GOAL SLIPPAGE: A MECHANISM FOR SONTANEOUS INSTRUMENTAL HELPING IN INFANCY?

John Michael
Department of Philosophy
University of Warwick
Coventry, UK
j.michael.2@warwick.ac.uk

&

Marcell Székely
Department of Cognitive Science
Central European University
Budapest, Hungary
szekelymarcell@gmail.com

Abstract: In recent years, developmental psychologists have increasingly been interested in various forms of prosocial behavior observed in infants and young children – in particular comforting, sharing, pointing to provide information, and spontaneous instrumental helping. We briefly review several models that have been proposed to explain the psychological mechanisms underpinning these behaviors. Focusing on spontaneous instrumental helping, we home in on models based upon what Paulus (2014) has dubbed 'goal-alignment', i.e. the idea that the identification of an agent's goal leads infants to take up that goal as their own. We identify a problem with the most well-known model based upon this idea, namely the 'goal contagion' model. The problem arises from the way in which the model specifies the content of the goal which is identified and taken up. We then propose an alternative way of specifying the content of the goal, and use this as a starting point for articulating an alternative model based upon the idea of alignment, namely the 'goal slippage' model. By elucidating the difference between goal contagion and goal slippage, we contribute to the articulation of experimental criteria for assessing whether and when the mechanisms specified by these two models are at work.

Keywords: prosocial behavior, helping, altruism, goals, joint action, common coding

1. Introduction

As a species, we humans are characterized by the pervasiveness and flexibility with which we cooperate. In attempting to account for this hallmark of human sociality, comparative and developmental psychologists have increasingly become interested in the emergence in infancy and early childhood of prosocial behavior, i.e. of 'behaviors benefiting another person without providing the helper an immediate payoff' (Paulus 2014). In particular, it has been observed that infants as early as the second year of life comfort others who are in distress (Bischof-Köhler, 1988, 1991; Johnson, 1982; Zahn-Waxler, Radke-Yarrow, Wagner, & Chapman, 1992; Dunfield et al., 2011; 2013; Vaish, Carpenter, & Tomasello, 2009), share food and other resources (Hay et al. 1991; & Levitt 1995), point to provide others with information (Liszkowski et al. 2006), and spontaneously help others to achieve their goals (Warneken & Tomasello, 2006; Svetlova, Nichols, & Brownell, 2010; Hepach et al., 2012; 2016; 2017). Given that these behaviors are exhibited prior to extensive enculturation, they may reflect the

operation of phylogenetically ancient psychological mechanisms underpinning human cooperation (Warneken & Tomasello, 2009). Explaining the mechanisms underpinning these behaviors may therefore shed light on how human cooperation emerged in evolution and what basic psychological mechanisms sustain it today.

While it may be tempting to seek a single explanation that covers all of the aforementioned varieties of prosociality in infancy, we should be wary of assuming that this will be possible, especially in light of recent research indicating that these types of behaviors exhibit 'dissociable developmental trajectories and distinct associations with individual difference factors early in life' (Dunfield, 2014: 1). In the following, therefore, we will limit our focus to one variety of infant prosociality, namely spontaneous instrumental helping (Warneken & Tomasello 2006; Warneken & Tomasello, 2007; Warneken et al., 2007; Cf. Svetlova et al., 2010; Dunfield, 2014).

While many different scenarios have been used in instrumental helping paradigms, they share the same basic structure: an agent is unable to achieve her goal because there is some obstacle or because an object is out of her reach. In one scenario, for example, an agent who wants to write a letter attempts to grasp a pencil which is out of her/his reach, but which is within reach of the infant participant. As Warneken & Tomasello (2007) have demonstrated, even 14-month-olds will typically grasp the pencil and hand it over to the agent. In a different scenario, 18-month-olds observing an agent who is unable to place a stack of books into a cabinet when the cabinet door falls shut (the agent's arms are full of books) will typically jump up, walk over, and open the cabinet door to help the agent (Warneken & Tomasello, 2006). In a third scenario, a book slides off of a stack as the agent attempts to place it on top of the stack; the 18-month-olds pick up the book and return it to the top of the stack (Warneken & Tomasello, 2006).

What leads children to perform these actions? We will begin (Section 2) with a brief overview of models that have been proposed to account for this, and refer to specific studies which either support or fail to support each model. Next (Section 3), we turn to our primary project, which is to illuminate a specific subset of models that have been proposed to account for spontaneous instrumental helping – namely models based upon 'goal alignment' (Paulus, 2014). The core idea behind goal alignment models that the identification of an agent's goal leads infants to take up that goal as their own (Barresi & Moore 1996; Kenward & Gredebäck 2013; Paulus, 2014; Köster et al., 2015; Michael, Sebanz & Knoblich, 2016). Our main aim will be to distinguish between two separate goal-alignment models: 'goal contagion' and 'goal slippage'. As we shall see, these two models differ in how they specify the content of

the goal which is identified and taken up. By elucidating this difference, we contribute to the articulation of experimental criteria for assessing whether and when the mechanisms specified by these two goal-alignment models are at work.

2. Modeling Instrumental Helping

2.1 Psychological Altruism

The first model to consider is the one originally proposed by Warneken and Tomasello (2006), namely that the kids in their studies are motivated by altruism. More specifically, what they have in mind is *psychological* altruism. Since they themselves do not use the qualifier 'psychological,' it is worth taking a moment to explain why we do so. Doing so will also help to clarify the explanatory target of the models we will be considering.

In using the term 'psychological altruism,' we are appealing to the general distinction between the evolutionary (ultimate) level of explanation and the psychological (proximate) level (Tinbergen 1963). In evolutionary biology, a behavior is considered altruistic if it raises the expected reproductive success of the recipient at the expense of the reproductive success of the agent performing it (Kitcher, 1998). Altruism is puzzling from an evolutionary perspective insofar as a disposition to act in a way that does not enhance the chances of one's own genes to be propagated should be expected to disappear from a population over time through natural selection (Axelrod & Hamilton 1981; Axelrod, 1984; Hamilton 1964, 1970; Maynard Smith, 1964, 1974). Much research in recent decades has accordingly been devoted to identifying evolutionary explanations (i.e. ultimate mechanisms) that would support the selection of altruistic behavior – e.g. kin selection, direct (Trivers, 1971) and indirect (Nowak and Sigmund, 1998) reciprocity, and the interdependency hypothesis (Roberts 2005; Tomasello 2016). Regardless of which of these ultimate mechanisms turns out to be correct, though, there is still a further question as to what the psychological mechanisms are that actually motivate the altruistic behavior. After all, humans (and other animals) often engage in (at least apparently) altruistic behavior without explicitly reasoning about kin relations or any other evolutionary rationale. What are the psychological mechanisms that motivate them to do so?

One possibility is *psychological* altruism, i.e. for altruism to feature as a proximate mechanism. What this means is that the agent's goal in performing the behavior is to provide a benefit to the recipient (Batson et al., 2008; Foster, Wenseleers, & Ratnieks, 2006), and that

the recipient's benefit must be perceived as an end in itself, not as a means to the achievement of some other goals (Kitcher, 1998) or to the attainment of an external reward (Piliavin & Charng, 1990). According to a narrower definition, the benefit to the recipient must come at a cost to the agent (Grusec, Davidov & Lundell 2002). This narrower definition is more rigorous in that it can serve to rule out the possibility that the altruistic behavior is performed at least in part because its performance is intrinsically rewarding (i.e. this potential reward is offset by a cost).

Is this the case for infants in instrumental helping paradigms such as the ones referred to above? In support of this conjecture, Warneken et al. (2007) were able to show that rewarding the infants for helping did not increase their helping behavior at all (experiment 1), and also (experiment 2) that 18-month olds were no less likely to help if it was made more costly for them (they had to get by an obstacle in order to do so, which is difficult for an 18-month-old). Building on this, Warneken & Tomasello (2008) reported the same pattern of findings when they raised the cost of the helping behavior, i.e. the helping required the infants to resist the attractive option to play with interesting toys in a different part of the space. Carrying this logic further, Svetlova et al. (2010) increased the cost still further by devising a scenario in which the child would have to (temporarily) give up a cherished object brought from home (such as a favorite hairclip) in order to help. They found that 30-month-olds were still willing to help, albeit to a lesser extent than when the help was not costly; 18-month-olds, in contrast, rarely helped in this condition. One other crucial finding from Warneken & Tomasello's (2008) study was that children who had received a material reward for helping at one time-point were less likely to help at a later time point than children who had not been rewarded. As Warneken & Tomasello (2009) note: 'this surprising finding provides even further evidence for the hypothesis that children's helping is driven by an intrinsic rather than an extrinsic motivation. Rewards are often not only superfluous, but can have even detrimental effects as they can undermine children's intrinsic altruistic motivation' (460).

Impressive as these findings are, they are not decisive. With respect to the question of *external* rewards, Dahl (2015) was able to show that, while infants may not be rewarded in the lab for performing these behaviors, 11-25 month-olds are commonly encouraged to help with similar activities at home and praised for doing so. Thus, it is possible that rewards form part of the developmental context in which instrumental helping emerges. There is also the issue of *internal* rewards – i.e. infants may experience a positive emotion as a result of helping and be motivated by this. In line with this prediction, Aknin et al. (2012) reported that 22 month-olds exhibit greater happiness when giving rewards to others than when receiving the rewards

themselves. Moreover, they were especially happy when engaging in costly giving – i.e. forfeiting their own resources to give to others. The notion of internal rewards underscores a difficulty in evaluating the psychological altruism model: it is not clear how to distinguish between internal rewards that may be included as components within the altruism model and internal rewards which present alternatives to it. On the face of it, the suggestion that infants are motivated by the prospect of a positive emotion ('warm glow') appears to be an alternative to the suggestion that their motivation stems from a desire that the agent be helped. On the other hand, the psychological altruism model must surely include some specific account of how the altruistic desire to bring about the observed agent's goal motivates the helping behavior. Could such an account appeal to the prospect of a warm glow as a motive? In order to decide this, and more generally to specify testable predictions derivable uniquely from psychological altruism model, it will be necessary to spell the model out in greater detail, and in particular to specify the motivational mechanisms that it includes.

2.2 A Preference for Joint Action

A further motivation for some forms of altruistic behavior is that it can be intrinsically pleasurable to engage in joint actions. Thus, young children who exhibit spontaneous instrumental helping behavior may do so at least in part because they like engaging in joint actions and are motivated to do so (Rheingold et al., 1982; Svetlova et al., 2010; Paulus & Moore, 2012[1]), i.e. not because of any benefit that their contribution brings to anyone else.

One finding in the literature that provides support for this model is from a study by Barragan & Dweck (2014). This study was motivated by the thought that the 'reciprocal play' phase used in many studies to familiarize infants with the experimenter may prime a cooperative (joint action?) mindset. To test this, they contrasted a condition in which the experimenter engages in reciprocal play with the infant (rolling a ball back and forth) with a condition in which the experimenter and infant play in parallel next to each other. The main finding was that infants were significantly more likely to help the experimenter after reciprocal play than after parallel play.

This model generates various testable predictions, some of which have indeed already been tested. First, it generates the prediction that infants will be less likely to help if doing so would not involve a joint action with the other agent. This prediction appears not to be

[1] Moore & Paulus (2012) use the term 'social interaction model' to refer to much the same idea.

supported by the results of a study by Hepach et al (2016), in which it was shown that 18-month-olds were no less likely to help in a condition in which the other agent was absent during the performance of the helping behavior. There are at least two ways in which this finding could be accommodated within the model however. First, the infants may experience the activity as a joint action even if the agent is temporarily not present – especially if a joint action mindset has been primed in a prior familiarization phase (cf. Barragan & Dweck, 2014. Second, the finding does not rule out the possibility that a preference for joint action might provide a further motivation.

Another important study by Hepach and colleagues (2012) also bears upon this model. In this study, 2 year-olds' pupil dilation was measured (as a proxy for arousal) at key moments during the experiment, for example when the agent was in distress as she dropped a crayon and was unable to reach it. The results showed that the children were aroused upon seeing the agent in distress, and then just as relieved to see that agent helped by some third party as they were when they helped the agent themselves. This suggests that, at least in these cases, they are motivated more strongly by the desire that the agent be helped than by the desire to perform a joint action together with the agent. However, it is possible that this is because in the case of distress, the goal of relieving the agent's distress is more salient than any other goal. Thus, it may in principle be helpful to apply a similar method to cases in which the agent is not in distress but is performing a mundane, everyday activity, such as putting books into a cabinet or reaching for a pencil – although measuring pupil dilation may not be appropriate methodologically for this question, since an agent's struggle to achieve her goal in a mundane situation may not be sufficiently arousing to elicit a change in pupil dilation. And, again, as we noted above in discussing Hepach and colleagues' (2016) study, the infants may experience the activity as a joint action (including the third party) even when they themselves do not have to make a contribution at the moment because the third party does so.

The joint action preference model also generates a slightly different prediction, namely that infants will help irrespective of the benefit to the helpee. For example, if infants were just as likely to help an agent to achieve a goal which was detrimental to her well-being, this would be difficult to explain as altruism, but would be unsurprising from the perspective of the joint action preference model. We know of no research testing this directly.

2.3 Aversion to Others' Distress

A further motivation for prosocial behavior is that seeing others nervous or upset (e.g. about not achieving a goal) can be aversive. It may, for example, make infants and young children nervous or upset, possibly because they fear negative consequences for themselves. To the extent that this is the case, it may provide a motivation to contribute to others' goals in order to avoid being confronted with others' distress.

The aversion to others' distress model has prima facie plausibility in light of Hoffman's (1975) influential stage theory, which posits that infants are capable of empathic distress in the first year of life but do not experience empathic concern until the second year. On the other hand, some studies designed to test Hoffman's theory have not corroborated the predictions that it generates. For example, Hay et al. (1981) reported that 6-month-olds tended to orient toward a peer who was expressing distress; i.e. they did not seem to be confused as to who was in distress or to focus on their own state of distress. Similarly, Roth-Hanania et al. (2011) observed signs of affective and cognitive empathy by 8-10-months (for a review and discussion, see Davidov et al. 2013).

In view of these findings, the aversion to others' distress model seems unlikely to fully explain the instrumental helping data. Of course, this does not rule out the possibility that it may identify one factor among others that can motivate infants' instrumental helping behavior. To explore this, one possibility would be to investigate whether young children would be contented to simply occlude their view of the agent in distress, or to exit from the scene.[2]

2.4 Reputation Management

As adults, we sometimes calculate the likely consequences of our actions on our reputations. If infants are to some extent motivated by similar concerns, this could provide an explanation of their instrumental helping behavior. As it happens, current research suggests that it is not until somewhat later in childhood (i.e. around 5) that children adapt their actions to manage their reputations. Specifically, Engelmann et al (2012; 2013) have shown that 5-year-olds share more and steal less when observed by a peer than when alone.

To test the model directly, it would be important to investigate how kids behave in situations in which they are not being observed. The reputation management model should

[2] This suggestion is based upon a paradigm used with adults by Batson et al. (1987).

predict that they would be less likely to help when they do not believe that they are being observed. As it happens, there are some findings that bear upon this prediction. For example, the study by Hepach and colleagues mentioned above (2016) provides evidence that 18 month-olds are equally likely to help in a scenario in which the agent is absent and does not know about the help being offered.

While these findings indicate that reputation management is unlikely to provide an exhaustive explanation of the motivation underlying instrumental helping in infancy, it is of course still possible that infants may be responsive to cues which are relevant to reputation management, even if they are not representing or reasoning about reputation per se. In other words, some more proximal mechanisms underpinning reputation management may play a role. This conjecture is motivated by the results of Rochat, Broesch & Jayne (2012), who administered a mirror self-recognition task designed to test the hypothesis that young children interpret their mirror image in reference to how others might perceive and evaluate them. To this end, they included a 'Norm Condition' in which the child, the experimenter and a parent were all marked prior to the mirror exposure. They found that children as young as 18 months old tended either to leave the mark on in the Norm Condition or at least to hesitate, indicating that they were aware of their self-image and acted to conform to the norms of their group.

With respect to instrumental helping, one possibility is that infants may be responding to the expectations which they take others to have of them: The infants may infer that they are expected to help and have a default preference to fulfill expectations that they take others to have of them. In many of the scenarios used in instrumental helping studies, the agent performs an action that is not only highly unlikely to lead to their apparent goals but also highly inefficient. For example, the experimenter in Warneken and Tomasello's seminal (2006) study walks towards a closed cabinet with an armload of books. It would be rational for the infants scenarios like this to infer that the experimenter is expecting them to help. This interpretation would be supported if it could be shown that making the other agent's expectation more salient increased the helping behavior (e.g., if the agent announced to some third party that she expected the participant to help, or if she made eye contact with the participant).

There are some findings in the literature which sit awkwardly with the conjecture that infants' instrumental helping behavior may constitute a response to expectations. First, Warneken (2013) found that 2 year-olds helped just as much in the absence of parental presence an encouragement – although the significance of this result should be qualified in light of Dahl's (2015) finding, discussed above, that encouragement and praise in helping

situations at home may instill a belief in the children that they are expected to help. The next awkward finding is due to Warneken, who in a recent study (Warneken 2013) reported that 2 year-olds were just as likely to help when the helpee was not yet aware of the accident necessitating the help, and thus did not expect any help. This, like the Hepach et al. (2016) report of anonymous helping discussed above, are not immediately reconcilable with the present conjecture about expectation. In both of these cases, though, the infant may anticipate that s/he will be expected to help in a moment when the agent notices the accident/returns to the scene, and be proactively sensitive to this expectation. Moreover, the point about Dahl's (2015) findings also applies here: the infants may already have learned that they are expected to help in such situations.

2.5 Compulsive Planning

A further model, which to our knowledge has not yet been discussed in the literature, takes its starting point from the observation that humans are also highly proficient at representing the instrumental structure of action – i.e. at constructing plans and flexibly adapting them during the course of actions (including *joint* actions) (Silk, 2009; Tomasello, 2009). Infants and young children, though, do not yet exhibit this characteristic human. This point is illustrated by a recent study by Beck and colleagues (2011). They found that 5-year-old children were not proficient at innovating tools in scenarios that were in fact quite similar to contexts in which some non-human animals, particularly corvids, perform well (Weir, Chappell, & Kacenik, 2002; Bird & Emery, 2009). We believe that these findings underscore the point that children must acquire a proficiency for reasoning flexibly about the instrumental structure of actions – given that such a proficiency is clearly highly characteristic of adult humans. So, when an infant or young child observes an agent performing an action and identifies the goal of the action (e.g. putting the books in the shelf), she may have a tendency to spontaneously engage in practice planning, i.e. to calculate the most efficient way of achieving the goal. In many cases the most efficient plan involves a contribution X from a second agent (for example opening the cabinet door). In sum, the identification of the goal leads the child to represent X (what she would need to do to contribute to bringing about the goal). Of course, this conjecture does not yet provide an explanation of why the infant would then act to carry out such a plan once she has constructed it, but there are ways of addressing this challenge.

One possibility, for example, is to speculate that if the infant works out an efficient plan for bringing about the goal, it could be unsatisfying for her to see the goal pursued in a

less efficient manner. In other words, children (and perhaps people in general) have a preference for things to be done in the most efficient way possible. As a source of preliminary support to motivate this conjecture, one might reflect that it can indeed be irritating to see people performing actions incompetently -- often one feels an urge to correct them and make them do it right. To probe the conjecture experimentally, it could be fruitful to investigate whether children would be content to contribute to a suboptimal strategy. Studies with children as young as 2 reveal a tendency to correct others who do not perform actions in a manner that conflicts with relevant norms, i.e. not in accordance with the rules of a game (Rakoczy, 2008; Rakoczy & Schmidt, 2013). Of course, failing to perform actions according to rules is different from failing to perform actions in an efficient manner, but these results at least do indicate that children are in some cases prepared to correct other agents who do not perform an action in the way the child believes they should. Perhaps the methods employed in these studies may therefore be extended to investigate whether children at this age would also be motivated to correct others who perform actions inefficiently and/or whether children may be less willing to contribute to inefficient action plans.

A further option for addressing the challenge would be to speculate that performing the actions in these scenarios may provide the infants with an opportunity for active learning, i.e. to test out the plan and learn from the consequences[3] (cf. Schulz, 2015; Buchsbaum et al., 2012). This could be in principle tested by manipulating the degree to which the goal and actions are familiar to the infant. More familiar goals and actions (perhaps due to a longer familiarization phase) should, on this model, elicit less helping behavior.

2.6 Goal Alignment

A further class of models, which Paulus (2014) has dubbed 'goal-alignment models', are based on the core idea that the identification of an agent's goal leads infants to take up that goal as their own. This may occur because of the lack of self–other differentiation in young infants (cf. Barresi & Moore, 1996) – i.e. having identified the goal, the infant lacks the resources to quarantine it from her own endogenous goals and simply treats it like any other goal that she has. As a result, she is motivated to perform the action just as she would be if the goal had arisen endogenously. In this subsection, we briefly present two different versions of this basic idea.

[3] We are grateful to Sam Clarke for this suggestion.

Goal Contagion: One way of thinking about goal contagion is in terms of behavioral mimicry at a relatively abstract level, i.e. not imitating the agent's specific movements but being primed to perform an action with the same goal: 'The representation of the observed goal may have primed behaviour resulting in that goal' (Kenward & Gredebäck, 2013). To motivate this conjecture, Kenward and Gredebäck refer to research showing that adults are motivated to perform actions with similar goals to the agents in vignettes they have read. In particular, Aarts et al. (2004) exposed participants to brief vignettes in which an agent appeared to be motivated to achieve a certain goal, such as making money, and then measured participants' motivation to pursue a similar goal, such as making money. They found that participants were indeed primed to exert more effort in pursuit of similar goals (making money in the experiment).

Goal Slippage: Michael, Sebanz & Knoblich (2016) have recently proposed an alternative explanation of young children's spontaneous helping behavior: when the child identifies the goal G of the agent's action (e.g. putting a stack of books into the cabinet), this causes the child to form the goal G, i.e. it elicits a motivation to complete the action and to achieve the goal. 'Goal slippage', as they term this process, may occur as a consequence of the way in which goals are represented at the most basic level, namely in an agent-neutral manner – i.e. as outcomes that are to be brought about, irrespective of *who* desires them or is *who* is attempting to bring them about. In other words, the identification of a goal as a goal has the effect that the goal slips from perception into action, and the child treats it as her or his own goal.

These two models are clearly quite similar insofar as they both seek to explain infants' motivation to help in instrumental helping paradigms as being of the same kind as their motivation to act upon endogenously generated goals. This is because both goal alignment models entail that the infants simply take up the goal as their own[4]. But, as we shall see in the

[4] It is worth noting that there is a way a third model goal alignment model, based upon affective contagion. Kärtner et al. (2010) hypothesize that in situations where an agent is distressed (e.g. when a doll's arm breaks), the infant will be infected with the agent's distress via affective contagion, and that the agent's object-directed behavior (e.g. toward the doll's broken arm) will indicate a cause of the distress. As they put it: 'the toddlers acquired a situation-specific understanding ("sad because of the broken teddy") although "sad" was not understood as the mental state of the other person. This experience-bound understanding allows toddlers to help the distressed other. Thus, we propose that *situational* helping behavior is an alternative to empathically motivated helping behavior in emotion-laden situations with a needy or distressed other' (912). Kärnter et al. (2010)'s study does not relate directly to instrumental helping, since the agent was not distressed because of her failure to

next section, there is also an important difference in how these two goal alignment models specify the content of the goal that the infant identifies and takes up.

3. Whose Goal?

3.1 Rich Goals and Lean Goals

While the concept of a goal is fundamental in psychology and elsewhere, there is surprisingly little consensus about how to define it. At a bare minimum, a goal is an outcome of an agent's movements. But clearly this is not enough to distinguish goals from incidental consequences of movements. For example, stepping on and killing a bug may be a consequence of walking across the room, whereas the goal may be to place some books in the cabinet. Intuitively, an outcome of an action is only a goal of that action if the action is performed because it is likely to bring about that outcome. There are various ways of articulating this idea. In particular, they differ with respect to whether or not they appeal to the mental representations of the agent carrying out the action. Butterfill & Apperly (2013), for example, offer a lean characterization of goals which avoids making appeal to the mental representations of the agent. They write:

> We stipulate that for an outcome, *g*, to be the goal of some bodily movements is for these bodily movements to occur in order to bring about *g*; that is, *g* is the function of this collection. Here "function" should be understood teleologically. On the simplest teleological construal of function, for an action to have the function of bringing about *g* would be for actions of this type to have brought about *g* in the past and for this action to occur in part because of this fact . . . The virtue of this way of representing goals is that it allows them to be inferred from actions without appealing to intentions, beliefs, preferences or other psychological states. (Butterfill & Apperly, 2013, p. 613)

This characterization (by design) eschews mentalistic talk of what an agent *intends* or *desires* to bring about, or is *trying* to bring about, or of what outcome the agent *represents*. Instead, it distinguishes the goal of an action from other outcomes of the action by appealing

---

achieve a goal, and the infants' accordingly tended to comfort the agent rather than instrumentally helping her. For this reason, we will not consider it further here. It would be worthwhile for further research to probe how the affective contagion model could be applied to instrumental helping, and how it would relate to goal contagion and goal slippage.

to the notion of a function (understood teleologically, cf. Millikan, 1984): the action is performed because on previous occasions performing the action led to the outcome. This characterization has the virtue of simplicity, and the absence of mentalistic language may well make it easier to operationalize.

On the other hand, it may be problematic in cases in which an action is performed for the first time, or where it is likely to lead to a different outcome than it has in the past. Moreover, the very same movements can function to bring about different outcomes in different situations, depending on features of the context, including various mental states of the observed agent (Jacob & Jeannerod 2005; Michael & Christensen, 2016). In order to address such cases, it may be useful to appeal to *intentions*, *desires*, *trying* or other mental *representations* that guide the action (Huang & Bargh, 2014; Aarts & Dijksterhuis, 2000). On such a richer view, an outcome of an action counts as a goal if the agent's actions are guided by a representation of that outcome. A representation of a particular outcome may, for example, make it possible to modify the action in light of feedback or of changing circumstances such as to increase the likelihood of efficiently bringing about the outcome.

With this distinction in hand, let us now return to our two goal alignment models. As well shall see, it is natural to think of goal slippage in terms of the lean notion of goals, to think of goal contagion in terms of the rich notion of goals. We will then use this distinction as a wedge to tease apart these models.

3.2 Rich Goals and Goal Contagion

To see how Kenward & Gredeback (2013) think about the notion of goals and in particular the relation between the agent and the goal within the content of the goal representation that infants identify in spontaneous helping paradigms, let us take a closer look at how they interpret the results of their instrumental helping study. They observed that infants lifted agent-like geometrical forms over a barrier and thereby helped them to reach their apparent goal of getting to the other side (Kenward & Gredebäck, 2013). Tellingly, they interpret this as evidence *against* the goal priming model:

> ‘One result, however, speaks against the goal-priming account. If goal-priming led to imitation of a non-human agent’s actions by infants, re-enactment of the agent’s original actions would be expected, at least in the control condition where there was no obvious incomplete action. Such re-enactment was observed only at very low frequencies,

> suggesting that goal- priming may not have been a strong motivator of the infants' actions.' (e75130)

What they seem to have in mind is that priming should lead an infant to take up the goal of getting to the other side of the barrier (i.e. for themselves) -- not the goal that the geometrical forms get to the other side. And indeed this is consistent with the aforementioned study by Aarts et al., (2004), in which participants were primed to take up the same goals (making money, attaining causal sex) as the protagonists of the vignettes with which they were presented: the participants were of course not motivated to make money for arrange causal sex for the protagonist in the vignette but for *themselves*.

It is apparent that the model is based upon the rich characterization of goals as representations of outcomes which guide an agent's actions, is to think of it as a form of priming. When an observer (an infant for example) perceives an agent (an experimenter for example) performing an action directed toward a particular outcome, she identifies the goal in the sense of a mental representation guiding the action. This mental representation then exercises the same functional role that it plays in the observed agent, and functions to guide *the observer's* actions to bringing about the outcome. In other words, identifying the experimenter's goal as a mental representation of a state of affairs that is to be brought about (e.g. putting the books into the cabinet) has the effect that the child observer will come to have this very same goal in the sense of a mental representation guiding her action, and will accordingly experience a tension until the state of affairs is achieved and will organize her actions in such a way as to bring about the state of affairs.

This is consistent with the 'selfish-goals' proposal offered by Huang & Bargh (2014), who argue that goals are autonomous in the sense that they organize an agent's thought and behavior such as to ensure that the represented outcome is brought about, and do so irrespective of whether that outcome is consistent with other goals (or interests) that the agent might have, and irrespective of whether the agent is aware of their working. They write:

> 'Priming (passively and temporarily activating) an individual's internal goal representation affects subsequent judgments and behaviors in a manner consistent with him or her being in a motivated state (Bargh et al. 2001; Bargh et al. 2010; Dijksterhuis & Aarts 2011).'

and:

> 'Research suggests that people who are unaware that they are pursuing a goal respond to the world in a way that maximizes the likelihood of goal completion, such as by paying more attention to objects in the environment that would assist with goal pursuit and becoming predisposed to like and physically approach those objects. Goals operate autonomously (i.e., independent of guidance from the conscious individual) through these mechanisms to encourage achievement of their associated end-states' (Huang & Bargh 2014, p. 123)

In applying the goal contagion model to spontaneous helping scenarios, then, we appear to arrive at the following account:

1. The infant identifies the goal *in the rich sense of a mental representation guiding the action* ('must get pencil').
2. This mental representation then exercises the same functional role that it plays in the observed agent, and functions to guide *the infant's* actions such as to bring about the outcome (i.e. to get the pencil).
3. This may lead the infant to get the pencil for herself.

For some actions, such as putting the books on the shelf, this will make no difference, but for some other ones, such as getting the pencil, it will. In other words, this model generates the empirically false prediction that the child will take up the goal of getting the pencil (for herself) rather than the goal of getting the experimenter the pencil.

3.3 Lean Goals and Goal Slippage

To avoid this consequence, we propose to consider how a goal slippage model may be articulated on the basis of the lean notion of goals. One way of doing so would be to draw upon the so-called 'common-coding approach', developed most prominently by Wolfgang Prinz (1997; cf. also Hommel, 2001; James, 1890). Common coding provides a framework within which to understand goal slippage on the basis of a thin characterization of goals. According to this common coding, perceptible events, such as a glass being filled with water, are represented in the same format as actions (such as the action of filling the glass with

water). As a consequence of this overlap in the representational formats of action and perception, the representation of an event activates the very same representations that would cause the motor system to initiate an action that would bring about that event. Thus, observing (or imagining) a glass being filled with water activates a motor program for filling the glass with water (see Hommel, 2001 for a review of evidence in support of the theory).

If this is correct, then the observation of an action may lead the observer (i.e. the infant in a spontaneous helping scenario) to identify the event that is the goal (i.e. putting the books in the cabinet), and as a result of activating this representation to initiate an action that will bring about this outcome.

The explanation that this provides of why an observer who identifies the outcome toward which an agent's movements are directed is motivated to perform an action directed toward bringing about that outcome does not appeal to the representations of the observed agent. This is an advantage. To see why, consider the example of the experimenter trying to grasp a pencil that is just beyond her reach. If the goal which the infant identifies in this case were characterized as a mental representation guiding the agent's action (i.e. in the rich sense of a goal), then she should take up a goal with the content ('get the pencil'), leading her to get the pencil for herself rather than for the experimenter. Instead, however, we arrive at the following account:

1. The infant observes an agent performing an action directed toward a particular outcome
2. The infant identifies the goal *in the lean sense of an outcome (event) toward which the agent's movements are directed* (i.e. the agent getting the pencil in her hand).
3. The representation of this possible outcome (event) activates the very same representations in the infant that would cause the motor system to initiate an action to bring about that event.
4. The infant initiates that action and brings it about that the agent winds up with the pencil in her hand.

In other words, by starting out from the lean notion of goals, the goal slippage model avoids the empirically false prediction generated by the goal contagion model and successfully explains a wide range of findings. It entails that once a goal is identified, the infant identifying the goal should become motivated to achieve the goal because s/he will treat it just like any other goal that s/he has. This explanation eliminates the need to postulate

any further motivational mechanism apart from those which move infants to act on endogenously arising goals.

As such, it generates a pattern of predictions that sets it apart from the models discussed above. Like the psychological altruism model, it predicts that rewards are not necessary for, and may even interfere with, instrumental helping. Unlike the psychological altruism model (but like the joint action preference model), it predicts that an infant would continue helping (or even protesting) if an agent were to become distracted, lose interest or otherwise abandon the goal. Unlike the joint action preference model, it predicts that infants will want to complete goals even if this does not involve joint action. Unlike the reputation management model, it correctly predicts anonymous helping (Hepach, 2017). Unlike the aversion to others' distress model, it does not generate the prediction that infants will look away to avoid being confronted with the agent in distress. Unlike the compulsive planning model, it does not predict that familiar actions should elicit less helping behavior.

There are however some problems with the goal slippage model as sketched here. One of them arises from the lean characterization of goals. As noted above, the very same movements can function to bring about different outcomes on different occasions, depending on many features of the context, including the mental states of the observed agent (Jacob & Jeannerod, 2005 Michael & Christensen, 2016). How, then, does the infant determine which possible outcome is the one toward which the agent's movements are directed? For example, how does an infant identify that the agent's goal is to put the books into the cupboard? One way in which this problem could in principle be solved would be to identify the goal as the outcome that one has most frequently observed as an end state of the type of movement that one is currently observing[5]. If the infants in instrumental helping scenarios rely on something like this, then it should be possible to influence infants' helping behavior by manipulating the frequency with which particular types of movements are paired with particular goals in a familiarization phase.

A second problem arises from the appeal made here to common coding as a means of articulating goal slippage. There are often many different motor possibilities for bringing about an outcome (grasping the pencil with the right or the left hand; grasping it towards the end or in the middle; holding it up and letting the other agent reach out to take it or extending the arm to place it in their hand). Selecting one action plan will likely to depend on various contextual details (Is one's right hand already occupied? How is the pencil currently situated?

[5] In order to evaluate this conjecture, it would be necessary to spell out how goals and movements types are individuated, and in particular at what fineness of grain they are represented in computing the frequency of movement-goal pairs).

How far away and how tall is the other agent?), so it is difficult to see how one action plan could be uniquely linked to the outcome, and therefore which motor command would be elicited by the anticipation of the outcome.

Third, this account does not provide an explanation that specifically addresses the motivation to perform actions leading to outcomes that are the goals of other agents. It provides a general explanation of why, whenever an agent represents any event that she could possibly bring about through an action in her repertoire, this should trigger an impulse to perform the action and to bring about the event.

Fourthly, and relatedly, it appears to predict that people will be motivated to perform actions to bring about any event at all that they are led to imagine, as long as it is one that they could bring about through an action of their own. This is too broad. After all, one thinks of possible events all the time without then bringing them about. Indeed, sometimes one thinks of events in order to *prevent* them. The goal slippage model therefore needs to build in a mechanism for inhibiting actions in a great many cases. If this is correct, then it should be possible to selectively interfere with the inhibiting mechanism and thereby increase spontaneous helping behavior. One way to test this prediction would be to increase cognitive load through the introduction of a secondary task: spontaneous helping behavior should be less likely to occur when executive resources are occupied (e.g. under cognitive load). Unfortunately, such an approach would probably not be suitable for an experiment with very young children. However, if some version of the goal alignment account is right, then it should be possible to increase helping behavior in adults through the imposition of cognitive load. In order to implement such a test, it would be useful to develop an instrumental helping paradigm that could – like the secondary task – be performed while seated at a computer.

4. Conclusions

We have compared and contrasted several distinct models of the psychological mechanisms underpinning spontaneous instrumental helping in infancy. We have aimed to illuminate the motivations for considering these models, and also the theoretical commitments and empirical predictions that they generate. In some cases, this has led us to formulate theoretical objections and to point out where empirical findings are not consistent with the predictions generated by a model. We do not believe that any of these objections or predictive shortcomings are decisive; our aim in discussing them has been to indicate how further research may evaluate them and/or to revise them.

Both of the goal alignment models that we have discussed, namely goal contagion and goal slippage, have in common that once a goal is identified, the observer identifying the goal should become motivated to achieve it because s/he will treat it just like any other goal that s/he has. Thus, in contrast to the models discussed in section 2, these models do not postulate any further motivational mechanism apart from those which move infants to act on endogenously arising goals. This may make them suitable for explaining some cognitively undemanding forms of prosocial behavior (i.e. which may qualify as altruistic in the evolutionary sense but not the psychological sense). The goal slippage model has the further attractive feature that it, in contrast to the goal contagion model, does not generate the empirically false prediction that observers (i.e. infants in instrumental helping scenarios) will wind up competing with the agents whose goals they take up. Instead, it generates a unique pattern of predictions that sets it apart from the other models, and which may provide a fruitful impulse for further research investigating the ontogeny of the psychological underpinnings of human cooperation.

Acknowledgements

We are grateful for the very generous and constructive comments we received from two anonymous reviewers, and from Hemdat Lerman, Jack Shardlow, Alex Green, James Brown, Sam Clarke, Ian Phillips, Nick Shea, and the participants of the 'Mind Work in Progress Seminar' at Oxford. John Michael and Marcell Székely were both supported by a Starting Grant from the European Research Council (nr 679092, SENSE OF COMMITMENT).

References

Aarts, H., & Dijksterhuis, A. P. (2000). The automatic activation of goal-directed behaviour: The case of travel habit. *Journal of Environmental Psychology*, *20*(1), 75-82.

Aarts H, Gollwitzer PM, Hassin RR (2004) Goal contagion: Perceiving is for pursuing. *Journal of Personality and Social Psychology* 87: 23–37.

Aknin, L. B., Hamlin, J. K., & Dunn, E. W. (2012). Giving leads to happiness in young children. *PLoS One*, *7*(6), e39211.

Axelrod, R. and Hamilton, W. D., 1981, 'The Evolution of Cooperation', *Science*, 211: 139-96.

Axelrod. R., 1984, *The Evolution of Cooperation*, New York: Basic Books.

Bargh, J. A., Gollwitzer, P. M., Lee-Chai, A., Barndollar, K., & Trötschel, R. (2001). The automated will: nonconscious activation and pursuit of behavioral goals. *Journal of personality and social psychology*, *81*(6), 1014.

Barresi, J., & Moore, C. (1996). Intentional relations and social understanding. *Behavioral and brain sciences*, *19*(01), 107-122.

Batson, C. D., Fultz, J. and Schoenrade, P. A. (1987), Distress and Empathy: Two Qualitatively Distinct Vicarious Emotions with Different Motivational Consequences. *Journal of Personality*, 55: 19–39. doi:10.1111/j.1467-6494.1987.tb00426.x

Batson, C. D., Ahmad, N., Powell, A. A., & Stocks, E. L. (2008). Prosocial motivation. In J. Y. Shah & W. L. Gardner (Eds.), *Handbook of motivation science* (pp. 135– 149). New York: Guilford.

Beck, S. R., Apperly, I. A., Chappell, J., Guthrie, C., & Cutting, N. (2011). Making tools isn't child's play. *Cognition*, *119*(2), 301-306.

Bird, C. D., & Emery, N. J. (2009). Insightful problem solving and creative tool modification by captive nontool-using rooks. *Proceedings of the National Academy of Science of the United States of America*, 106, 10370–10375.

Bischof-Köhler, D. (1994). Self object and interpersonal emotions. Identification of own mirror image, empathy and prosocial behavior in the 2nd year of life. *Zeitschrift für experimentelle und angewandte Psychologie*, 202, 349–377.

Brownell, C. A., Svetlova, M., & Nichols, S. (2009). To share or not to share: When do toddlers respond to another's needs? *Infancy*, 14, 117–130. doi:10.1080/15250000802569868

Buchsbaum, D., Bridgers, S., Weisberg, D. S., & Gopnik, A. (2012). The power of possibility: Causal learning, counterfactual reasoning, and pretend play. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1599), 2202-2212.

Butterfill, S. (2012). Joint action and development. *Philosophical Quarterly,* 62 (246), 23-47.

Butterfill, S. and Apperly, I. (2013). How to construct a minimal theory of mind. *Mind and Language*, 28 (5), 606-637.

Dahl, A. (2015). The developing social context of infant helping in two US samples. *Child development*, *86*(4), 1080-1093.

Davidov, M., Zahn-Waxler, C., Roth-Hanania, R., & Knafo, A. (2013). Concern for others in the first year of life: Theory, evidence, and future directions. *Child Development Perspectives*, 7, 126–131. doi:10.1111/cdep.12028

Dunfield, K. (2014). A construct divided: prosocial behavior as helping, sharing and comforting subtpyes. *Frontiers in Psychology*, volume5, article 958, doi: 10.3389/fpsyg.2014.00958

Dunfield, K. A., & Kuhlmeier, V. A. (2013). Classifying prosocial behavior: Children's responses to instrumental need, emotional distress, and material desire. *Child Development*, 84, 1766–1776. doi:10.1111/cdev.12075

Dunfield, K., Kuhlmeier, V. A., O'Connell, L., & Kelley, E. (2011). Examining the diversity of prosocial behavior: Helping, sharing, and comforting in infancy. *Infancy*, 16, 227–247. doi:10.1111/j.1532-7078.2010.00041.x

Eisenberg, N., Guthrie, I. K., Murphy, B. C., Shepard, S. A., Cumberland, A., & Carlo, G. (1999). Consistency and development of prosocial dispositions: A longitudinal study. *Child Development*, 70, 1360–1372. doi:10.1111/1467-8624.00100

Emery, N. J., & Clayton, N. S. (2009). Tool use and physical cognition in birds and mammals. *Current Opinion in Neurobiology*, 19, 27–33.

Engelmann, J. M., Herrmann, E., & Tomasello, M. (2012). Five-year olds, but not chimpanzees, attempt to manage their reputations. *PLoS One*, *7*(10), e48433.

Engelmann, J. M., Over, H., Herrmann, E., & Tomasello, M. (2013). Young children care more about their reputation with ingroup members and potential reciprocators. *Developmental Science*, *16*(6), 952-958.

Foster, K. R., Wenseleers, T., & Ratnieks, F. (2006). Kin selection is the key to altruism. *Trends in Ecology and Evolution*, 21(2), 57–60.

Grusec, J. E. (2006). The development of moral behavior and conscience from a socialization perspective. In M. Killen & J. G. Smetana (Eds.), *Handbook of moral development* (pp. 243–265). Mahwah, NJ: Erlbaum.

Grusec, J. E., Davidov, M., & Lundell, L. (2002). Prosocial and helping behavior. In P. K. Smith & C. H. Hart (Eds.), *Blackwell handbook of childhood social development* (pp. 457–474). Malden, MA: Blackwell.

Hamilton, W. D., 1964, 'The Genetical Evolution of Social Behaviour I and II', *Journal of Theoretical Biology*, 7: 1–16, 17–32.

——, 1970, 'Selfish and Spiteful Behaviour in an Evolutionary Model', *Nature*, 228: 1218-1220.

——, 1972, 'Altruism and Related Phenomena, mainly in the Social Insects', *Annual Review of Ecology and Systematics*, 3: 193–232.

Hay, D. F., Caplan, M., Castle, J., & Stimson, C. A. (1991). Does sharing become increasingly" rational" in the second year of life?. *Developmental Psychology*, *27*(6), 987.

Hay, D. F., Nash, A., & Pederson, J. (1981). Responses of six-month- olds to the distress of their peers. *Child Development*, 52, 1071– 1075. doi:10.2307/1129114

Hepach, R., Vaish, A., & Tomasello, M. (2012). Young children are intrinsically motivated to see others helped. *Psychological Science*, 23(9), 967-972.

Hepach, R., Vaish, A., Grossmann, T., & Tomasello, M. (2016). Young children want to see others get the help they need. *Child Development*, *87*(6), 1703-1714.

Hepach, R., Haberl, K., Lambert, S., & Tomasello, M. (2017). Toddlers Help Anonymously. *Infancy*, *22*(1), 130-145.

Hoffman, M. L. (1975). Developmental synthesis of affect and cognition and its interplay for altruistic motivation. *Developmental Psychology*, 11, 607–622. doi:10.1037/0012-1649.11.5.607

Hommel, B., Müsseler, J., Aschersleben, G., & Prinz, W. (2001). Codes and their vicissitudes. *Behavioral and brain sciences*, *24*(05), 910-926.

Huang, J. Y., & Bargh, J. A. (2014). The Selfish Goal: Autonomously operating motivational structures as the proximate cause of human judgment and behavior. *Behavioral and Brain Sciences*, *37*(02), 121-135.

Jacob, P., & Jeannerod, M. (2005). The motor theory of social cognition: a critique. *Trends in cognitive sciences*, *9*(1), 21-25.

James, W. (1890/1981) *The principles of psychology.* Macmillan/Harvard University Press. (Original work published 1890).

Kärtner, J., & Keller, H. (2012). Culture-specific developmental path- ways to prosocial behavior: A comment on Bischof-Köhler's univer- salist perspective. *Emotion Review*, 4, 49–50. doi:10.1177/ 1754073911421383

Kärtner, J., Keller, H., & Chaudhary, N. (2010). Cognitive and social influences on early prosocial behavior in two sociocultural con- texts. *Developmental Psychology*, 46, 905–914. doi:10.1037/ a0019718

Kenward, B., & Gredebäck, G. (2013). Infants help a non-human agent. *PLoS ONE*, 8, e75130. doi:10.1371/journal.pone.0075130

Levitt, M. J., Weber, R. A., Clark, M. C., & McDonnell, P. (1985). Reciprocity of exchange in toddler sharing behavior. *Developmental Psychology*, *21*(1), 122.

Liszkowski U, Carpenter M, Tomasello M. 2007. Pointing out new news, old news, and absent referents at 12 months of age. *Developmental. Science,* 10, F1–7

Martin, A., & Olson, K. R. (2013). When kids know better: paternalistic helping in 3-year-old children. *Developmental psychology*, *49*(11), 2071.

Maynard Smith, J., 1964, 'Group Selection and Kin Selection', *Nature*, 201: 1145–1147.

——, 1974, 'The Theory of Games and the Evolution of Animal Conflicts, *Journal of Theoretical Biology*, 47: 209–21.

——, 1982, *Evolution and the Theory of Games*, Cambridge: Cambridge University Press.

——, 1998, 'The Origin of Altruism', *Nature*, 393: 639–640.

Michael, J, & Christensen, W. (2016), Flexible goal attribution in early mindreading, *Psychological Review* Vol. 123, No. 2, 219–227.

Michael, J., Sebanz., N. & Knoblich, K. (2016). The Sense of Commitment: A Minimal Approach, *Frontiers in Psychology* 6, 1968, DOI: 10.3389/fpsyg.2015.01968

Millikan, R. G. (1984). *Language, thought, and other biological categories: New foundations for realism*. MIT press.

Paulus, M. (2014). The emergence of prosocial behavior: why do infants and toddlers help, comfort, and share? *Child Development Perspectives*, *8*(2), 77-81.

Paulus, M., & Moore, C. (2012). Producing and understanding prosocial actions in early

childhood. *Advances in Child Development and Behavior*, 42, 275–309. doi:10.1016/B978-0-12-394388-0.00008-3

Paulus, M., & Moore, C. (2014). The development of sharing behavior and expectations about other people's sharing in preschool children. *Developmental Psychology*, 50, 914–921. doi:10.1037/a0034169 11

Paulus, M., Gillis, S., Li, J., & Moore, C. (2013). Preschool chil- dren involve a third party in a dyadic sharing situation based on fairness. *Journal of Experimental Child Psychology*, 116, 78–85. doi:10.1016/j.jecp.2012.12.014

Piliavin, J. A., & Charng, H.-W. (1990). Altruism: A review of recent theory and research. *Annual Review of Sociology*, 16, 27–65.

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

Rakoczy, H., & Schmidt, M. F. (2013). The early ontogeny of social norms. *Child Development Perspectives*, *7*(1), 17-21.

Rheingold, H. L. (1982). Little children's participation in the work of adults, a nascent prosocial behavior. *Child Development*, 53, 114–125.

Roberts, G. (2005). Cooperation through interdependence. *Animal Behaviour*, *70*(4), 901-908.
Rochat, P., Broesch, T., & Jayne, K. (2012). Social awareness and early self-recognition. *Consciousness and cognition*, *21*(3), 1491-1497.

Roth-Hanania, R., Davidov, M., & Zahn-Waxler, C. (2011). Empathy development from 8 to 16 months: Early signs of concern for others. *Infant Behavior and Development*, 34, 447–458. doi:10.1016/ j.infbeh.2011.04.007

Schmidt, M. F., Rakoczy, H., & Tomasello, M. (2012). Young children enforce social norms selectively depending on the violator's group affiliation. *Cognition*, *124*(3), 325-333.

Schmidt, M. F. H., & Sommerville, J. A. (2011). Fairness expectations and altruistic sharing in 15-month-old human infants. *PLoS ONE*, 6, 23223. doi:10.1371/journal.pone.0023223

Schulz, L. (2015). Infants explore the unexpected. Comment on Observing the unexpected enhances infants' learning and exploration. *Science*, 3 (6230), 42-43.

Silk, J., Brosnan, S., Vonk, J., Henrich, J., Povinelli, D., Richardson, A. S., et al. (2005). Chimpanzees are indifferent to the welfare of unrelated group members. *Nature*, 437, 1357–1359.

Silk, J. (2009). Nepotistic cooperation in non-human primate groups. *Philosophical Transactions of the Royal Society* B, 364, 3243–3254.

Svetlova, M., Nichols, S. R., & Brownell, C. A. (2010). Toddlers' prosocial behavior: From instrumental to empathic to altruistic helping. *Child development*, *81*(6), 1814-1827.

Tomasello, M. (2016). *A natural history of human morality*. Harvard Univ. Pr.

---- (2009), *Why we cooperate*, Cambridge: MIT Press.

Tomasello, M., Hare, B., Lehmann, H., & Call, J. (2007). Reliance on head versus eyes in the gaze following of great apes and human infants: the cooperative eyehypothesis. *Journal of Human Evolution*, 52(3), 314-320.

Tinbergen, Niko (1963) "On Aims and Methods of Ethology," *Zeitschrift für Tierpsychologie*, 20: 410–433.

Warneken, F., Chen, F., & Tomasello, M. (2006). Cooperative activities in young children and chimpanzees. *Child development*, 77(3), 640-663.

Warneken F, Tomasello M (2007). Helping and cooperation at 14 months of age. *Infancy,* 11, 271–294. doi: 10.1111/j.1532-7078.2007.tb00227.x

Warneken, F., Hare, B., Melis, A. P., Hanus, D., & Tomasello, M. (2007). Spontaneous

altruism by chimpanzees and young children. *PLoS Biol*, *5*(7), e184.

Warneken, F. (2013). Young children proactively remedy unnoticed accidents. *Cognition*, 126(1), 101–108.

Warneken, F. (2015). Insights into the biological foundation of human altruistic sentiments. *Current Opinion in Psychology*, 7, 51–56.

Warneken, F., & Tomasello, M. (2006). Altruistic helping in human infants and young chimpanzees. *Science*, 311, 1301–1303.

Warneken, F., & Tomasello, T. (2008). Extrinsic rewards undermine altruistic tendencies in 20-month-olds. *Developmental Psychology*, 44(6), 1785–1788.

Warneken, F., & Tomasello, M. (2009). Varieties of altruism in children and chimpanzees. *Trends in Cognitive Science*, 13, 397–402.

Warneken, F., & Tomasello, M. (2013). Parental presence and encouragement do no influence helping in young children. *Infancy*, *18*(3), 345-368.

Weir, A. A. S., Chappell, J., & Kacenik, A. (2002). Shaping of hooks in New Caledonian Crows. *Science*, 291.

Zahn-Waxler, C., Radke-Yarrow, M., Wagner, E., & Chapman, M. (1992). Development of concern for others. *Developmental Psychology*, 28, 126–136. doi:10.1037//0012-1649.28.1.126