):1–14. [MG, aMT]
- Gilbert, M. (1992) *On social facts*. Princeton University Press. [PHL]
- Gilbert, M. (1996) *Living together: Rationality, sociality, and obligation*. Rowman & Littlefield. [MG]
- Gilbert, M. (2003) The structure of the social atom: Joint commitment as the foundation of human social behavior. In: *Social metaphysics*, ed. F. Schmitt, pp. 39–64. Rowman & Littlefield. [MG]
- Gilbert, M. (2006) *A theory of political obligation: Membership, commitment, and the bonds of society*. Oxford University Press. [MG]
- Gilbert, M. (2014) *Joint commitment: How we make the social world*. Oxford University Press. [MG, PP aMT]
- Gilbert, M. (2018a) Remarks on joint commitment and its relation to moral thinking. *Philosophical Psychology* 31(5):755–66. [MG]
- Gilbert, M. (2018b) *Rights and demands: A foundational inquiry*. Oxford University Press. [MG]
- Gneezy, A. & Epley, N. (2014) Worth keeping but not exceeding: Asymmetric consequences of breaking versus exceeding promises. *Social Psychological and Personality Science* 5(7):796–804. [RAA]
- Göckeritz, S., Schmidt, M. F. H. & Tomasello, M. (2014) Young children's creation and transmission of social norms. *Cognitive Development* 30:81–95. [aMT]
- Goldberg, A. E. (2006) *Constructions at work: The nature of generalization in language*. Oxford University Press. [PHL]
- Goodwin, G. P. & Darley, J. M. (2012) Why are some moral beliefs perceived to be more objective than others?. *Journal of Experimental Social Psychology* 48(1):250–56. [RAA]
- Goodwin, M. H. (2006) *The hidden life of girls: Games of stance, status, and exclusion*. Blackwell. [ZL]
- Gopnik, A. (2009) *The philosophical baby: What children's minds tell us about truth, love & the meaning of life*. Random House. [JM]
- Gopnik, A. & Wellman, H. M. (2012) Reconstructing constructivism: Causal models, Bayesian learning mechanisms, and the theory-theory. *Psychological Bulletin* 138(6):1085–108. <https://doi.org/10.1037/a0028044>. [MR]
- Gould, L. (1992) Alloparental care in free-ranging *Lemur catta* at Berenty Reserve, Madagascar. *Folia Primatologica* 58(2):72–83. [RBR]
- Goulding, B. W. & Friedman, O. (2018) The development of territory-based inferences of ownership. *Cognition* 177:142–49. doi:10.1016/j.cognition.2018.04.013. [MK]
- Gräfenhain, M., Behne, T., Carpenter, M. & Tomasello, M. (2009) Young children's understanding of joint commitments. *Developmental Psychology* 45(5):1430–33. [KB, aMT]
- Gräfenhain, M., Carpenter, M. & Tomasello, M. (2013) Three-year-olds' understanding of the consequences of joint commitments. *PLOS ONE* 8(9):e73039. [aMT]
- Graham, J., Haidt, J., Koleva, S., Motyl, M., Iyer, R., Wojcik, S. P. & Ditto, P. H. (2013) Moral foundations theory: The pragmatic validity of moral pluralism. In: *Advances in experimental social psychology*, vol. 47, pp. 55–130. Academic Press. [BF]
- Grassmann, S. & Tomasello, M. (2010) Young children follow pointing over words in interpreting acts of reference. *Developmental Science* 13(1):252–63. [PHL]
- Greenberg, J. R., Hamann, K., Warneken, F. & Tomasello, M. (2010) Chimpanzee helping in collaborative and noncollaborative contexts. *Animal Behaviour* 80(5):873–80. [aMT]

- Greene, J. D. (2008) The secret joke of Kant's soul. *Moral Psychology 3*: 35–79. [AS]
- Greene, J. D. (2013) *Moral tribes: Emotion, reason, and the gap between us and them*. Atlantic Books. [BF]
- Grice, H. P. (1975) Logic and conversation. In: *Syntax and semantics, vol. 3: Speech acts*, ed. P. Cole & J. L. Morgan, pp. 41–58. Academic Press. [JIMC]
- Grocke, P., Rossano, F. & Tomasello, M. (2015) Procedural justice in children: Preschoolers accept unequal resource distributions if the procedure provides equal opportunities. *Journal of Experimental Child Psychology 140*:197–210. [aMT]
- Grocke, 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. [aMT]
- Grosse, G., Behne, T., Carpenter, M. & Tomasello, M. (2010) Infants communicate in order to be understood. *Developmental Psychology 46*(6):1710–22. [PHL]
- Guglielmo, S. & Malle, B. F. (2019) Asymmetric morality: Blame is more differentiated and more extreme than praise. *PLOS ONE 14*(3):e0213544. <https://doi.org/10.1371/journal.pone.0213544>. [RAA]
- Habermas, J. (1990) *Moral consciousness and communicative action*. MIT Press. (Original work published in 1983). [JIMC]
- Hage, P. & Harary, F. (1996) *Island networks: Communication, kinship, and classification structures in Oceania*. Cambridge University Press. [AB]
- Haidt, J. & Baron, J. (1996) Social roles and the moral judgement of acts and omissions. *European Journal of Social Psychology 26*(2):201–18. [RAA, MSC, JM]
- Hall, M. S., Franks, B., Bauer, M. & Bangerter, A. (2019, in preparation) The coexistence of commonsense and conspiracist views in conspiratorial worldviews. Department of Psychological and Behavioural Science, London School of Economics. [BF]
- Hamann, K., Warneken, F., Greenberg, J. R. & Tomasello, M. (2011) Collaboration encourages equal sharing in children but not in chimpanzees. *Nature 476* (7360):328–31. doi:10.1038/nature10278. [JC, aMT]
- Hamann, K., Warneken, F. & Tomasello, M. (2012) Children's developing commitments to joint goals. *Child Development 83*(1):137–45. [aMT]
- Hamlin, J. K. (2013) Moral judgment and action in preverbal infants and toddlers: Evidence for an innate moral core. *Current Directions in Psychological Science 22* (3):186–93. doi:10.1177/0963721412470687. [RBR, RAT]
- Hardecker, S., Schmidt, M. & Tomasello, M. (2017) Children's developing understanding of the conventionality of rules. *Cognition and Development 18*(2):163–88. [aMT]
- Harris, P. L., Koenig, M. A., Corriveau, K. H. & Jaswal, V. K. (2017) Cognitive foundations of learning from testimony. *Annual Review of Psychology 69*(1):251–73. [PHL]
- Haun, D. & Over, H. (2014) Like me: A homophily-based account of human culture. In: *Cultural evolution*, ed. P. J. Richerson & M. Christiansen, pp. 75–85. MIT Press. [aMT]
- Haun, D. B. M. & Tomasello, M. (2011) Conformity to peer pressure in preschool children. *Child Development 82*(6):1759–67. [aMT]
- Haward, P., Wagner, L., Carey, S. & Prasada, S. (2018) The development of principled connections and kind representations. *Cognition 176*:255–68. <https://doi.org/10.1016/j.cognition.2018.02.001>. [LC, MR]
- Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Gintis, H., McElreath, R., Alvard, M., Barr, A., Ensminger, J., Henrich, N. S., Hill, K., Gil-White, F., Gurven, M., Marlowe, F. W., Patton, J. Q. & Tracer, D. (2005) "Economic man" in cross-cultural perspective: Behavioral experiments in 15 small-scale societies. *Behavioral and Brain Sciences 28*:795–855. [AB]
- Henrich, J., Ensminger, J., McElreath, R., Barr, A., Barrett, C., Bolyanatz, A., Cardenas, J. C., Gurven, M., Gwako, E., Henrich, N., Lesorogol, C., Marlowe, F., Tracer, D. & Ziker, J. (2010) Markets, religion, community size, and the evolution of fairness and punishment. *Science 327*(5972):1480–84. doi:10.1126/science.1182238. [AB, DD]
- Hetherington, C., Hendrickson, C. & Koenig, M. (2014) Reducing an in-group bias in preschool children: The impact of moral behavior. *Developmental Science 17* (6):1042–49. [PHL]
- Higgins, E. T. (1987) Self-discrepancy: A theory relating self and affect. *Psychological Review 94*(3):319–40. [DD]
- Hill, K. (2002) Altruistic cooperation during foraging by the Ache, and the evolved human predisposition to cooperate. *Human Nature 13*(1):105–28. [rMT]
- Hill, K. R., Walker, R. S., Božičević, M., Eder, J., Headland, T., Hewlett, B., Hurtado, A. M., Marlowe, F., Wiessner, P. & Wood, A. (2011) Co-residence patterns in hunter-gatherer societies show unique human social structure. *Science 331*:1286–89. [BF]
- Hinchman, E. (2005) Telling as inviting to trust. *Philosophy and Phenomenological Research 70*(3):562–87. [PHL]
- Hirsh, J. B. & Kang, S. K. (2016) Mechanisms of identity conflict: Uncertainty, anxiety, and the behavioral inhibition system. *Personality and Social Psychology Review 20* (3):223–44. [JBH]
- Hirsh, J. B., Lu, J. G. & Galinsky, A. D. (2018) Moral utility theory: Understanding the motivation to behave (un)ethically. *Research in Organizational Behavior 38*(1):43–59. [JBH]
- Hogh-Olesen, H. (Ed.) (2009) *Human morality and sociality: Evolutionary and comparative perspectives*. Palgrave Macmillan. [AK]
- Hohfeld, W. N. (1923) *Fundamental legal concepts as applied in judicial reasoning and legal essays*. Yale University Press. [aMT]
- Holton, R. (1994) Deciding to trust, coming to believe. *Australasian Journal of Philosophy 72*(1):63–76. [PHL]
- House, B. & Tomasello, M. (2018) Modeling social norms increasingly influences costly sharing in middle childhood. *Journal of Experimental Child Psychology 171*:84–98. [aMT]
- 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. [aMT]
- 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*:14586–591. [aMT]
- Hrdy, S. B. (2009) *Mothers and others: The evolutionary origins of mutual understanding*. Harvard University Press. [RAT]
- Hume, D. (1751/1957) *An enquiry concerning the principles of morals*. Bobbs-Merrill. [aMT]
- Ibbotson, P. (2014) Little dictators: A developmental meta-analysis of prosocial behavior. *Current Anthropology 55*(6):814–21. [aMT]
- Janoff-Bulman, R., Sheikh, S. & Hepp, S. (2009) Proscriptive versus prescriptive morality: Two faces of moral regulation. *Journal of Personality and Social Psychology 96*(3):521–37. [RAA]
- Jaswal, V. K. (2004) Don't believe everything you hear: Preschoolers' sensitivity to speaker intent in category induction. *Child Development 75*(6):1871–85. [PHL]
- Jensen, K., Call, J. & Tomasello, M. (2007) Chimpanzees are rational maximizers in an ultimatum game. *Science 318*(5847):107–09. [aMT]
- Jin, K.-S. & Baillargeon, R. (2017) Infants possess an abstract expectation of in-group support. *Proceedings of the National Academy of Sciences 114*(31):8199–04. <https://doi.org/10.1073/pnas.1706286114>. [LC]
- 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*:1691–703. [aMT]
- 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. [aMT]
- Kachel, U. & Tomasello, M. (2019) 3- and 5-year-old children's adherence to explicit and implicit joint commitments. *Developmental Psychology 55*(1):80–8. [arMT]
- Kalish, C. W. (2005) Becoming status conscious: Children's appreciation of social reality. *Philosophical Explorations 8*(3):245–63. [CK]
- Kalish, C. W. & Lawson, C. A. (2008) Development of social category representations: Early appreciation of roles and deontic relations. *Child Development 79*:577–93. [JM]
- Kang, S. K. & Bodenhausen, G. V. (2015) Multiple identities in social perception and interaction: Challenges and opportunities. *Annual Review of Psychology 66*(1):547–74. [JBH]
- Kanngieser, P., Rossano, F. & Tomasello, M. (2019) Children, but not great apes, respect ownership. *Developmental Science 23*(1):e12842. [aMT]
- Kaplan, H., Hill, K., Cadellina, R. V., Hayden, B., Hyndman, D. C., Preston, R. J., Smith, E. A., Stuart, D. E. & Yesner, D. R. (1985) Food sharing among ache foragers: Tests of explanatory hypotheses. *Current Anthropology 26*(2):223–46. [AK]
- Kaufmann, A. (2015) Animal mental action: Planning among chimpanzees. *Review of Philosophy and Psychology 6*(4):745–60. [AK]
- Kaufmann, A. (2017) Joint distal intentions: Who shares what? In: *Routledge handbook of philosophy of the social mind*, ed. J. Kiverstein, pp. 343–56. Routledge. [AK]
- Kelley, H. H. (1982) *Personal relationships: Their structures and processes*. Psychology Press. [MSC]
- Kelley, H. H., Holmes, J. G., Kerr, N. L., Reis, H. T., Rusbult, C. E. & Van Lange, P. A. M. (2003) *An atlas of interpersonal situations*. Cambridge University Press. [MSC]
- Ketelaer, T. & Au, W. T. (2003) The effects of feelings of guilt on the behaviour of uncooperative individuals in repeated social bargaining games: An affect-as-information interpretation of the role of emotion in social interaction. *Cognition and Emotion 17*(3):429–53. [AR]
- Keupp, S., Behne, T. & Rakoczy, H. (2013) Why do children over-imitate? Normativity is crucial. *Journal of Experimental Child Psychology 116*:392–406. [aMT]
- Killen, M. (2007) Children's social and moral reasoning about exclusion. *Current Directions in Psychological Science 16*(1):32–6. doi:10.1111/j.1467-8721.2007.00470.x. [MK]
- Killen, M. & Smetana, J. G. (2015) Origins and development of morality. In: *vol. 3, Handbook of child psychology and developmental science* (7th edition), ed. R. M. Lerner & M. Lamb, pp. 701–49. Wiley-Blackwell. <https://psycnet.apa.org/record/2015-15586-017>. [MK]
- Kinzler, D. K. & Shutts, K. (2008) Memory for "mean" over "nice": The influence of threat on children's face memory. *Cognition 107*:775–83. [RAA]
- Kinzler, K. D., Dupoux, E. & Spelke, E. S. (2007) The native language of social cognition. *Proceedings of the National Academy of Sciences 104*(30):12577–80. <https://doi.org/10.1073/pnas.0705345104>. [LC, ZL]

- Klein, N. & Epley, N. (2014) The topography of generosity: Asymmetric evaluations of prosocial actions. *Journal of Experimental Psychology: General* **143**(6):2366–79. [RAA]
- Knobe, J. (2003a) Intentional action and side effects in ordinary language. *Analysis* **63** (279):190–94. [RAA]
- Knobe, J. (2003b) Intentional action in folk psychology: An experimental investigation. *Philosophical Psychology* **16**(2):309–24. [RAA]
- Kochanska, G. (2002) Mutually responsive orientation between mothers and their young children: A context for the early development of conscience. *Current Directions in Psychological Science* **11**(6):191–95. doi:10.1111/1467-8721.00198. [RAT]
- Kochanska, G., Boldt, L. J. & Goffin, K. C. (2019) Early relational experience: A foundation for the unfolding dynamics of parent-child socialization. *Child Development Perspectives* **13**(1):41–7. doi:10.1111/cdep.12308. [RAT]
- Kochanska, G., Coy, K. C. & Murray, K. T. (2001) The development of self-regulation in the first four years of life. *Child Development* **72**:1091–111. [RAA]
- Koenig, M. A. & Echols, C. H. (2003) Infants' understanding of false labeling events: The referential roles of words and the speakers who use them. *Cognition* **87**(3):179–208. [PHL]
- Kohlberg, L. (1969) Stage and sequence: The cognitive developmental approach to socialization. In: *Handbook of socialization theory and research*, ed. D. Goslin, pp. 347–480. Rand McNally. [RAT]
- Kohlberg, L. (1971) From is to ought: How to commit the naturalistic fallacy and get away with it in the study of moral development. In: *Psychology and genetic epistemology*, ed. T. Mischel, pp. 347–480. Academic Press. [MK]
- Korsgaard, C. (1996) *The sources of normativity*. Cambridge University Press. [MK, arMT]
- Koymen, B., Schmidt, M., Lieven, E. & Tomasello, M. (2015) Teaching versus enforcing norms in preschoolers' peer interactions. *Journal of Experimental Child Psychology* **135**:93–101. [aMT]
- Köymen, S. B. & Kyrtatzis, A. (2014) Dialogic syntax and complement constructions in toddlers' peer interactions. *Cognitive Linguistics* **25**(3):1497–521. doi:10.1515/cog-2014-0028. [ZL]
- Krebs, D. L. (2008) Morality: An evolutionary account. *Perspectives on Psychological Science* **3**(3):149–72. <https://doi.org/10.1111/j.1745-6924.2008.00072.x>. [SB-D]
- Kushnir, T. & Koenig, M. A. (2017) What I don't know won't hurt you: The relation between professed ignorance and later knowledge claims. *Developmental Psychology* **53**(5):826–35. [PHL]
- Lawler, E. J. (2001) An affective theory of social exchange. *American Journal of Sociology* **107**(2):321–52. [DD]
- Lazare, A. (2004) *On apology*. Oxford University Press. [RAA]
- Legare, C. H. & Gelman, S. A. (2008) Bewitchment, biology, or both: The co-existence of natural and supernatural explanatory frameworks across development. *Cognitive Science* **32**(4):607–42. [BF]
- Lemay, E. P., Overall, N. C. & Clark, M. S. (2010) Experiences and interpersonal consequences of hurt and anger. *Journal of Personality and Social Psychology* **103**(6):982–1006. [MSC]
- Lerner, J. S. & Tetlock, P. E. (1999) Accounting for the effects of accountability. *Psychological Bulletin* **125**(2):255–75. [JBH]
- Leslie, A. M., Knobe, J. & Cohen, A. (2006) Acting intentionally and the side-effect effect: Theory of mind and moral judgment. *Psychological Science* **17**(5):421–27. [RAA]
- Levine, E. E. & Schweitzer, M. E. (2014) Are liars ethical? On the tension between benevolence and honesty. *Journal of Experimental Social Psychology* **53** :107–117. [AS]
- Levine, S., Mikhail, J. & Leslie, A. M. (2018) Presumed innocent? How tacit assumptions of intentional structure shape moral judgment. *Journal of Experimental Psychology: General* **147** :1728–47. [AS]
- Liberman, Z., Woodward, A. L. & Kinzler, K. D. (2017a) The origins of social categorization. *Trends in cognitive sciences* **21**(7):556–68. [ZL]
- Liberman, Z., Woodward, A. L. & Kinzler, K. D. (2017b) Preverbal infants infer third-party social relationships based on language. *Cognitive Science* **41**(3):622–34. [ZL]
- Liberman, Z., Woodward, A. L., Sullivan, K. R. & Kinzler, K. D. (2016) Early emerging system for reasoning about the social nature of food. *Proceedings of the National Academy of Sciences* **113**(34):9480–85. [ZL]
- Lyons, D. E., Young, A. G. & Keil, F. C. (2007) The hidden structure of overimitation. *Proceedings of the National Academy of Sciences* **104**(50):19751–56. [aMT]
- Maccoby, E. E. (1984) Socialization and developmental change. *Child Development* **55** (2):317–28. doi:10.2307/1129945. [RAT]
- MacWhinney, B. (2000) *The CHILDES project: Tools for analyzing talk, vol. 2, The database* (3rd edition). Erlbaum. [KB]
- 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**(2):254–62. [KB, aMT]
- Maranges, H. M. & Ainsworth, S. E. (2020) Moralization of obesity: Social error management concerns about self-control and cooperation. Unpublished manuscript. Florida State University, Tallahassee, FL. [HMM]
- Marshall, J., Mermin-Bunnell, N., Gollwitzer, A., Shinomiya, M., Retelsdorf, J. & Bloom, P. (in preparation) The developmental foundations of relational obligations across 5 societies. [JM, rMT]
- Marshall, J., Wynn, K. & Bloom, P. (in press) Do children and adults take social relationship into account when evaluating other peoples' actions? *Child Development*. [JM]
- McGraw, A. P. & Tetlock, P. E. (2005) Taboo trade-offs, relational framing, and the acceptability of exchanges. *Journal of Consumer Psychology* **15**(1):2–15. [MSC]
- McMyler, B. (2011) *Testimony, trust, and authority*. Oxford University Press. [PHL]
- Melis, A. P., Floedl, A. & Tomasello, M. (2015) Non-egalitarian allocations among preschool peers in a face-to-face bargaining task. *PLOS ONE* **10**(3):e0120494. [aMT]
- Melis, A. P., Hare, B. & Tomasello, M. (2006) Engineering cooperation in chimpanzees: Tolerance constraints on cooperation. *Animal Behaviour* **72**(2):275–86. [aMT]
- 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):485–93. [aMT]
- Miller, J. G. & Bersoff, D. M. (1992) Culture and moral judgment: How are conflicts between justice and interpersonal responsibilities resolved? *Journal of Personality and Social Psychology* **62**(4):541–54. [EEB, AS]
- Miller, J. G., Bersoff, D. M. & Harwood, R. L. (1990) Perceptions of social responsibilities in India and in the United States: Moral imperatives or personal decisions? *Journal of Personality and Social Psychology* **58**(1):33–47. [EEB, JM]
- Miller, J. G., Das, R. & Chakravarthy, S. (2011) Culture and the role of choice in agency. *Journal of Personality and Social Psychology* **101**(1):46–61. <https://doi.org/10.1037/a0023330>. [EEB]
- Misch, A., Over, H. & Carpenter, M. (2014, October) Stick with your group: Young children's attitudes about group loyalty. *Journal of Experimental Child Psychology* **126**:19–36. <https://doi.org/10.1016/j.jecp.2014.02.008>. [LC, aMT]
- Mitani, J. C., Merriwether, D. A. & Zhang, C. (2000) Male affiliation, cooperation and kinship in wild chimpanzees. *Animal Behaviour* **59**(4):885–93. [RBR]
- Moore, C. & Barresi, J. (2017) The role of second-person information in the development of social understanding. *Frontiers in Psychology* **8**. doi:10.3389/fpsyg.2017.01667. [JC]
- Moran, R. (2005) Problems of sincerity. *Proceedings of the Aristotelian Society* **105** (1):325–45. [PHL]
- Mulvey, K. L. (2016a) Children's reasoning about social exclusion: Balancing many factors. *Child Development Perspectives* **10**(1):22–7. doi:10.1111/cdep.12157. [MK]
- Mulvey, K. L. (2016b) Evaluations of moral and conventional intergroup transgressions. *British Journal of Developmental Psychology* **34**(4):489–501. doi:10.1111/bjdp.12145. [MK]
- Nagel, T. (1970a) *The possibility of altruism*. Clarendon Press. [HR]
- Nagel, T. (1970b) *The possibility of altruism*. Oxford University Press. [arMT]
- Nesdale, D. (2004) Social identity processes and children's ethnic prejudice. In: *The development of the social self*, ed. M. Bennett & F. Sani, pp. 219–45. Psychology Press. [MK]
- Nesse, R. M. & Ellsworth, P. C. (2009) Evolution, emotions, and emotional disorders. *American Psychologist* **64**(2):129–39. <https://doi.org/10.1037/a0013503>. [SB-D]
- Newton, E. K., Thompson, R. A. & Goodman, M. (2016) Individual differences in toddlers' prosociality: Experiences in early relationships explain variability in prosocial behavior. *Child Development* **87**(6):1715–26. doi:10.1111/cdev.12631. [RAT]
- Norenzayan, A., Shariff, A. F., Gervais, W. M., Willard, A. K., McNamara, R. A., Slingerland, E. & Henrich, J. (2016) The cultural evolution of prosocial religions. *Behavioral and Brain Sciences* **39**, E1. <https://doi.org/10.1017/S0140525X14001356>. [EEB]
- Nowak, M. (2006) Five rules for the evolution of cooperation. *Science* **314**:1560–63. [PP]
- Noyes, A. & Dunham, Y. (2017) Mutual intentions as a causal framework for social groups. *Cognition* **162**:133–42. [aMT]
- Nucci, L. & Weber, E. K. (1995) Social interactions in the home and the development of young children's conceptions of the personal. *Child Development* **66**(5):1438–52. [MK]
- Ohtsubo, Y. (2007) Perceived intentionality intensifies blameworthiness of negative behaviors: Blame-praise asymmetry in intensification effect. *Japanese Psychological Research* **49**:100–10. [RAA]
- Over, H. (2018, April) The influence of group membership on young children's prosocial behaviour. *Current Opinion in Psychology* **20**:17–20. <https://doi.org/10.1016/j.copsyc.2017.08.005>. [LC, aMT]
- Over, H., Vaish, A. & Tomasello, M. (2016, Oct.–Dec.) Do young children accept responsibility for the negative actions of ingroup members? *Cognitive Development* **40**:24–32. <https://doi.org/10.1016/j.cogdev.2016.08.004>. [LC, aMT]
- Parfit, D. (2011) *On what matters: Vol. I*. Oxford University Press. [DO]
- Pea, R. D. (1982) Origins of verbal logic: Spontaneous denials by two- and three-year olds. *Journal of Child Language*, **9**(3):597–626. [PHL]
- Perner, J. & Roessler, J. (2010) Teleology and causal reasoning in children's theory of mind. In: *Causing human action: New perspectives on the causal theory of action*, ed. J. Aguilar & A. Buckareff, pp. 199–228. MIT Press. [CK]
- Pesch, A. & Koenig, M. A. (2018) Varieties of trust in preschoolers' learning and practical decisions. *PLOS ONE* **13**(8):e0202506. [PHL]
- Petersen, G. (2000) Indigenous island empires: Yap and Tonga considered. *Journal of Pacific History* **35**:5–27. [AB]
- Pettit, P. (2018a) *The birth of ethics: Reconstructing the role and nature of morality*. Oxford University Press. [rMT, PP]

- Pettit, P. (2018b) Naturalizing Tomasello's history of morality. *Philosophical Psychology* 31:722–35. [PP]
- Piaget, J. (1932) *Moral judgment of the child*. Free Press. (1965; original work published in 1932). [MK, aMT, RAT, JIMC]
- Piaget, J. (1965) *The moral judgment of the child*. Free Press. (Original work published in 1932). [JIMC]
- Pizarro, D., Uhlmann, E. & Salovey, P. (2003) Asymmetry in judgments of moral blame and praise: The role of perceived metadesires. *Psychological Science* 14(3):267–72. [RAA]
- Powell, L. J. & Spelke, E. S. (2013) Preverbal infants expect members of social groups to act alike. *Proceedings of the National Academy of Sciences* 110(41):E3965–72. [ZL]
- Prasada, S. & Dillingham, E. M. (2009) Representation of principled connections: A window onto the formal aspect of common sense conception. *Cognitive Science* 33(3):401–48. <https://doi.org/10.1111/j.1551-6709.2009.01018.x>. [LC]
- Quinn, P. C., Yahr, J., Kuhn, A., Slater, A. M. & Pascalis, O. (2002) Representation of the gender of human faces by infants: A preference for female. *Perception* 31(9):1109–21. <https://doi.org/10.1068/p3331>. [LC]
- Rai, S. T. & Fiske, A. P. (2011) Moral psychology is relationship regulation: Moral motives for unity, hierarchy, equality, and proportionality. *Psychological Review* 118(1):57–75. [MSC]
- 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. [aMT]
- Rakoczy, H. & Tomasello, M. (2009) Done wrong or said wrong? Young children understand the normative directions of fit of different speech acts. *Cognition* 113(2):205–12. [PHL]
- Rakoczy, H., Warneken, F. & Tomasello, M. (2007) "This way!", "No! That way!" – 3-year-olds know that two people can have mutually incompatible desires. *Cognitive Development* 22(1):47–68. [CK]
- 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):875–81. [KB, aMT]
- Reddy, R. B. & Mitani, J. C. (2019) Social relationships and caregiving behavior between recently orphaned chimpanzee siblings. *Primates* 60(5):389–400. [RBR]
- Reis, H. T. (2008) Reinvigorating the concept of situation in social psychology. *Personality and Social Psychology Review* 12(4):311–29. [MSC]
- Rekers, Y., Haun, D. & Tomasello, M. (2011) Children, but not chimpanzees, prefer to forage collaboratively. *Current Biology* 21:1756–58. [KB, aMT]
- Rest, J. R. (1986) *Moral development: Advances in research and theory*. Praeger. [JBH]
- Rhodes, M. (2012) Naïve theories of social groups. *Child Development* 83(6):1900–16. <https://doi.org/10.1111/j.1467-8624.2012.01835.x>. [LC, MR, aMT]
- Rhodes, M. (2013) How two intuitive theories shape the development of social categorization. *Child Development Perspectives* 7(1):12–16. <https://doi.org/10.1111/cdep.12007>. [LC]
- Rhodes, M. (2014) Children's explanations as a window into their intuitive theories of the social world. *Cognitive Science* 38:1687–97. [aMT]
- Rhodes, M. & Chalik, L. (2013) Social categories as markers of intrinsic interpersonal obligations. *Psychological Science* 24(6):999–1006. doi:10.1177/0956797612466267. [MK, MR, aMT]
- Rhodes, M., Leslie, S. J., Bianchi, L. & Chalik, L. (2018) The role of generic language in the early development of social categorization. *Child Development* 89(1):148–55. <https://doi.org/10.1111/cdev.12714>. [LC]
- Rhodes, M. & Wellman, H. M. (2017, October) Moral learning as intuitive theory revision. *Cognition* 167:191–200. [RBR, aMT]
- Richerson, P., Baldini, R., Bell, A. V., Demps, K., Frost, K., Hillis, V., Mathew, S., Newton, E. K., Naar, N., Newson, L., Ross, C., Smaldino, P. E., Waring, T. M. & Zefferman, M. (2016) Cultural group selection plays an essential role in explaining human cooperation: A sketch of the evidence. *Behavioral and Brain Sciences* 39:1–68. [BF]
- Rizzo, M. T., Cooley, S., Elenbaas, L. & Killen, M. (2018) Young children's inclusion decisions in moral and social-conventional group norm contexts. *Journal of Experimental Child Psychology* 165:19–36. doi:10.1016/j.jecp.2017.05.006. [MK]
- Roberts, S. O., Gelman, S. A. & Ho, A. K. (2017) So it is, so it shall be: Group regularities license children's prescriptive judgments. *Cognitive Science* 41:576–600. <https://doi.org/10.1111/cogs.12443>. [MR]
- Roccas, S. & Brewer, M. B. (2002) Social identity complexity. *Personality and Social Psychology Review* 6(2):88–106. [JBH]
- Rochat, P., Dias, M. D., Liping, G., Broesch, T., Passos-Ferreira, C., Winning, A. & Berg, B. (2009) Fairness in distributive justice by 3- and 5-year-olds across seven cultures. *Journal of Cross-Cultural Psychology* 40:416–42. [AS]
- Rochat, P., Robbins, E., Passos-Ferreira, C., Donato Oliva, A., Dias, M. D. G. & Guo, L. (2014) Ownership reasoning in children across cultures. *Cognition* 132(3):471–84. doi:10.1016/j.cognition.2014.04.014. [MK]
- Roese, N. J. & Olson, J. M. (1997) Counterfactual thinking: The intersection of affect and function. *Advances in Experimental Social Psychology* 29:1–59. [RAA]
- Ronfard, S., Nelson, L., Dunham, Y. & Blake, P. R. (2019) How children use accuracy information to infer informant intentions and to make reward decisions. *Journal of Experimental Child Psychology* 177:100–18. doi:10.1016/j.jecp.2018.07.017. [PHL]
- Ross, A. (1986) Why do we believe what we are told? *Ratio* 28(1):69–88. [PHL]
- Ross, H. & den Bak-Lammers, I. (1998) Consistency and change in children's tattling on their siblings: Children's perspectives on the moral rules and procedures in family life. *Social Development* 7:275–300. [KB]
- Ross, H. S., Friedman, O. & Field, A. (2014) Toddlers assert and acknowledge ownership rights. *Social Development* 24(2):341–56. <https://doi.org/10.1111/sode.12101>. [MK]
- Rotella, A., Jung, J., Chinn, C. & Barclay, P. (in preparation) Observation and moral ambiguity matter: A meta-analysis on moral licensing. [AR]
- Rottman, J. & Kelemen, D. (2012) Aliens behaving badly: Children's acquisition of novel purity-based morals. *Cognition* 124(3):356–60. <https://doi.org/10.1016/j.cognition.2012.06.001>. [SB-D]
- Rottman, J., Young, L. & Kelemen, D. (2017) The impact of testimony on children's moralization of novel actions. *Emotion* 17(5):811–27. <https://doi.org/10.1037/emo0000276>. [SB-D]
- Rousseau, D. M., Sitkin, S. B., Burt, R. & Camerer, C. (1998) Not so different after all: A cross-discipline view of trust. *Academy of Management Review* 23(3):393–404. [DD]
- Rutland, A. & Killen, M. (2015) A developmental science approach to reducing prejudice and social exclusion: Intergroup processes, social-cognitive development, and moral reasoning. *Social Issues and Policy Review* 9(1):121–54. doi:10.1111/sjpr.12012. [MK]
- Rutland, A., Mulvey, K. L., Hitti, A., Abrams, D. & Killen, M. (2015) When does the in-group like the out-group? Bias among children as a function of group norms. *Psychological Science* 26(6):834–42. doi:10.1177/0956797615572758. [MK]
- Scanlon, T. (1998) *What we owe to each other*. Belknap Press of Harvard University Press. [aMT]
- Schaefer, M., Haun, D. & Tomasello, M. (2015) Fair is not fair everywhere. *Psychological Science* 26:1252–1260. [aMT]
- Schlösser, T., Fetschenhauer, D. & Dunning, D. (2016) Against all odds? The emotional dynamics underlying trust. *Decision* 3(3):216–30. [DD]
- Schlösser, T., Mensching, O., Dunning, D. & Fetschenhauer, D. (2015) Trust and rationality: Shifting normative analyses in risks involving other people versus nature. *Social Cognition* 33(5):459–82. doi:10.1521/soco.2015.33.5.459. [DD]
- Schmidt, M. F. H., Butler, L. P., Heinz, J. & Tomasello, M. (2016a) Young children see a single action and infer a social norm. *Psychological Science* 27:1360–70. [aMT]
- Schmidt, M. F. H., Cabrera, I. & Tomasello, M. (2017) Children's developing metaethical judgments. *Journal of Experimental Child Psychology* 164:163–77. [aMT]
- Schmidt, M. F. H., Rakoczy, H. & Tomasello, M. (2011) Young children attribute normativity to novel actions without pedagogy or normative language. *Developmental Science* 14:530–39. [aMT]
- Schmidt, M. F. H., Rakoczy, H. & Tomasello, M. (2012) Young children enforce social norms selectively depending on the violator's group affiliation. *Cognition* 124(3):325–33. [HR, aMT]
- Schmidt, M. F. H., Rakoczy, H. & Tomasello, M. (2016b) Young children understand the role of agreement in establishing arbitrary norms – but unanimity is key. *Child Development* 87(2):612–26. [aMT]
- Schmidt, M. F. H., Rakoczy, H. & Tomasello, M. (2019) Eighteen-month-old infants correct non-conforming actions by others. *Infancy* 24(4):613–35. [ZL]
- Schmidt, M. F. H. & Tomasello, M. (2012) Young children enforce social norms. *Current Directions in Psychological Science* 21(4):232–36. [aMT]
- Schniter, E., Sheremeta, R. M. & Sznycer, D. (2012) Building and rebuilding trust with promises and apologies. *Journal of Economic Behavior & Organization* 94:242–56. [AR]
- Searle, J. R. (1995a) *The construction of social reality*. Free Press. [aMT]
- Searle, J. R. (1995b) *The construction of social reality*. Penguin. [ZL]
- Searle, J. R. (2010) *Making the social world: The structure of human civilization*. Oxford University Press. [ZL, aMT]
- Sellars, W. (1962) Philosophy and the scientific image of man. In: *Frontiers of science and philosophy*, ed. R. Colodny, pp. 35–78. University of Pittsburgh Press. [HR]
- Sen, A. (2009) *The idea of justice*. Harvard University Press. [MK]
- Shaw, A., Barakzai, A. & Keysar, B. (in press) When negative reciprocity is more fair than positive reciprocity. *Cognitive Science*. [AS]
- Shaw, A., DeScioli, P. & Olson, K. R. (2012) Fairness versus favoritism in children. *Evolution and Human Behavior* 33:736–45. [AS]
- Shaw, A. & Olson, K. (2014) Fairness as partiality aversion: The development of procedural justice. *Journal of Experimental Child Psychology* 119:40–53. [aMT]
- Shteynberg, G. (2012) Intersubjectivity, agency, and idiosyncratic identity. *Journal of the Anthropological Society of Oxford* 4(1):1–21. [JBH]
- Shteynberg, G. (2015) Shared attention. *Perspectives on Psychological Science* 10(5):579–90. [JBH]
- Shteynberg, G. (2018) A collective perspective: Shared attention and the mind. *Current Opinion in Psychology* 23(1):93–97. [JBH]

- Shun, K.-L. & Wong, D. B. (2004) Introduction. In: *Confucian ethics: A comparative study of self, autonomy, and community*, ed. K.-L. Shun & D. B. Wong, pp. 1–7. Cambridge University Press. [EEB]
- Shwe, H. I. & Markman, E. M. (1997) Young children's appreciation of the mental impact of their communicative signals. *Developmental Psychology* 33(4):630–36. [PHL]
- Simpson, A. & Laham, S. M. (2015, February) Individual differences in relational construal are associated with variability in moral judgment. *Personality and Individual Differences* 74:49–54. [MSC]
- Simpson, A., Laham, S. M. & Fiske, A. P. (2016) Wrongness in different relationships: Relational context effects on moral judgment. *The Journal of Social Psychology* 156(6):594–609. [MSC]
- Siposova, B., Tomasello, M. & Carpenter, M. (2018) Communicative eye contact signals a commitment to cooperate for young children. *Cognition* 179:192–201. [arMT]
- Smetana, J.G., Rote, W., Jambon, M., Tasopoulos-Chan, M., Villalobos, M. & Comer, J. (2012) Developmental changes and individual differences in young children's moral judgments. *Child Development* 83(2):683–96. doi:10.1111/j.1467-8624.2011.01714.x. [MK]
- Smith, A. (2013) Moral blame and moral protest. In: *Blame: Its nature and norms*, ed. D. J. Coates & N. A. Tognazzini, pp. 27–48. Oxford University Press. [aMT]
- Song, R., Over, H. & Carpenter, M. (2015) Children draw more affiliative pictures following priming with third-party ostracism. *Developmental Psychology* 51(6):831–40. doi:10.1037/a0039176. [MK]
- Sorce, J. F., Emde, R. N., Campos, J. J. & Klinnert, M. D. (1985) Maternal emotional signaling: Its effect on the visual cliff behavior of 1-year-olds. *Developmental Psychology* 21(1):195–200. https://doi.org/10.1037/0012-1649.21.1.195. [SB-D]
- Spaemann, R. (2006) *Persons: The difference between "someone" and "something."* Oxford University Press. (Original work published in 1996). [JIMC]
- Spivak, A. (2016) Dynamics of young children's socially adaptive resolutions of peer conflict. *Social Development* 25(1):212–31. doi:10.1111/sode.12135. [MK]
- Starmans, C. & Bloom, P. (2016) When the spirit is willing, but the flesh is weak: Developmental differences in judgments about inner moral conflict. *Psychological Science* 27 :1498–1506. [AS]
- Steger, M. F., Kashdan, T. B. & Oishi, S. (2008) Being good by doing good: Daily eudaimonic activity and well-being. *Journal of Research in Personality* 42(1):22–42. https://doi.org/10.1016/j.jrp.2007.03.004. [EEB]
- Sterponi, L. (2004) Construction of rules, accountability and moral identity by high-functioning children with autism. *Discourse Studies* 6(2):207–28. [ZL]
- Stewart, A. (2016) The Trolley problem and me: A study of identity and its effect on moral decision-making. Unpublished MSc Dissertation, Department of Psychological and Behavioural Science, London School of Economics. [BF]
- Stewart, A., Franks, B. & I. Gleibs (2019, in preparation) The divided we and moral choices: How shifts within an identity can change decision making. Department of Psychological and Behavioural Science, London School of Economics. [BF]
- Stivers, T., Mondada, L. & Steensig, J., eds. (2011) *The morality of knowledge in conversation*. Cambridge University Press. [ZL]
- Storbeck, J. & Clore, G. L. (2008) Affective arousal as information: How affective arousal influences judgments, learning, and memory. *Social and Personality Psychology Compass* 2(5):1824–43. https://doi.org/10.1111/j.1751-9004.2008.00138.x. [SB-D]
- Strawson, P. F. (1962) Freedom and resentment. *Proceedings of the British Academy* 48:1–25. [aMT]
- Strawson, P. F. (1968) Freedom and resentment. In: *Studies in the philosophy of thought and action*. Oxford University Press. [SD]
- Sylwester, K. & Roberts, G. (2010) Cooperators benefit through reputation-based partner choice in economic games. *Biology Letters* 6(5):659–62. [AR]
- Sylwester, K. & Roberts, G. (2013) Reputation-based partner choice is an effective alternative to indirect reciprocity in solving social dilemmas. *Evolution and Human Behavior* 34(3):201–6. [AR]
- Sznycer, D., Tooby, J., Cosmides, L., Porat, R., Shalvi, S. & Halperin, E. (2016) Shame closely tracks the threat of devaluation by others, even across cultures. *Proceedings of the National Academy of Sciences* 113(10):2625–30. [AR]
- Sznycer, D., Xygalatas, D., Agey, E., Alami, S., An, X. F., Ananyeva, K. I., Atkinson, Q. D., Broitman, B. R., Cotne, T.J., Flores, C., Fuskushima, S., Hitokoto, H., Kharitonov, A. N., Onyishi, C. N., Onyishi, I.E., Romero, P. P., Schrock, J. S., Snodgrass, J., Sugiyaman, L. S., Takemura, K., Townsend, C., Zhuang, J., Aktipis, C. A., Cronk, L., Cosmides, L. & Tooby, J. (2018) Cross-cultural invariances in the architecture of shame. *Proceedings of the National Academy of Sciences* 115(39):9702–707. [AR]
- Tajfel, H. & Turner, J. C. (1986) *The social identity theory of inter-group behavior*. In: *Psychology of intergroup relations*, ed. S. Worchel & W. Austin, pp. 7–24. Nelson-Hall. [MK]
- Tamir, D. I. & Mitchell, J. P. (2012) Disclosing information about the self is intrinsically rewarding. *Proceedings of the National Academy of Sciences* 109:8038–43. [HMM]
- Telzer, E. H., Fuligni, A. J., Lieberman, M. D. & Galván, A. (2014) Neural sensitivity to eudaimonic and hedonic rewards differentially predict adolescent depressive symptoms over time. *Proceedings of the National Academy of Sciences* 111(18):6600–05. https://doi.org/10.1073/pnas.1323014111. [EEB]
- Tepe, B. & Adymi-Karakulak, A. (2018) Beyond harmfulness and impurity: Moral wrongness as a violation of relational motivations. *Journal of Personality and Social Psychology* 117(2):310–37. [MSC]
- Terry, D. J., Hogg, M. A. & White, K. M. (1999) The theory of planned behaviour: self-identity, social identity and group norms. *British Journal of Social Psychology* 38(3):225–44. [JBH]
- Thompson, M. (2004) What is it to wrong someone?: A puzzle about justice. In: *Reason and value: Themes from the philosophy of Joseph Raz*, ed. J. Wallace, P. Pettit, S. Scheffler & M. Smith. Oxford University Press. [aMT]
- Thompson, R. A. (2012) Wither the pre-conventional child? Toward a life-span moral development theory. *Child Development Perspectives* 6(4):423–29. doi:10.1111/j.1750-8606.2012.00245.x. [RAT]
- Thompson, R. A. (2019) Early moral development and attachment theory. In: *The Oxford handbook of parenting and moral development*, ed. D. Laible, L. M. Padilla-Walker & G. Carlo, pp. 21–39. Oxford University Press. [RAT]
- Thomson, J. J. (1985) The trolley problem. *Yale Law Journal* 94:1395–1415. [BF]
- Ting, F., Dawkins, M. B., Stavans, M. & Baillargeon, R. (2019) Principles and concepts in early moral cognition. In: *The social brain: A developmental perspective*, ed. J. Decety. MIT Press. [MR]
- Ting, F., He, Z. & Baillargeon, R. (2019) Toddlers and infants expect individuals to refrain from helping an ingroup victim's aggressor. *Proceedings of the National Academy of Sciences* 116(13):6025–34. https://doi.org/10.1073/pnas.1817849116. [LC]
- Tomasello, M. (1988) The role of joint attentional processes in early language development. *Language Sciences* 10(1):69–88. doi:10.1016/0388-0001(88)90006-X. [ZL]
- Tomasello, M. (2014a) *A natural history of human thinking*. Harvard University Press. [MG, aMT]
- Tomasello, M. (2014b) The ultra-social animal. *European Journal of Social Psychology* 44(3):187–94. [ZL]
- Tomasello, M. (2016a) *A natural history of human morality*. Harvard University Press. [SB-D, MG, PP, arMT]
- Tomasello, M. (2016b) Cultural learning redux. *Child Development* 87(3):643–53. https://doi.org/10.1111/cdev.12499. [EEB]
- Tomasello, M. (2018) The normative turn in early moral development. *Human Development* (Special Issue) 61:248–63. [aMT]
- Tomasello, M. (2019) *Becoming human: A theory of ontogeny*. Harvard University Press. doi:10.4159/9780674988651. [JC, aMT]
- 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(5):675–91. doi:10.1017/S0140525X05000129. [JC, aMT]
- Tomasello, M., Kruger, A. C. & Ratner, H. H. (1993) Cultural learning. *Behavioral and Brain Sciences* 16(3):495–511. https://doi.org/10.1017/S0140525X0003123X. [EEB]
- Tomasello, M., Melis, A. P., Tennie, C., Wyman, E. & Herrmann, E. (2012) Two key steps in the evolution of human cooperation: The interdependence hypothesis. *Current Anthropology* 53(6):673–92. [rMT]
- Tomasello, M. & Vaish, A. (2013) Origins of human cooperation and morality. *Annual Review of Psychology* 64(1):231–55. [ZL]
- Tripathi, R., Cervone, D. & Savani, K. (2018) Are the motivational effects of autonomy-supportive conditions universal? Contrasting results among Indians and Americans. *Personality and Social Psychology Bulletin* 44(9):1287–1301. https://doi.org/10.1177/0146167218764663. [EEB]
- Turiel, E. (1983) *The development of social knowledge: Morality and convention*. Cambridge University Press. [HR]
- Turiel, E. (2002) *The culture of morality: Social development, context, and conflict*. Cambridge University Press. [MK]
- Turiel, E. (2007) The development of morality. In: *Handbook of child psychology, vol. 3: Social, emotional, and personality development*, ed. W. Damon, R. M. Lerner, & N. Eisenberg, pp. 789–857. Wiley. doi:10.1002/9780470147658.chpsy0313 [aMT]
- Turiel, E. & Dahl, A. (2019) The development of domains of moral and conventional norms, coordination in decision-making, and the implication of social opposition. In: *The normative animal: On the anthropological significance of social, moral, and linguistic norms*, ed. K. Bayertz & N. Roughley, pp. 195–213. Oxford University Press. [MK]
- Turnbull, W. (2003) *Language in action: Psychological models of conversation*. Psychology Press. [JIMC]
- Turner, J. C., Hogg, M. A., Oakes, P. J., Reicher, S. D. & Wetherell, M. S. (1987) *Rediscovering the social group: A self-categorization theory*. Basil Blackwell. [AS]
- 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. [aMT]
- Ullman-Margalit, E. (1977) *The emergence of norms*. Oxford University Press. [WS-A]
- Vaish, A., Carpenter, M. & Tomasello, M. (2016) The early emergence of guilt-motivated prosocial behavior. *Child Development* 87:1772–82. [aMT]
- Van Lawick-Goodall, J. (1968) The behavior of free-living chimpanzees in the Gombe Stream Reserve. *Animal Behavior Monographs* 1(Part 3):161–311, IN1–IN12. [RBR]

- Van Tongeren, D. R., Newbound, H. & Johnson, E. (2016) The interactive effects of religiosity and priming religion following recall of a values violation. *Sexual Addiction & Compulsivity* **23**(2–3):211–24. [BF]
- Van Tongeren, D. R., Welch, R. D., Davis, D. E., Green, J. D. & Worthington Jr., E. L. (2012) Priming virtue: Forgiveness and justice elicit divergent moral judgments among religious individuals. *The Journal of Positive Psychology* **7**:405–15. [BF]
- Wainryb, C., Brehl, B. & Matwin, S. (2005) Being hurt and hurting others: Children's narrative accounts and moral judgments of their own interpersonal conflicts. *Monographs of the Society for Research in Child Development* **70**(3):1–125. Wiley-Blackwell. [MK]
- Walker, R., Hill, K., Kaplan, H. & McMillan, G. (2002) Age dependency of hunting ability among the Ache of eastern Paraguay. *Journal of Human Evolution* **42**(6):639–57. [AK]
- Wallace, J. (2019) *The moral nexus*. Princeton University Press. [aMT]
- Wallace, R. J. (1994) *Responsibility and the moral sentiments*. Harvard University Press. [aMT]
- Walster, E., Walster, G. W. & Berscheid, E. (1978) *Equity: Theory and research*. Allyn & Bacon. [MSC]
- Warneken, F., Chen, F. & Tomasello, M. (2006) Cooperative Activities in Young Children and Chimpanzees. *Child Development* **77**(3):640–63. [aMT]
- 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. [aMT]
- Warneken, F., Lohse, K., Melis, A. P. & Tomasello, M. (2011) Young children share the spoils after collaboration. *Psychological Science* **22**(2):267–73. doi:10.1177/0956797610395392. [JC, aMT]
- Warneken, F. & Tomasello, M. (2013) The emergence of contingent reciprocity in young children. *Journal of Experimental Child Psychology* **116**:338–50. [aMT]
- Waytz, A., Dungan, J. & Young, L. (2013) The whistleblower's dilemma and the fairness loyalty tradeoff. *Journal of Experimental Social Psychology* **49**:1027–33. [AS]
- Wellman, H. & Gelman, S. A. (1992) Cognitive development: Foundational theories of core domains. *Annual Review of Psychology* **43**(1):337–75. <https://doi.org/10.1146/annurev.psych.43.1.337>. [LC]
- Wellman, H. M., Song, J.-H. & Peskin-Shepherd, H. (2019) Children's early awareness of comprehension as evident in their spontaneous corrections of speech errors. *Child Development* **90**(1):196–209. [PHL]
- Wilks, M., Kirby, J. & Nielsen, M. (2018) Developmental changes in young children's willingness to copy the antisocial actions of ingroup members in a minimal group context. *Developmental Psychology* **55**(4):709–21. [PHL]
- Williams, B. (1981) Practical necessity. In: *Moral luck: Philosophical papers 1973–1980*, pp. 124–32. Cambridge University Press. [DO]
- Williams, B. (1993) Moral incapacity. *Proceedings of the Aristotelian Society* **3**:59–70. [DO]
- 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):324–37. [aMT]
- Wright, J. C. & Bartsch, K. (2008) Portraits of early moral sensibility in two children's everyday conversations. *Merrill-Palmer Quarterly* **54**(1):56–85. [KB]
- Wu, J., Balliet, D. & Van Lange, P. A. (2016) Gossip versus punishment: The efficiency of reputation to promote and maintain cooperation. *Scientific Reports* **6**:23919. [AR]
- Yucel, M., Hepach, R. & Vaish, A. (in press) Young children and adults show differential arousal to moral and conventional transgressions. *Frontiers in Psychology*. [SB-D]