

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.

Q80 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 2019). 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 this commentary, 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 them 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 and 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, e0

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., [under review](#); 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>