Explaining the U-Shaped Development of Intent-Based Moral Judgments

Francesco Margoni* and Luca Surian

Department of Psychology and Cognitive Sciences, University of Trento, Rovereto, Italy

When preschoolers evaluate actions and agents, they typically neglect agents' intentions and focus on action outcomes instead. By contrast, intentions count much more than outcomes for older children and adults. This phenomenon has traditionally been seen as evidence of a developmental change in children's concept of what is morally good and bad. However, a growing number of studies shows that infants are able to reason about agents' intentions and take them into account in their spontaneous socio-moral evaluations. Here we argue that this puzzling U-shaped trajectory in children's judgments is best accounted for by a model that posits developmental continuity in moral competence and emphasizes the effect of immature executive function skills on preschoolers' performance.

Keywords: moral judgment, moral development, intention, executive function, theory of mind

Mental state reasoning is required in several tasks, from inferential communication and the interpretation of social situations to the socio-moral evaluation of actions and agents. Children at an early age start to accuse peers by crying loudly "you did it on purpose!," and legal systems typically distinguish harmful acts that are performed intentionally from acts that accidentally produce personal harm. A large body of developmental research investigated when and how children acquire the ability to attend to agents' intentions and action outcomes in their socio-moral judgments, but the conclusions one can draw from infant studies seem at odds with the conclusions one can draw from studies on older children. Infants seem to possess abilities that young preschoolers' responses do not reveal. In the present work, we address this puzzle by first reviewing relevant results on socio-moral reasoning in infants and children. Then, we evaluate different proposals put forward to explain the reported developmental changes and the apparent contradiction between infants and preschool children's responses.

THE OUTCOME-TO-INTENT SHIFT IN PRESCHOOLERS' MORAL REASONING

Since Piaget's (1932/1965) seminal work, a large body of studies has shown that a crucial developmental change from an outcome- to an intent-based moral evaluation occurs in the late preschool years. A typical Piagetian task would consist of evaluating which of two characters is more naughty and deserves to be punished. Piaget presented children with a story in which a supposedly well-intentioned character accidentally caused serious material damage (e.g., he broke 15 cups), and another story in which a bad-intentioned character caused, also by accident, less serious damage (e.g., he broke one cup). Younger children (aged 6–7) judged the character that

produced serious material damage to be more naughty and punishable, whereas older children judged the bad-intentioned one to be more naughty and punishable. These and other similar findings were taken as evidence of a shift from an initial outcome-based ('objective') moral judgment to a later intent-based ('subjective') moral judgment.

Subsequent research overcame several methodological limitations of Piaget's work (King, 1971; Farnill, 1974; Karniol, 1978; Nelson, 1980), but confirmed the occurrence of an outcome-to-intent shift. During the 1970s and 1980s, many different tasks were developed to investigate this phenomenon. By reducing the cognitive processing necessary to answer experimenters' questions, scholars found that even preschoolers, at age 3, can attend to agents' intentions in their moral evaluations (e.g., Armsby, 1971; Farnill, 1974; Yuill, 1984; Yuill and Perner, 1988). Nevertheless, Piaget's main claim concerning the outcome-to-intent shift found further support, since children older than 4–5 years relied more on intention and less on outcome, whereas younger children showed the opposite pattern (e.g., Costanzo et al., 1973; Imamoglu, 1975; Keasey, 1978; Moran and O'Brien, 1983; Baird and Astington, 2004; Wainryb et al., 2005; Nobes et al., 2009; Cushman et al., 2013).

When intentions and outcomes lead to conflicting responses, as in the cases of *failed attempts to harm* and *accidental harm,* young preschoolers attend to outcome more than older children, relying mostly on outcome (e.g., Helwig et al., 1995), or equally on intention and outcomes (e.g., Killen et al., 2011; Cushman et al., 2013). With age, the condemnation of attempted but failed harm increases (Helwig et al., 1995), whereas the condemnation of accidental harm decreases (Cushman et al., 2013; see also Killen et al., 2011 on the development of an intent-based punishability evaluation).

While intentions dominate adults' attribution of moral goodness and badness, adults often rely on both intent and outcomes to evaluate the punishability of agents (Cushman, 2008). A recent dual-process model explains why this is so: adults' moral reasoning is generated by the work of two independent and sometimes conflicting processes, one that attributes value to actions and assesses agents' mental states, and the other that evaluates the causal responsibility for action outcomes (Cushman, 2013, 2015). This proposal contradicts the Piagetian view, which posited a full replacement of the outcome-based judgment by an intent-based judgment.

INFANTS' INTENT-BASED SOCIO-MORAL EVALUATIONS

Extrapolating the developmental trajectory found in preschoolers, one may predict that infants and toddlers would rely mostly on action outcome rather than agents' intention, assuming that they can produce a moral judgment. However, recent evidence shows that this is not the case. Several studies suggest that, in the first year, infants are able to distinguish between intentions and outcomes, they evaluate helping, harming and distributive actions, and they rely, for these evaluations, on intentions rather than outcomes.

Experimental studies used both elicited-response tasks and spontaneous-response tasks in the *violation-of-expectation paradigm*. Both research strategies found that, by the end of the first year, infants are able to attend to agents' intentions and understand successful as well as failed actions (e.g., Woodward, 1998; Gergely and Csibra, 2003; Brandone and Wellman, 2009). This early understanding of failed attempts generalizes to first- as well as third-party socio-moral evaluations (Behne et al., 2005; Dunfield and Kuhlmeier, 2010; Hamlin, 2013; Lee et al., 2015).

In studies on *first-party* evaluations, infants were engaged in interactions with an experimenter and were presented with actors that were either unwilling or unable to please them (e.g., Behne et al., 2005; Dunfield and Kuhlmeier, 2010; Marsh et al., 2010). While the outcomes were identical in both conditions, intentions were different (negative for 'unwilling agents,' positive for 'unable agents'). Infants responded differently to these two cases, showing that they used intention cues to guide their *first-party* evaluations and preferences. Nine-month-olds' spontaneous signals of impatience (such as reaching and banging or looking and turning away) revealed that they become more agitated when they interact with actors unwilling to provide them with a toy (Behne et al., 2005). Moreover, using a manual choice measure (infants have to choose between two contrasted individuals), some studies found that by the second year of life, infants choose to help an unable over an unwilling actor, when asked to help someone. By contrast, infants were equally likely to help able agents, who displayed positive intentions and successful actions, and unable agents, who displayed positive intentions and unsuccessful actions (Dunfield and Kuhlmeier, 2010). Overall, these studies show that infants process information about intention and use it to evaluate others' behavior.

Further studies on infants' representations of harm and help examined *third-party* socio-moral evaluations (e.g., Kuhlmeier et al., 2003; Hamlin et al., 2007, 2013; Hamlin and Wynn, 2011; Meristo and Surian, 2014). Infants observed events in which an agent either helps or hinders the goal-directed action of another agent. Their evaluations of prosocial and antisocial actors were typically tested using a manual preference task. Early in their first year of life, infants consistently prefer the helper over the hinderer (Hamlin et al., 2007; Wynn, 2008). Also, at 16 months they prefer agents performing fair over unfair distributive actions (e.g., Geraci and Surian, 2011). When evaluating an agent's behavior, infants are able to take into account not only a person's intention, but also other relevant mental states such as informational states and beliefs (Hamlin et al., 2013; Meristo and Surian, 2013; Choi and Luo, 2015; for a review: Baillargeon et al., 2015).

Hamlin (2013; see also Hamlin et al., 2013) played a puppet show in which puppets either try but fail or succeed to help (or hinder) someone's goal-directed action. Eight-month-olds preferred a helper (failed or successful) over a hinderer, but, most importantly here, infants did not prefer the successful helper (displaying both intention and relevant outcome) over the puppet that attempted to help, but failed (showing a good intention, but no relevant outcome). This suggests that infants' preferences were guided by agents' intentions rather than outcomes. Moreover, studying expectations by measuring

spontaneous looking behavior, scholars recently found that by the end of the first year, infants infer agents' socio-moral preferences by taking into account the agents' information about others' prosocial and antisocial intentions (Lee et al., 2015). They expect that an agent would prefer to approach a second agent who has previously shown a good intention, no matter what the consequences of the second agent's action were.

How can we reconcile the classic results of preschoolers' outcome-to-intent shift with these recent results of infants' intent-based expectations and evaluations? A lesson may be learned from the literature on theory of mind.

HOW TO ACCOUNT FOR SEEMINGLY CONFLICTING RESULTS: THE CASE OF FALSE BELIEFS TASKS

The description of the intent-based judgment development sketched above, that is, an initial intent-based evaluation developing from an outcome-based evaluation that in turn shifts again toward an intent-based evaluation, resembles the '*puzzle about belief*' (Perner and Roessler, 2012), that is, the puzzle regarding the development of Theory of Mind. Using traditional elicited-response tasks to study children's attributions of false beliefs, researchers initially concluded that the ability to attribute false beliefs does not emerge until about the fourth birthday (e.g., Wimmer and Perner, 1983; Baron-Cohen et al., 1985; Wellman et al., 2001). However, using *violation-of-expectation* and *anticipatory-looking* spontaneous-response tasks, researchers began to study also infants' mentalizing abilities. Using these and others tasks, scholars demonstrated that babies at least in their second year of life are able to attribute reality congruent and incongruent mental representations across several situations (Southgate et al., 2007; Surian et al., 2007; Buttelmann et al., 2009; Baillargeon et al., 2010; Luo, 2011; Low and Perner, 2012; Surian and Geraci, 2012).

Why do 3-year-olds fail to attribute false beliefs when their abilities are tested on elicited-response tasks? There are two possible answers to this question. First, one may posit continuity during development and argue that preschoolers fail because they do not have the necessary executive function skills to pass an elicited-response task. Second, one may posit a conceptual change during development and argue that the representations and processes involved in resolving spontaneous-response tasks are fundamentally different from the ones involved in resolving elicited-response tasks.

Young preschoolers may succeed in representing the agent's false belief, as infants do, but fail to select the right response and inhibit the wrong response when they are questioned via an elicited-response task (Leslie and Polizzi, 1998; Baillargeon et al., 2010). An *innate modular account* posits that from an early age babies are able to represent and use others' mental states to understand social situations (Leslie et al., 2005; Surian et al., 2007; Kovács et al., 2010). What really develops is the set of cognitive abilities that children need to exploit their representational skills. At 3 years, executive function skills are not sufficiently developed to meet the processing demands of the

elicited-response tasks (Thoermer et al., 2012). The continuity account is receiving growing experimental support, but it is still controversial. Many argue for a *conceptual shift account* according to which infants that pass a spontaneous-response task show a qualitatively different level of understanding compared to children who pass an elicited-response task (e.g., Wellman, 2014).

RECONCILING RESULTS ON INFANTS AND CHILDREN AT THE PROCESSING LEVEL

As in the literature on false–belief understanding, we can draw a distinction between two main positions. First, an *emergence* view posits that during preschool years a conceptual change occurs in moral competence. The construction of a novel conceptual competence explains why school-aged children's judgments differ from preschoolers. This view does not deny the role of executive function skills, as these are certainly involved in theory construction and revision processes (Cushman et al., 2013). Second, an *expression* view posits conceptual continuity during development and sees the role of executive function in a very different way. It claims that developmental differences result solely from changes in executive function, or theory of mind, that are external to the moral competence (Zelazo et al., 1996; Chandler et al., 2001; Killen et al., 2011). We argue that the studies on infants' spontaneous socio-moral evaluations we briefly reviewed above favor the latter view and challenge the former, assuming that a "rich interpretation" (Aslin, 2000) of the infant studies is the correct one. Studies on infants' evaluations suggest that infants can employ an intent-based concept of moral goodness and badness in their socio-moral evaluations. Therefore, the development of intent-based moral judgment is unlikely to derive from a conceptual change occurring in the preschool years.

If infants already possess an intent-based concept of moral badness and goodness, can executive limitations account for preschoolers' outcome-based judgments? The expression view claims that young preschoolers fail at weighting intentions more than outcomes because of processing demands of the task. The additional processing demands of the elicited-response tasks compared to the spontaneous-response task used in the infant literature, lead kindergartners to produce outcome-based evaluations. With the acquisition of sufficient executive function (roughly at 4), children's responses on elicited-response tasks can gradually match infants' spontaneous ones, and become mostly intent-based. When judging an action or an agent in elicited-response tasks, for preschoolers it is difficult to suppress cues concerning action outcomes, while older children may have the sufficient executive function abilities to inhibit an outcome-based judgment and select an intent-based response (or *set shift* to an intent-based response, see Miyake et al., 2000; Diamond, 2013).

Highlighting how different tasks tap different forms of evaluation, and distinguishing between elicited and spontaneous responses may provide the key to solving the *puzzle about intent-based moral judgment*, and avoiding two conclusions

that appear highly implausible (Hamlin, 2013). First, it would be implausible that an early tendency to privilege intentions over outcomes emerges during infancy only to be replaced during preschool years by the opposite tendency to privilege outcomes over intentions, or to weight these cues equally, and eventually be again replaced with a final tendency to privilege intentions. Second, it would be also very odd to posit that infants' evaluation system is not related to the later evaluation system, so that we would have two intent-based moral evaluations mutually independent. Conversely, it is likely that if an evaluation system emerges, it will not be replaced later in the development by a new system that serves the exact same function. In order to test directly the expression account, future research should also investigate young preschoolers' generation of intent-based moral evaluations in spontaneous-response tasks.

ADDITIONAL FACTORS THAT MAY AFFECT THE OUTCOME-TO-INTENT SHIFT

In an expression view, several internal and external factors may promote the emergence of an intent-based elicited response. Among the internal factors, we can include the frontal lobe maturation underlying the acquisition of executive functions (Huttenlocher and Dabholkar, 1997; Benes, 2001; Miller and Cohen, 2001; Moriguchi and Hiraki, 2009, 2011; Moriguchi, 2014; but see also Knight and Stuss, 2002; Lepsien and Nobre, 2006). Among the plausible external factors, one could include interactions with adults and peers (e.g, Tomasello et al., 2005). Preschoolers start to be considered somewhat responsible for their actions by their parents, and parents correct their behaviors by pointing to actions outcomes (Piaget, 1932/1965) or negligence (Nobes et al., 2009). However, this may not be true for infants and older school-aged children. While infants' actions outcomes are limited in their valence and severity, and parents do not deem their children fully responsible for what they cause, older children develop a more controlled behavior, and parents or peers now privilege a comprehensive evaluation of children's intentions and outcomes.

Now, moving from an explanation concerning proximal causes (the processing level discussed in the previous section), to an explanation concerning distal causes (the evolutionary level), one promising perspective is offered by the *life-history theory*. Life-history theory is an approach in evolutionary biology that seeks to explain the timing of the organism's ontogenesis by linking it to relevant evolutionary pressures (Kaplan and Gangestad, 2005). The emergence of a trait in the phenotype has both costs and benefits for the organism with regards to its reproductive fitness. The timing of such emergence would optimize the costs/benefits trade-off by on-setting a certain trait at a particular age, rather than earlier or later. Originally, this perspective was employed to explain the timing of morphological and physiological traits, such

as sexual maturation, but recently it has been argued that it may also explain children's delay in acting accordingly to fairness principles (Sheskin et al., 2014). In fact, while infants appear to evaluate others following an implicit understanding of fairness and harm, only some years later they consistently apply those moral principles during their social interactions (Siegal, 1982).

How can one apply life-history theory to the development of intent-based moral reasoning? Advocates of life-history theory may want to claim that the elicited intent-based moral reasoning emerges roughly at 5–6 years because at this time-point children increasingly engage in social interactions with peers. To understand and properly evaluate others' intentions is fundamental in forming and maintaining such relationships. Attending to agents' intentions, rather than actions outcomes, in the evaluations of agents, may become crucial just at the age in which children, in the evolutionary past, could not rely anymore on 'free' resources provided by parents and had to rely on their interactions with peers, avoiding potentially dangerous conflicts (Marlowe, 2005). Therefore, life-history theory may explain both the growing concern for fairness and the growing reliance on agents' intention in preschoolers. The defendants of this position may then conclude that the human mind is wired with an innate ability to understand and evaluate others' intentions, but it is only during the late preschool years that this ability is systematically recruited by children in a variety of social interactions.

CONCLUSION

In sum, we have seen that young preschoolers' outcome-based judgment is preceded by an early capacity to evaluate intentions that is revealed in spontaneous-response tasks. Drawing a parallel with the literature on the acquisition of mental state reasoning, we argued that the outcome-to-intent shift is best explained by an expression account that posits an early emerging infant socio-moral competence and explains preschoolers' outcome-based judgments as due to immature domain-general executive function. Current evidence is more consistent with a view that assumes developmental continuity than with the opposite view based on conceptual changes.

AUTHOR CONTRIBUTIONS

All authors listed have made substantial, direct and intellectual contribution to the work, and approved it for publication.

ACKNOWLEDGEMENTS

REFERENCES

Armsby, R. E. (1971). A reexamination of the development of moral judgments in children. *Child Dev.* 42, 1241–1248. doi: 10.2307/1127807

Aslin, R. N. (2000). Why take the cog out of infant cognition? *Infancy* 1, 463–470. doi: 10.1207/S15327078IN0104_6

Baillargeon, R., Scott, R., and He, Z. (2010). False-belief understanding in infants. *Trends Cogn. Sci.* 14, 110–118. doi: 10.1016/j.tics.2009.12.006

Baillargeon, R., Scott, R. M., He, Z., Sloane, S., Setoh, P., Jin, K., et al. (2015). "Psychological and sociomoral reasoning in infancy," in *APA Handbook of Personality and Social Psychology: Attitudes and Social Cognition*, Vol. 1, eds M. Mikulincer, P. R. Shaver, E. Borgida, and J. A. Bargh (Washington, DC: American Psychological Association), 79–150.

Baird, J. A., and Astington, J. W. (2004). The role of mental state understanding in the development of moral cognition and moral action. *New Dir. Child Adolesc. Dev.* 103, 37–49. doi: 10.1002/cd.96

Baron-Cohen, S., Leslie, A. M., and Frith, U. (1985). Does the autistic child have a "theory of mind"? *Cognition* 21, 37–46. doi: 10.1016/0010-0277(85)90022-8

Behne, T., Carpenter, M., Call, J., and Tomasello, M. (2005). Unwilling versus unable: infants' understanding of intentional action. *Dev. Psychol.* 41, 328–337. doi: 10.1037/0012-1649.41.2.328

Benes, F. (2001). "The development of frontal cortex: the maturation of neurotransmitter systems and their interactions," in *Handbook of Developmental Cognitive Neuroscience*, eds C. Nelson and M. Luciana (Cambridge, MA: MIT Press), 79–92.

Brandone, A., and Wellman, H. M. (2009). You can't always get what you want: infants understand failed goal-directed actions. *Psychol. Sci.* 20, 85–91. doi: 10.1111/j.1467-9280.2008.02246.x

Buttelmann, D., Carpenter, M., and Tomasello, M. (2009). Eighteen-month-old infants show false belief understanding in an active helping paradigm. *Cognition* 112, 337–342. doi: 10.1016/j.cognition.2009.05.006

Chandler, M. J., Sokol, B. W., and Hallett, D. (2001). "Moral responsibility and the interpretive turn: children's changing conceptions of truth and rightness," in *Intentions and Intentionality: Foundations of Social Cognition*, eds B. F. Malle, L. J. Moses, and D. A. Baldwin (Cambridge, MA: MIT Press), 345–365.

Choi, Y., and Luo, Y. (2015). 13-month-olds' understanding of social interactions. *Psychol. Sci.* 26, 274–283. doi: 10.1177/0956797614562452

Costanzo, P. R., Coie, J. D., Grumet, J. F., and Farnill, D. (1973). A reexamination of the effects of intent and consequence on children's moral judgments. *Child Dev.* 44, 154–161. doi: 10.2307/1127693

Cushman, F. (2008). Crime and punishment: distinguishing the role of causal and intentional analyses in moral judgment. *Cognition* 108, 353–380. doi: 10.1016/j.cognition.2008.03.006

Cushman, F. (2013). Action, outcome, and value: a dual-system framework for morality. *Pers. Soc. Psychol. Rev.* 17, 273–292. doi: 10.1177/1088868313495594

Cushman, F. (2015). Deconstructing intent to reconstruct morality. *Curr. Opin. Psychol.* 6, 97–103. doi: 10.1016/j.copsyc.2015.06.003

Cushman, F., Sheketoff, R., Wharton, S., and Carey, S. (2013). The development of intent-based moral judgment. *Cognition* 127, 6–21. doi: 10.1016/j.cognition.2012.11.008

Diamond, A. (2013). Executive functions. *Annu. Rev. Psychol.* 64, 135–168. doi: 10.1146/annurev-psych-113011-143750

Dunfield, K. A., and Kuhlmeier, V. A. (2010). Intention-mediated selective helping in infancy. *Psychol. Sci.* 21, 523–527. doi: 10.1177/0956797610364119

Farnill, D. (1974). The effects of social judgment set on children's use of intent information. *J. Pers.* 42, 276–289. doi: 10.1111/j.1467-6494.1974.tb00674.x

Geraci, A., and Surian, L. (2011). The developmental roots of fairness: infants' reactions to equal and unequal distributions of resources. *Dev. Sci.* 14, 1012–1020. doi: 10.1111/j.1467-7687.2011.01048.x

Gergely, G., and Csibra, G. (2003). Teleological reasoning in infancy: the naïve theory of rational action. *Trends Cogn. Sci.* 7, 287–292. doi: 10.1016/S1364-6613(03)00128-1

Hamlin, J. K. (2013). Failed attempts to help and harm: intention versus outcome in preverbal infants' social evaluations. *Cognition* 128, 451–474. doi: 10.1016/j.cognition.2013.04.004

Hamlin, J. K., Ullman, T., Tenenbaum, J., Goodman, N., and Baker, C. (2013). The mentalistic basis of core social cognition: experiments with preverbal infants and a computational model. *Dev. Sci.* 16, 209–226. doi: 10.1111/desc.12017

Hamlin, J. K., and Wynn, K. (2011). Young infants prefer prosocial to antisocial others. *Cogn. Dev.* 26, 30–39. doi: 10.1016/j.cogdev.2010.09.001

Hamlin, J. K., Wynn, K., and Bloom, P. (2007). Social evaluation by preverbal infants. *Nature* 450, 557–559. doi: 10.1038/nature06288

Helwig, C. C., Hildebrandt, C., and Turiel, E. (1995). Children's judgments about psychological harm in social context. *Child Dev.* 66, 1680–1693. doi: 10.2307/1131903

Huttenlocher, P., and Dabholkar, A. (1997). "Developmental anatomy of prefrontal cortex," in *Development of the Prefrontal Cortex: Evolution, Neurobiology, and Behavior*, eds N. Krasnegor, G. Lyon, and P. Goldman-Rakic (Baltimore: Brookes Publishing), 69–83.

Imamoglu, E. O. (1975). Children's awareness and usage of intention cues. *Child Dev.* 46, 39–45. doi: 10.2307/1128831

Kaplan, H. S., and Gangestad, S. W. (2005). "Life history theory and evolutionary psychology," in *The Handbook of Evolutionary Psychology*, ed. D. M. Buss (Hoboken, NJ: Wiley), 68–95.

Karniol, R. (1978). Children's use of intention cues in evaluating behavior. *Psychol. Bull.* 85, 76–85. doi: 10.1037/0033-2909.85.1.76

Keasey, C. B. (1978). Young children's attribution of intentionality to themselves and others. *Child Dev.* 48, 261–264. doi: 10.2307/1128907

Killen, M., Mulvey, K. L., Richardson, C., Jampol, N., and Woodward, A. (2011). The accidental transgressor: morally-relevant theory of mind. *Cognition* 119, 197–215. doi: 10.1016/j.cognition.2011.01.006

King, M. (1971). The development of some intention concepts in young children. *Child Dev.* 42, 1145–1152. doi: 10.2307/1127799

Knight, R., and Stuss, D. (2002). "Prefrontal cortex: the present and the future," in *Principles of Frontal Lobe Function*, eds D. Stuss and R. Knight (New York, NY: Oxford University Press), 573–597.

Kovács, Á. M., Téglás, E., and Endress, A. D. (2010). The social sense: susceptibility to others' beliefs in human infants and adults. *Science* 330, 1830–1834. doi: 10.1126/science.1190792

Kuhlmeier, V., Wynn, K., and Bloom, P. (2003). Attribution of dispositional states by 12-month-olds. *Psychol. Sci.* 14, 402–408. doi: 10.1111/1467-9280.01454

Lee, Y., Yun, J. E., Kim, E. Y., and Song, H. (2015). The development of infants' sensitivity to behavioral intentions when inferring others' social preferences. *PLoS ONE* 10:e0135588. doi: 10.1371/journal.pone.0135588

Lepsien, J., and Nobre, A. C. (2006). Cognitive control of attention in the human brain: insights from orienting attention to mental representations. *Brain Res.* 1105, 20–31. doi: 10.1016/j.brainres.2006.03.033

Leslie, A. M., German, T. P., and Polizzi, P. (2005). Belief-desire reasoning as a process of selection. *Cogn. Psychol.* 50, 45–85. doi: 10.1016/j.cogpsych.2004.06.002

Leslie, A. M., and Polizzi, P. (1998). Inhibitory processing in the false belief task: two conjectures. *Dev. Sci.* 1, 247–253. doi: 10.1111/1467-7687.00038

Low, J., and Perner, J. (2012). Implicit and explicit theory of mind: state of the art. *Br. J. Dev. Psychol.* 30, 1–13. doi: 10.1111/j.2044-835X.2011.02074.x

Luo, Y. (2011). Do 10-month-old infants understand false beliefs? *Cognition* 121, 289–298. doi: 10.1016/j.cognition.2011.07.011

Marlowe, F. W. (2005). Hunter-gatherers and human evolution. *Evol. Anthropol.* 14, 54–67. doi: 10.1002/evan.20046

Marsh, H., Stavropoulos, J., Nienhuis, T., and Legerstee, M. (2010). Six- and nine-month-old infants discriminate between goals despite similar action patterns. *Infancy* 15, 94–106. doi: 10.1111/j.1532-7078.2009.00002.x

Meristo, M., and Surian, L. (2013). Do infants detect indirect reciprocity? *Cognition* 129, 102–113. doi: 10.1016/j.cognition.2013.06.006

Meristo, M., and Surian, L. (2014). Infants distinguish antisocial actions directed towards fair and unfair agents. *PLoS ONE* 5:e110553. doi: 10.1371/journal.pone.0110553

Miller, E. K., and Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annu. Rev. Neurosci.* 24, 167–202. doi: 10.1146/annurev.neuro.24.1.167

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., and Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: a latent variable analysis. *Cogn. Psychol.* 41, 49–100. doi: 10.1006/cogp.1999.0734

Moran, J. D., and O'Brien, G. (1983). The development of intention-based moral judgments in three– and four-year-old children. *J. Genet. Psychol.* 143, 175–179. doi: 10.1080/00221325.1983.10533549

Moriguchi, Y. (2014). The early development of executive function and its relation to social interaction: a brief review. *Front. Psychol.* 5:388. doi: 10.3389/fpsyg.2014.00388

Moriguchi, Y., and Hiraki, K. (2009). Neural origin of cognitive shifting in young children. *Proc. Natl. Acad. Sci. U.S.A.* 106, 6017–6021. doi: 10.1073/pnas.0809747106

Moriguchi, Y., and Hiraki, K. (2011). Longitudinal development of prefrontal function during early childhood. *Dev. Cogn. Neurosci.* 1, 153–162. doi: 10.1016/j.dcn.2010.12.004

Nelson, S. A. (1980). Factors influencing young children's use of motives and outcomes as moral criteria. *Child Dev.* 51, 823–829. doi: 10.2307/1129470

Nobes, G., Panagiotaki, G., and Pawson, C. (2009). The influence of negligence, intention, and outcome on children's moral judgments. *J. Exp. Child Psychol.* 104, 382–397. doi: 10.1016/j.jecp.2009.08.001

Perner, J., and Roessler, J. (2012). From infants' to children's appreciation of belief. *Trends Cogn. Sci.* 16, 519–525. doi: 10.1016/j.tics.2012.08.004

Piaget, J. (1932/1965). *The Moral Judgment of the Child*. London: Kegan Paul.

Sheskin, M., Chevallier, C., Lambert, S., and Baumard, S. (2014). Life-history theory explains childhood moral development. *Trends Cogn. Sci.* 18, 613–615. doi: 10.1016/j.tics.2014.08.004

Siegal, M. (1982). *Fairness in Children. A Social-Cognitive Approach to the Study of Moral Development*. New York, NY: Academic Press.

Southgate, V., Senju, A., and Csibra, G. (2007). Action anticipation through attribution of false belief by 2-year-olds. *Psychol. Sci.* 18, 587–592. doi: 10.1111/j.1467-9280.2007.01944.x

Surian, L., Caldi, S., and Sperber, D. (2007). Attribution of beliefs by 13-month-old infants. *Psychol. Sci.* 18, 580–586. doi: 10.1111/j.1467-9280.2007.01943.x

Surian, L., and Geraci, A. (2012). Where will the triangle look for it? Attributing false beliefs to a geometric shape at 17 months. *Br. J. Dev. Psychol.* 30, 30–44. doi: 10.1111/j.2044-835X.2011.02046.x

Thoermer, C., Sodian, B., Vuori, M., Perst, H., and Kristen, S. (2012). Continuity from implicit to an explicit understanding of false belief from infancy to preschool age. *Br. J. Dev. Psychol.* 30, 172–187. doi: 10.1111/j.2044-835X.2011.02067.x

Tomasello, M., Carpenter, M., Call, J., Behne, T., and Moll, H. (2005). Understanding and sharing intentions: the origins of cultural cognition. *Behav. Brain Sci.* 28, 675–735. doi: 10.1017/S0140525X05000129

Wainryb, C., Brehl, B., and Matwin, S. (2005). Being hurt and hurting others: children's narrative accounts and moral judgments of their own interpersonal conflicts. *Monogr. Soc. Res. Child Dev.* 70, 1–114.

Wellman, H. M. (2014). *Making Minds: How Theory of Mind Develops*. New York, NY: Oxford University Press.

Wellman, H. M., Cross, D., and Watson, J. (2001). Meta-analysis of theory-of-mind development: the truth about false belief. *Child Dev.* 72, 655–684. doi: 10.1111/1467-8624.00304

Wimmer, H., and Perner, J. (1983). Beliefs about beliefs: representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition* 13, 103–128. doi: 10.1016/0010-0277(83)90004-5

Woodward, A. L. (1998). Infants selectively encode the goal object of an actor's reach. *Cognition* 69, 1–34. doi: 10.1016/S0010-0277(98)00058-4

Wynn, K. (2008). "Some innate foundations of social and moral cognition," in *The Innate Mind: Foundations and the Future*, Vol. 3, eds P. Carruthers, S. L. Laurence, and S. Stich (Oxford: Oxford University Press), 330–347.

Yuill, N. (1984). Young children's coordination of motive and outcome in judgments of satisfaction and morality. *Br. J. Dev. Psychol.* 2, 73–81. doi: 10.1111/j.2044-835X.1984.tb00536.x

Yuill, N., and Perner, J. (1988). Intentionality and knowledge in children's judgments of actors responsibility and recipients emotional reaction. *Dev. Psychol.* 24, 358–365. doi: 10.1037/0012-1649.24.3.358

Zelazo, P. D., Helwig, C. C., and Lau, A. (1996). Intention, act, and outcome in behavioral prediction and moral judgment. *Child Dev.* 67, 2478–2492. doi: 10.2307/1131635