

Decoding the Intent-to-Outcome Developmental Shift in Moral Judgment:

Unraveling the Puzzle from Infancy to Preschool Age

Marine Buon¹ and Francesco Margoni^{2,3}

¹ Laboratoire EPSYLON, Paul Valéry University, France

² Department of Social Studies, University of Stavanger, Norway

³ Department of Psychology, University of Oslo, Norway

Author Note

Marine Buon ORCID ID 0000-0002-2091-1587

Francesco Margoni ORCID ID 0000-0002-8109-228X

We have no known conflict of interest to disclose.

Contact: Marine Buon, Laboratoire EPSYLON, EA 4656, Paul Valéry University, Montpellier, France. Email: marine.buon@univ-montp3.fr

Acknowledgements: We thank Luca Surian and Renée Baillargeon for their helpful comments and suggestions.

Abstract

For decades, researchers in moral judgment development have described and tried to explain the so-called ‘outcome-to-intent shift’ that occurs by late preschool age: preschoolers predominantly base their moral judgment on the outcomes of an agent’s action and it is only at the age of 5 or later that most children start to generate intent-based moral judgment. However, recent research in the domain of early moral competencies has reported intriguing findings: in their socio-moral evaluations, infants are sensitive to agents’ intentions and are even able to disregard information about the consequences of agents’ actions. What are the processes underlying this surprising developmental pattern? We first aim to critically review recent attempts to explain this developmental pattern that focus on the factors that may impact infants’ and children’s ability to represent information about *intentions*. We argue that these propositions explain a part of the findings, but they underestimate the importance of *outcomes* processing. Thus, we analyze the factors that may influence the way children represent and react to information about outcomes, and conclude that a combination of developmental, experimental, cognitive and experiential factors is likely to play a key role in preschoolers’ over-consideration of outcomes in moral judgment (which we call ‘*outcome bias*’). In addition to shedding new light on the understanding of the discrepancy between infants’ and children’s moral competencies, the implications of this proposition for future developmental psychology research in the field of moral cognition are discussed.

Keywords: morality, socio-moral cognition, child development, infancy, dual-process model, harmful interactions, theory of mind

1. Introduction and General Context

When we morally judge someone for a bad outcome, two important factors are considered: whether they intended this outcome and whether they were causally responsible for what happened (Abelard, 1971). Although both intentions and the caused outcomes carry weight in adults' moral judgment, evidence demonstrates that agents' intentions are largely prioritized in adults' moral judgments (Cushman, 2008; Monroe & Malle, 2017; even if some cross-cultural variations have been reported, e.g., Barrett et al., 2016; McNamara et al., 2019).

For decades, researchers in moral development have described (and tried to explain) the so-called 'outcome-to-intent shift' that occurs by late preschool age: preschoolers predominantly base their moral judgments on the outcomes caused by an agent, and it is only at the age of 5 or later that most children start to generate intent-based moral judgments (for classical research, see: e.g., Baird & Astington, 2004; Chandler et al., 1973; Costanzo et al., 1973; Farnill, 1974; Hebble, 1971; Helwig et al., 1995; Imamoglu, 1975; Karniol, 1978; King, 1971; Moran & O'brien, 1983; Nelson, 1980; Piaget, 1932; for more recent research, see: Amir et al., 2021; Cushman et al., 2013; Gvozdic et al., 2016; Killen et al., 2011; Li & Tomasello, 2018; Li et al., 2020; Margoni & Surian, 2017, 2020; Nobes et al., 2009, 2016, 2017; Proft & Rakoczy, 2018; Van de Vondervoort & Hamlin, 2018). For example, in a 'classical' moral judgment task, when presented with vignettes showing cases of *accidental harm* (neutral intention, negative outcome; e.g., "John is running when he trips on a rock and accidentally pushes somebody over.") and *attempted but failed harm* (negative intention, neutral outcome; e.g., "Mark attempts to push somebody over when he trips on a rock and misses.") and asked to judge the protagonists (e.g., "Is John/Mark a bad, naughty boy?"; in Cushman et al., 2013), younger children mainly rely on the presence or absence of negative outcomes and say that John, the accidental transgressor, is a bad boy while they are more lenient toward Mark, the character

who failed to cause harm. However, recent research in the domain of early moral competencies has reported interesting and quite surprising findings: preverbal infants are sensitive to agents' intentions in their socio-moral evaluations and are even able to disregard information about the consequences of agents' actions (e.g., Geraci et al., 2022; Hamlin, 2013a; Hamlin et al., 2013; Kanakogi et al., 2017; Lee et al., 2015; Strid & Meristo, 2020; Woo & Spelke, 2023; Woo et al., 2017; for a review, see Woo et al., 2023). What are the processes underlying this U-shaped developmental trajectory?

At first glance, infants' sensitivity to agents' socio-moral intentions fits quite well with the literature on infants' early psychological and moral competencies and the claim that infants come into the world equipped with a quite sophisticated "innate moral core" (Hamlin, 2013b; Thomsen & Carey, 2013; Woo et al., 2022; see also Thomas, 2024). First, a number of studies have shown that infants are able to master the basic "mentalistic" tools enabling them to attribute mental states to agents, such as intentions, motives, preferences and desires (e.g., Behne et al., 2005; Brandone & Wellman, 2009; Gergely et al., 1995; Woodward, 1998; for a review, see Baillargeon et al., 2016). Second, infants have been shown to be sensitive to the violation of several key moral principles (Ting et al., 2020), such as the principle of *fairness* (e.g., Lucca et al., 2018; Meristo & Surian, 2013; Sloane et al., 2012), the principle of *harm* (e.g., Buon et al., 2014; Ting & Baillargeon, 2021) or *hinder avoidance* (e.g., Hamlin et al., 2007), the principle of *ingroup support & loyalty* (e.g., Bian et al., 2018; Jin & Baillargeon, 2017; Pun et al., 2021), and the principle of *authority*, which specifies moral obligations for both leaders and followers (e.g., Margoni et al., 2018; Margoni & Thomsen, 2023; Stavans & Baillargeon, 2019; Thomas et al., 2018). Last, the existence of intent-based socio-moral evaluations in infancy is also consistent with the intuitionist view about adults' moral competencies: adults' moral judgments or decision-making mostly rely on moral intuitions,

that is the intuitive gut feeling that arises when we perceive an (im)moral action (Haidt, 2001; see also Stratton-Lake, 2014). According to this position, adults' moral intuitions rely on automatic processes that are either unlearned or learned very quickly through domain-specific learning mechanisms (Margolis & Laurence, 2013) and may thus exist very early in development (e.g., Mikhail, 2007).

Children's protracted shift from outcome-based to intent-based moral judgment was initially at the core of Piaget's stage theory of moral development (Piaget, 1932). According to this model, young children's initial stage of morality (the "heteronomous" stage) is based on an egocentric stance and an inability to adopt the perspective of others. As a result, preschoolers and children up to the age of 8 or 9 mainly focus on the outcomes of moral transgressions and do not consider others' subjective internal states (e.g., intentions) in their moral judgments. As they grow older, and with the progress from concrete to abstract thinking, egocentric thought to allocentric thought, children start to use agents' mental states as the main criteria for generating moral judgments. Even though Piaget was a pioneer in the study of moral development, several studies have reconsidered these initial results from both methodological and theoretical perspectives (e.g., Baird & Astington, 2004; Buon et al., 2013; Chandler et al., 1973; Cushman et al., 2013; Imamoglu, 1975; Karniol, 1978; Nelson, 1980; Margoni & Surian, 2017, 2020; Nobes et al., 2016, 2017; Proft & Rakoczy, 2019).

First, various studies, employing different methodologies, have refined the understanding about the age and context in which children become able to incorporate information about intentions into their moral judgment. It has been found that preschoolers are far from being unable to incorporate information about intentions into their moral judgment. When agents' intentions are presented in a salient and explicit way, and/or when intentions and consequences *do not* co-vary, 5- or even 3-year-old preschoolers may be

capable of judging intentional transgressions (negative intention, negative outcome) more severely and more punishable than accidental ones (neutral intention, negative outcome) (Baird & Astington, 2004; Chernyak & Sobel, 2016; Nelson-Le Gall, 1985; Nelson, 1980; Nobes et al., 2009; Proft & Rakoczy, 2019; Van de Vondervoort & Hamlin, 2018; see also Armsby, 1971; Farnill, 1974; Yuill & Perner, 1988). However, in contrast to school-age children, (younger) preschoolers still tend to struggle to prioritize information about intentions over information about outcomes, in particular when the valence (negative or positive/neutral) of the two pieces of information co-varies. For instance, when asked to judge *attempted but failed transgressions* (negative intention, neutral outcome) and *accidental transgressions* (neutral intention, negative outcome), 3- and 4-year-olds tend to judge the former less severely than the latter and thus fail to generate intent-based moral judgments (however, for evidence that 3-year-olds can already judge attempted transgressions more severely and punishable than accidental transgressions, see Margoni & Surian, 2020; see also Vaish et al., 2010, for a similar pattern when children's prosocial behaviors toward the agents are measured).

From a theoretical perspective, Piaget's stage theory of moral development has been challenged by dual-process models of moral judgment and decision making which have predominantly guided investigations on adult moral competencies (Buon et al., 2016; Cushman, 2008, 2013; Cushman et al., 2007; Greene et al., 2004; Rand, 2016). According to Greene's model (Greene et al., 2004), adults' moral judgments rely on a complex interaction between fast, automatic processes and more reflective, deliberative and controlled processes, each contributing variably depending on the individual and circumstances. More generally, this type of model claims that adult moral judgment results from distinct evaluative systems

that can act in concert, in competition or in conflict¹, each relying on specific cognitive processes (Cushman, 2013; Cushman et al., 2007). One specific prediction of dual-process models of moral judgment is that evaluative processes at work in early development remain intact in adulthood, though more or less inhibited depending on the conditions, which contradicts Piaget's stage theory of moral judgment development that posits a full replacement of outcome-based judgment by intent-based judgment.

Testing this prediction of 'continuity', Buon and collaborators have explored adults' reliance on causation and intention in their moral judgment about harmful actions while performing a concurrent verbal shadowing task (i.e., high cognitive load that limits participants' use of verbal and executive resources to perform the moral judgment task) (Buon et al., 2013). Under high cognitive load, whereas participants were equally able to decode information about agents' intentions and causal roles, their judgments mostly relied on information about whether the agent caused harmful outcomes, and thus they were unable to integrate information about intentions into their moral judgment as they would have normally done. These results suggest that children's outcome-based moral judgment can "reappear" in adulthood under specific conditions. Moreover, they suggest that the integration of information about causal responsibility into moral judgment relies on fast, automatic processes ("cause-based heuristic"), whereas the integration of information about agents' intentions relies on more controlled and reflective processes. Consistent with this

¹ We draw a distinction between 'conflict' and 'competition'. A conflict occurs when the two evaluative processes are simultaneously activated (they are both 'occupied') but they give rise to two different, conflicting outputs (e.g., for accidental harm: causal responsibility for the harmful outcome = bad vs. innocent intentions = not bad). Instead, a competition (for limited cognitive resources) occurs when one evaluative process wins the competition for resources and as a result becomes the only one 'occupied' or activated (e.g., for attempted but failed harm: no causal responsibility for harmful outcome = the evaluative process for causation becomes empty; harmful intention = the evaluative process for intention becomes 'full').

evidence, a number of studies found evidence that older adults (age > 65 years), a population characterized for its limitations in cognitive abilities (declines in executive functioning skills in old age are well-documented; see Salthouse et al., 2003), tend to generate outcome-based moral judgments (Margoni, Geipel et al., 2018, 2019, 2023; Moran et al., 2012; but see also Margoni, Cho et al., 2023; Minton et al., 2024) and to make outcome-based decisions when asked to accept or reject either intentional or unintentional low offers in a modified version of the Ultimatum Game (Cho et al., 2020; Margoni et al., 2021).

The ETIC model (E as *emotional arousal*, T as *theory of mind*, IC as *inhibitory control*; Buon et al., 2016) is an attempt to precisely characterize which computations could underlie each evaluative system and when each system would develop. According to this model, whereas the moral evaluation of harmful outcomes caused by the agent would mostly rely on *emotional arousal* (i.e., the automatic affective response triggered by the observation or representation of someone else's distress), the moral evaluation of agent's intentions would rely on *theory of mind skills* (i.e., the ability to represent, attribute and use agents' mental states to explain their behaviors; Premack & Woodruff, 1978). When the two evaluative systems produce incongruent outputs (in the case of accidental harm: the agent has caused harm = s/he is bad; but s/he did not want to = s/he is not bad), *inhibitory control skills* (i.e., cognitive processes that enable the regulation and control of other cognitive processes) are required to inhibit the negative evaluation arising from the representation of the harmful outcomes caused to the victim. Thus, according to the ETIC model, the outcome-to-intent shift would be tied to the developmental time course of the above-mentioned processes: indeed, whereas emotional arousal – responsible for the evaluation of the harmful outcomes – would be 'online' early during development, flexible theory of mind and executive functioning skills

– responsible for a consideration and prioritization of information about intentions – would emerge later, at late preschool age or the beginning of school age.

According to another dual-process model (Cushman, 2008, 2013), distinct psychological processes also underlie evaluations about agents' intentions and actions' outcomes, but each of these evaluations relates differently to different types of judgment, wrongness/permissibility vs. punishment/blame (for the distinction between judgment types, see Malle, 2021, 2024). Whereas judgments of moral wrongness/permissibility principally rely on information about mental states, judgments of punishment/blame principally rely on information about causal responsibility. From a developmental perspective, it is proposed that young children's evaluative systems take as the main input information about the agents' harmful causal role, and yield judgments of wrongness, naughtiness and punishment that initially do not seem differentiated (Cushman et al., 2013). As children get older, and during the outcome-to-intent shift (5 to 6 years), they acquire a new concept of wrongness based on the emerging ability to evaluate the agents' mental states, which will then, in turn, constrain punishment judgment (to rely, at least partially, on intentions). Even though empirical findings that test this theoretical proposition remain sparse, some experiments conducted in adult (Martin, Buon, et al., 2021) and child populations (Cushman et al., 2013; Martin, Leddy, et al., 2021) have provided evidence at least consistent with this proposition. In particular, it has been reported that US children's intent-based judgments emerge first when children have to decide whether an agent is naughty, which then constrains judgments of deserved punishment (Cushman et al., 2013). Importantly, however, this last crucial bit of evidence in favor of the 'constrain view' has not been replicated in a sample of Italian children and with tests more powered than those in the original research (see Margoni & Surian, 2017).

2. Why Do Infants Ostensibly Outperform Preschoolers?

Taken separately, findings from infant and child experiments may find some theoretical foundations in the literature on moral judgment. However, if we attempt to combine them into a more global and coherent picture of *moral judgment development*, the task becomes more complicated. Indeed, preschoolers appear to inexplicably ‘lose’ an ability initially mastered at an early age, only to recover it a few years later, which is quite surprising (Margoni & Surian, 2020). When observed in other domains, Piaget used to refer to this kind of pattern as ‘vertical *décalage*’ (Piaget, 1941). In addition, the dual-process models described above would predict that very young children, i.e., infants and toddlers, would initially be sensitive to information about causation / outcomes and not to information about intentions, as it was reported by recent studies conducted with infants and toddlers (see e.g., Hamlin, 2013a; Hamlin et al., 2013; Woo & Spelke, 2023).

Researchers have recently tried to address this developmental puzzle, and most of them (Hamlin, 2013a; Hilton & Kuhlmeier, 2019; Margoni & Surian, 2016, 2020; Proft & Rakoczy, 2019; Van de Vondervoort & Hamlin, 2018) tend to explain the discrepancy between infants’ socio-moral cognition and children’s moral judgment by appealing to the differences that exist between the tasks used to probe moral understanding in these two populations. In infant studies, the tasks are obviously mostly non-verbal: the stimuli consist of short movies or puppet shows, and researchers measure infants’ *non-verbal reactions*, either *spontaneous* (e.g., reactions of surprise revealing infants’ expectations about the presented events; for a methodological review, see Margoni et al., 2024) or minimally *elicited*² (manual reaching behavior where the infant preferentially approaches one of the two protagonists, interpreted

² We use the term ‘minimally elicited’ because infants’ grasping behaviors are not entirely spontaneous; they are prompted after a simple request from the experimenter.

as a preference for it; for a methodological review, see Choi & Luo, 2023). In contrast, in more ‘classical’ moral judgment tasks used with children, most of the time participants judge interactions based on vignettes and verbal descriptions, and queries are also mostly verbal, with children being asked to generate explicit moral judgments about agents or actions, which arguably requires participants to use more verbal skills and inferential abilities. As a result, the type of tasks traditionally used with preschoolers and older children would be more demanding in executive skills, theory of mind abilities, inferential abilities and response-generation skills than the tasks used with infants or toddlers. These demands might hinder their ability to properly represent and integrate agents’ intentions into their moral judgments. In short, the tasks used with children may not allow them to *express* their true moral competences (Margoni & Surian, 2016, 2020).

In the following sections, we disentangle and critically assess the various factors that have been proposed to explain why preschoolers might seem to underperform compared to infants and toddlers: the task eliciting non-verbal and spontaneous / minimally elicited responses vs. verbal and elicited responses (section 2.1); the way in which the intentionality (or lack thereof) of the actions is presented to the participants (section 2.2); and the salience of information about intentions and outcomes respectively (section 2.3). Next, we question the relevance of including another potential and, so far, neglected candidate in the present debate. We wonder whether the way actions’ outcomes are represented in the stimuli used with infants and children, and the way that characters’ emotional reactions are processed by these different populations, may substantially affect the propensity to generate intent-based moral evaluations.

2.1. *Verbal vs. Non-verbal Spontaneous or Minimally Elicited Response Tasks*

According to some researchers, the type of task used to probe infants' and children's moral competencies is a key factor determining the likelihood that the participants will (or will not) be able to *express* their sensitivity to agents' moral intentions (Margoni & Surian, 2016, 2020; Van de Vondervoort & Hamlin, 2018). According to the account put forward by Margoni and Surian (2016, 2020), compared to the non-verbal spontaneous or minimally elicited response tasks used with infants and toddlers, the classical verbal moral judgment tasks used with preschoolers are cognitively too demanding for their age, preventing them from incorporating information about agents' intentions into their moral judgments and expressing their true moral competence. Following this hypothesis, Margoni and Surian (2020) conducted a study in which children were presented with stimuli which varied as to the intentions of the agent and the consequences of their action (attempted but failed harm vs. accidental harm). Critically, children did not have to provide any verbal judgments about the agents but had to enact their responses by placing agents in the most appropriate box (i.e., "the box where the good agents go" vs. "the box where the bad agents go"). In a first experiment, children were previously *trained* to use this type of response procedure (reduced response generation demands condition), whereas in a second experiment, they were not (*not*-reduced response generation demands condition). Three-year-old children were at chance level in the *not*-reduced response generation demands condition, whereas they were able to generate intent-based judgments when the response generation demands were reduced. These results are consistent with other studies that have been shown that young preschoolers are able to integrate information about intentions in their moral judgments when experimenters use probes that are more implicit than the ones used in the classical moral judgment tasks, for

example studies testing prosocial behavior (Vaish et al., 2010) or spontaneous protest (Chernyak & Sobel, 2016).

This account of the apparent discontinuity between infants' and preschoolers' moral evaluations closely aligns with another account put forward to explain why infants show the ability to represent false beliefs whereas younger preschoolers do not (see Baillargeon et al., 2010; Scott & Baillargeon, 2017; cf. also Perner & Roessler, 2012). Indeed, whereas preschoolers under the age of 4 or 5 years fail at classical elicited-prediction tasks when they have to answer a test question that requires them to predict the behavior of an agent acting upon a false belief (Wimmer & Perner, 1983), infants pass non-verbal implicit false-belief tasks (e.g., Onishi & Baillargeon, 2005). As for toddlers, they pass both non-verbal and verbal implicit tasks and even explicit elicited-prediction tasks of the kind used with older children when, however and importantly, the processing demands are reduced (Setoh et al., 2016; see also Buttelmann et al., 2009; Southgate et al., 2010). According to Baillargeon, Scott, and colleagues, both infants, toddlers and preschoolers do have the capacity to attribute false beliefs to others, but preschoolers appear to 'lost' it because they are tested with high-demand tasks which hinder its expression (Scott & Baillargeon, 2017). In particular, it is posited that when children in standard elicited-prediction false-belief tasks are asked, "Where will Sally look for her ball?", they initially think about where the ball currently is (true belief) and to pass the task, they must *inhibit* this initial, prepotent response and select the response that takes into account Sally's false belief about the true location of the ball (see also Carlson & Moses, 2011). Thus, because of their limited executive functioning abilities, younger preschoolers fail when tested with classical false belief tasks. In a similar way and for similar reasons, this population may fail at exculpating accidental harm when tested with classical moral judgment tasks (Margoni & Surian, 2016, 2020). When asked about whether the story's

protagonist is a bad, naughty boy and/or deserved to be punished, they may activate the representation of the negative outcomes brought about by the protagonist's actions and fail to inhibit this initial, prepotent outcome-based response. Thus, according to the account put forward by Margoni and Surian (2016, 2020), preschoolers won't 'lose' any competence that can be attributed to infants or toddlers, and all the shifts in moral judgment development can be accounted for by the differences between the tasks being used.

Whereas the different results evoked above (Chernyak & Sobel, 2016; Margoni & Surian, 2020; Vaish et al., 2010) can be taken as an argument that the type of task used to probe children's moral sensitivity is most likely critical for younger preschoolers³ to generate intent-based moral judgments, Martin, Leddy, et al. (2021) explain these effects of the task by highlighting the role of the *type of judgment* being asked. For Martin and collaborators, whereas non-verbal response tasks (typically used with infants) would be perceived as a partner choice by participants, verbal responses typically asked from children in classical moral judgment tasks may be viewed as endorsing the punishment of the evaluated agent. These two types of judgments (partner choice and punishment) supposedly rely differently on information about intentions and outcomes (e.g., Cushman, 2008; Cushman et al., 2009; Hackel et al., 2015; Raihani & Barclay, 2016). Whereas an assessment of the true willingness to act prosocially is what matters the most for partner choice, outcomes seem to have a greater weight in punishment decisions. In support of this hypothesis, it has been reported that children were able to generate intent-based moral judgments earlier (at 4-5 years) when

³ It should be noted that in Margoni and Surian (2020), children from 4 years of age were able to generate intent-based judgments, regardless of whether the response generation demands were reduced or not. As discussed below, this suggests that most likely other aspects of the changes made in this task compared to more traditional moral judgment tasks may too explain preschoolers' ability to show sensitivity to information about intentions.

explicitly prompted to decide verbally whether or not they would like a given agent as a playmate (partner choice), compared to when they were asked to decide whether a given agent deserved punishment (Martin, Leddy, et al., 2021). In a similar direction with respect to probe wording, the work of Nobes and collaborators has shown that asking to evaluate the story's protagonist ("Is Kevin good, bad, or just okay?") substantially increases the likelihood that preschoolers will generate an intent-based judgment compared to asking about the protagonist's action ("Is it okay for Kevin to do that?") (Nobes et al., 2016, 2017).

Thus, even if we do not negate the critical role played by the type of response requested from children (e.g., verbal versus non-verbal), the precise origin of the observed effects needs to be explored further and more deeply. Importantly, in the studies suggesting a key role for the task type we referred to above, several aspects of the procedure, the stimuli and the probes were changed at the same time to help children use intentions in their judgments. All these changes made the task administered to preschoolers more similar to the tasks employed with infants and toddlers and more different from the classical moral judgment tasks used with children. For instance, Vaish et al. (2010) also used real life interactions to depict moral events to children, as sometimes is done in infant research (e.g., Dunfield & Kuhlmeier, 2010), and Margoni and Surian (2020) also used movies (vs. vignettes) staging puppets interacting with each other on a puppet show, again similarly to what is done with infants (e.g., Surian & Margoni, 2020). In these studies, the stimuli used made the agents' intentions particularly salient and explicit thanks to perceptual and emotional cues (Hilton & Kuhlmeier, 2019). As we see in the next section, these cues may also improve children's ability to generate intent-based responses. It is thus difficult to disentangle the precise reason or set of reasons why children in these studies are better at integrating information about intention than in the majority of the other studies.

2.2. *Variability in the Display and Types of Intentional Actions*

Another methodological difference that has been pointed out to explain the contrasting abilities of infants and preschoolers in integrating information about agents' intentions into their evaluations concerns the way in which intentions are displayed in the experimental stimuli. According to Hilton and Kuhlmeier (2019), in most infant experiments and mainly because the events shown to participants cannot be accompanied by complex verbal explanations, there is a close correspondence between *external cues* and the actors' intentions (for instance, an accidental harm can be accompanied by a "Whoops!" and cues of surprise or regret by the agent producing the action). Thus, intentional cues can often be interpreted simply on the basis of perceptual cues. By contrast, in most child and adult experiments, agents' intentions have to be *inferred* from the agents' actions or the verbal description of the context, and remain hidden mental states (e.g., Young et al., 2011). On one side, this is closer to real life situations where is often the case that the mental states of others have to be inferred from their actions, and no close correspondence between external cues and internal mental states is present. However, on the other side, the cognitive cost required to grasp the agents' malevolent or innocent intentions is higher in child experiments than in infant experiments.

In a similar vein, Proft and Rakoczy (2019) focus on the way accidents are depicted in child experiments, as accidents are the situations with which preschoolers struggle the most, and where the outcome-to-intent shift is most evident (e.g., Cushman et al., 2013; Margoni & Surian, 2017; Nobes et al., 2009). They highlight the distinction between accidents based on unintentional actions (e.g., when a child stumbles and consequently accidentally breaks another child's preferred toy) and false-belief-based accidents (i.e., intentional actions based on a false belief; e.g., when a child throws a paper bag containing another child's favorite

snack into the trash while falsely thinking that the bag contained a piece of trash). What critically distinguishes these two cases is the chain of reasoning that is required to capture and assess the moral status of the agents' intentions. Whereas in the latter case, one needs to engage in two-step inferential reasoning (false belief \rightarrow unintended \rightarrow not bad), in the former case one only needs to engage in one-step inferential reasoning (unintended \rightarrow not bad). Thus, according to Proft and Rakoczy (2019), children's apparent difficulty in generating intent-based moral judgment might (at least in part) rest on the inferential complexity of the stimuli used. Consistent with this hypothesis, some studies have tested children's sensitivity to agents' intentions using stimuli that could have been used with infants, in which (the lack of a negative) intention can be captured solely based on the unintentional nature of the action, and not on false-belief inferences (Margoni & Surian, 2020; Vaish et al., 2010; Van de Vondervoort & Hamlin, 2018). They have shown that under these conditions (but other conditions as well) most preschoolers are sensitive to intentions (or lack thereof) in their moral judgments. Moreover, Proft and Rakoczy (2019)'s own elegant study on 5-year-olds has shown that children were able to integrate agents' intentions into their moral judgments reliably more when the stimuli depicted accidents based on unintentional actions than when the stimuli depicted accidents caused by false beliefs. And when preschoolers were primed for the intentional structure of the false-belief accidents – a procedure that helped them perform the first step (false belief \rightarrow unintended) of the two-step inferential reasoning (false belief \rightarrow unintended \rightarrow not bad) – preschoolers were better able to integrate information about intentions into their moral judgments. Thus, according to the authors, the inferential complexity of the stimuli used by past child research is one promising candidate for explaining preschoolers' difficulties in generating intent-based moral judgment. Unlike older children, probably due to fragile working memory, limited executive functioning skills and/or a

restricted theory of mind, younger preschoolers are not able to generate intent-based judgments if tested with complex stimuli.

Despite these interesting findings, two points need to be mentioned. First, infants' and toddlers' sensitivity to agents' moral intentions has been reported even when using stimuli in which intentions can be captured solely based on false belief inferences (Hamlin, 2013a; Woo et al., 2017; Woo & Spelke, 2023). Second, child studies in which agents' intentions can be captured without the need to engage in a complex inferential reasoning chain (because the agents' mental states can be discerned simply based on the intentional or unintentional nature of the agents' actions) or in which intentions were explicitly stated (e.g., 'Jessica wants to help take care of the garden' vs. 'Jessica wants the sandcastle her brother has built to collapse' – in Baird and Astington, 2004) did not systematically find that younger preschoolers were sensitive to agents' intentions (e.g., Baird & Astington, 2004; Nelson, 1980; Nelson-Le Gall, 1985; for the same result on punishment judgment, see also Martin, Leddy, et al., 2021). All these studies have also consistently revealed a substantial developmental increase in the ability to focus on agents' intentions with age (for a developmental increase in the use of intentions, see also Nobes et al., 2009). Thus, the way in which agents' mental states are differently displayed in infant and child studies may explain a part of the discrepancy between infants' and children's performances, but it is unlikely to explain the whole picture.

2.3. Salience of Intentions over Outcomes

According to Hilton and Kuhlmeier (2019), verbal (vs. non-verbal) presentation of information about intentions in the stimuli used with children also substantially alters their ability to construct an accurate representation of intentions, by decreasing the salience of intentions over outcomes (because mental states are not visible whereas outcomes are explicit). However, the difficulty in integrating information about intention is not restricted to

verbal stimuli, but is also reported in studies which used audio-video stimuli (see e.g., Gvozdic et al., 2016). Still, several experiments have showed that when the salience of intentions (over outcomes) is artificially increased, preschoolers' capacity to integrate information about intentions in their moral judgment also increases: This is true when the order of presentation of intentions and outcomes is modified and intentions are presented last (see Study 2 in Nelson, 1980), when – right before the moral judgment probe – children are systematically asked to recall what the intentions of the protagonist were (see Margoni & Surian, 2020; see also Margoni, Guglielmetti, et al., 2019), and when information about agents' intentions is accompanied with emotional cues (e.g., Vaish et al., 2010), or perceptual cues like “Oops!” (e.g., Chernyak & Sobel, 2016) or thought bubbles visually representing agents' intentions (e.g., Nelson-Le Gall, 1985).

However, increasing the saliency of the information about intentions does not completely erase the outcome-to-intent shift. Several of the studies that have used stimuli and procedures that enhance the saliency of intention information and have tested children at different ages still report that younger preschoolers rely both on intentions and outcomes, and report both an increase in the use of intentions and a decrease in the use of outcomes with age (e.g., Nelson, 1980; Nelson-Le Gall, 1985; Nobes et al., 2009, 2016, 2017). In addition, whereas increasing the relative salience of information about intentions vs. outcomes seems to positively impact children's ability to integrate that information into their judgments – according to Hilton and Kuhlmeier (2019), it would help children construct an accurate representation of the agents' intentions⁴ – it remains unclear what is the process (or set of processes) impacted by these manipulations.

⁴ Note however that preschoolers' difficulties in recalling relevant information in classical moral judgment tasks do not seem to be limited to information about intentions but extend

One could first endorse a lower level of explanation. By making information about intentions salient and explicit, we might also attach *emotionally charged cues* to it, on which children might in turn rely to form their judgments, without a genuine consideration of the origin and nature of these positively or negatively connotated pieces of information. To make an agent's positive intention more salient and explicit, one could say for instance, "She *likes* her, she wants to *play* with her". The two italicized words arguably carry a positive connotation for children. An agent's negative intention could instead be described as an agent who "is *angry at her* and wants to *destroy her toy*", a description containing two sets of words with negative connotations. However, if the child relies solely or primarily on statements with positive or negative connotations to generate moral judgments, it might give the impression that the child's judgments are relying mostly on an assessment of the agents' intentions, when this might not actually be the case (especially when the child is asked to evaluate cases of attempted but failed harm, where often no salient emotional cue is attached to the description of the final outcome). Moral judgment cannot be reduced to an assessment of the emotional states of the transgressor, rather it is assessment of whether the transgressor's intentions were to respect or violate moral principles and norms.

A second hypothesis could be that increasing the salience of the information about agents' intentions may "spare" children's need to deploy theory of mind abilities and/or executive resources that are required to "win" the competition against information about outcomes (Buon et al., 2016; Hilton & Kuhlmeier, 2019; Margoni & Surian, 2016, 2020). This explanation would be in line with studies conducted on adults linking mental reasoning skills to moral judgment (e.g., Buon et al., 2013; Moran et al., 2011; Young et al., 2007; Young &

to information about outcomes (e.g., Nelson, 1980). For the same type of evidence in adults under cognitive load, see Buon et al. (2013).

Saxe, 2009a,b) and recent dual-process models of moral cognition described above (e.g., Buon et al., 2016). For instance, in the case of accidental harm, to generate an intent-based moral judgment, one would need well-established theory of mind capacities (Buon et al., 2013; Moran et al., 2011) and/or inhibition skills (Buon et al., 2016; Gvozdic et al., 2016; Margoni & Surian, 2020; see also Rosset & Rottman, 2014) to override the negative emotional arousal or intuitive negative reaction arising from the perception of harmful outcomes. By contrast, in the case of attempted but failed harm, the demands on mental state reasoning (Young & Saxe, 2009a) and/or executive functioning skills are weaker (Martin, Buon, et al., 2021), probably because the information about outcomes – when there are neutral or positive consequences and these are presented to participants – is much less arousing than for accidental harm cases. This hypothesis is consistent with different strands of evidence. First, preschoolers (Cushman et al., 2013; see also Margoni & Surian, 2017, on help scenarios) and adults under cognitive load (Martin, Buon, et al., 2021) find it especially difficult to generate intent-based moral judgment when faced with accidental harm vs. attempted harm. Second, preschoolers' sensitivity to information about intentions appears to increase when the outcomes of the different actions are not explicitly displayed in the stimuli used (Baird & Astington, 2004). Third, in some studies, children's ability to generate intent-based moral judgment has been found to be linked to their theory of mind competencies (Baird & Astington, 2004; Fu et al., 2014; Gönültaş et al., 2021; Killen et al., 2011; for convergent evidence on older adults, see Margoni et al., 2018) and/or their executive functioning skills (Baker et al., 2021; Gvozdic et al., 2016; see also Ochoa et al., 2022).

Whereas this latter hypothesis finds some support in the literature, it still does not appear to fully explain the contrasting performance of infants and preschoolers. As mentioned earlier, when information about intentions is explicit and salient (e.g., Karniol, 1978; Nelson,

1980; Nelson-Le Gall, 1985; Nobes et al., 2009, 2016, 2017; see also Study 2 in Margoni & Surian, 2020), but also when information about outcomes is absent from the stimuli (Baird & Astington, 2004), a developmental outcome-to-intent shift is still perceptible, though less markedly from preschool age to middle school age. Making intentions more salient than outcomes may thus weaken the conflict arising between information about intentions and outcomes without erasing it. Children would still need to override the prepotent response arising from the perception of the outcomes. In the only study that so far reported that 3-year-olds fully exculpate accidental harm and condemn attempted but failed harm (Study 1 in Margoni & Surian, 2020), as discussed previously, it is likely that what determined the results was not only the increase in the salience of intention information, but a broader set of factors and manipulations (that, according to the authors of that study, reduced the overall cognitive demands of the task).

Moreover, in the studies investigating infants' intent-based socio-moral evaluations, infants do not appear to be much *bothered* by information about outcomes. In Hamlin (2013a), 8-month-old infants' evaluative patterns appeared comparable to the adult ones, that is, infants appeared to be sensitive to agents' intentions and *unconcerned* by actions' outcomes: in addition to preferring helpers over hinderers regardless of whether helpers were successful or failed, infants did *not* prefer a successful helper/hinderer over a failed helper/hinderer, thus appearing to completely disregard the information about outcome⁵ (see

⁵ Note however that a re-analysis of Hamlin (2013a)'s results done by us using Bayesian statistics shows only *anecdotal evidence* in favor of the null hypothesis of no preference between a successful helper and a failed helper (10 and 6 8-month-old infants respectively in Exp. 3A), $BF_{01} = 2.08$, and the null hypothesis of no (dis)preference between a successful hinderer and a failed hinderer (7 and 9 8-month-olds infants respectively in Exp. 3B), $BF_{01} = 2.97$. These Bayesian binomial tests were done with JASP, using default priors (JASP Team, 2022) and preclude drawing strong conclusions from these null results.

also Hamlin et al., 2013, for complementary evidence that pre-verbal infants do not prefer accidental helpers over accidental non-helpers). Consistently, the only study that so far tested infants' reactions to accidental vs. attempted outcomes reported that 15-month-olds prefer failed helpers over accidental helpers (Woo & Spelke, 2023; for evidence that 10-month-olds prefer accidental transgressors over intentional ones, and intentional over accidental helpers, see Woo et al., 2017, and Exp. 6 in Kanakogi et al., 2017, respectively)⁶. Last, to complement these findings, violation-of-expectation studies reported that 10-month-olds expect others to prefer distributors who attempt but fail to be fair over distributors who attempt but fail to be unfair (Strid & Meristo, 2020; for the same pattern in infants' own preferences, see Geraci et al., 2022) and 16-month-olds expect others to approach failed helpers over successful hinderers (Lee et al., 2015).

We think that part of the solution to explain the discrepancies between infant and preschooler findings could be to focus both on the way in which *outcomes* are presented in the stimuli used with infants and children, and on the way in which they are processed by these different populations. To date, attempts to explain the discrepancies between the results obtained by testing infants and those obtained by testing preschoolers have mainly focused on agents' intentions, with no or little consideration for actions' outcomes. However, outcomes cannot be considered as trivial and should not be neglected in our understanding of infants' and preschooler's findings. In the next section, we analyze the factors that we

⁶ Note however that, (a) in Woo and Spelke (2023) the accidental helper also had a negative intention (to help the protagonist retrieve the non-preferred toy); (b) no study so far has shown that infants prefer an accidental harm-doer/hinderer to an attempted one; and (c) in Woo et al. (2017), it remains unclear whether infants were preferring the accidental transgressor over the intentional one because they were indeed exculpating the former (as when you ask a 5-year-old to say whether the accidental transgressor is good or bad) or because they were simply approaching the least bad puppet, that is, the one who did not show a bad intention.

believe might play a role in the different reaction to outcomes by children and infants, and how these could impact their propensity to integrate information about intentions and outcomes in their socio-moral evaluations.

3. Could Outcomes Be Another Potential Candidate for Explaining the Intent-to-Outcome Shift from Infancy to Preschool Age?

Instead of focusing on agents' intentions, in this section we question infants' apparent insensitivity and preschoolers' potential over-sensitivity to actions' *outcomes* when faced with socio-moral interactions. In order to explain why infants seemingly outperform preschoolers in socio-moral tasks, we argue that preschoolers do not 'lose' competences that they had in infancy but instead they may exhibit an outcome bias triggered by a combination of methodological, developmental and experiential factors.

3.1. Moral Outcomes

To begin with, let us clarify what we refer to with the term 'moral outcomes'. Transgressions are initially deemed as moral mainly based on their negative consequences for the well-being of (relevant) others (Nucci & Turiel, 1978; see also Killen & Smetana, 2015). Emotional *empathy* – the capacity to perceive and vicariously experience the emotion of another person – or *sympathy/empathic concern* – the ability to feel concerned for another person's emotional state – are thought to be processes commonly involved in the individuals' perception of transgressions as moral (Blair, 1995; Decety & Holvoet, 2021), in the motivation to engage in prosocial behavior (Decety, 2010; Eisenberg, 2000; but cf. Bloom, 2017), and in the motivation to inhibit antisocial actions (Blair, 1995, 2017). Here, whether emotions are activated because individuals intuitively judge and perceive that a moral principle has been violated or, vice versa, the activation of moral thought and judgment follows from an emotional activation is irrelevant to the discussion (though we favor the first possibility) (see

Huebner et al., 2009, for a relevant discussion on the role of emotion in moral judgment). The crucial point is that, for the better or for the worse, the processing of a moral outcome is often associated with the activation of an emotional reaction in the individual morally judging the situation and an understanding of the emotional reaction of the victim (e.g., Blair, 1995; Young et al., 2012). Indeed, emotional empathy and/or *sympathy* (Smith, 1759) is supposed to be directed toward the victim of the transgression, who is expected to express distress following the action she was the target of. However, the availability of cues we can respond emotionally varies depending on the situation and the type of evaluated transgression. Sometimes, moral transgressions (i.e., physically harmful actions such as hitting, pulling hair, etc.) are intrinsically and systematically associated with overt emotional displays (e.g., overt expressions of pain, distress) from the victim, so that it would be difficult to show the transgression to participants without also showing cues of distress in the victim. Sometimes, despite transgressions can be certainly considered immoral (e.g., hindering someone by preventing her from reaching her goals, distributing resources unfairly), the consequences of the action (an agent who does not achieve her goals, e.g., to play with her preferred toy, or an agent who receives fewer resources than others) are not systematically associated with *overt* negative cues from the victim (e.g., distress, anger, sadness), so that it would be possible to show the immoral action without explicitly showing an emotional reaction from the victim (as in the stimuli typically used with infants, see section 3.2). When these cues are not available, adults and children can of course and to some extent infer them and empathize and/or sympathize with the victim through various cognitive processes, among which verbal mediation, social scripts or affective perspective taking (Shamay-Tsoory, 2011; Shamay-tsoory & Aharon-peretz, 2007; Shamay-Tsoory et al., 2009; Vaish, 2016; Vaish et al., 2009; Warneken & Tomasello, 2006). That being

said, we now take a closer look at the stimuli used in infant and child studies, with a focus on the way in which outcomes are displayed.

3.2. *Moral Outcomes in Infant and Child Studies*

In infant studies, the stimuli used tend to be minimal and abstract, made in a way that participants cannot use their familiarity with the agents and the actions presented to them to process the relevant social interactions (e.g., Mascaro & Csibra, 2012; Tatone et al., 2023; Thomsen et al., 2011). The stimuli used in the infant literature we are assessing represent either actions of helping vs. hindering (Hamlin, 2013a; Hamlin et al., 2013; Lee et al., 2015; Woo & Spelke, 2023; Woo et al., 2017) or distributing resources equally vs. unequally (Geraci et al., 2022; Strid & Meristo, 2020), with the only exception of Kanakogi et al. (2017)'s stimuli that were about intervening (either intentionally or accidentally) to stop an act of physical aggression (i.e., collision). As mentioned above, all these actions can be considered to fall within the (broadly defined) moral domain of *harm* since they impact the *patient's welfare*⁷ (Gray et al., 2012; Sousa & Piazza, 2014), though they are not directly associated with overt emotional cues from the patient (i.e., the stimuli do not show the physical, psychological or emotional state impact of the actions on the patients). Consider for example the stimuli used in Strid and Meristo (2020), where recipients affected by unequal distributions do not react negatively (or react at all), or consider the stimuli used in Woo and Spelke (2023), where the outcomes were accompanied by an external sound but the patient itself does not react in any way. By contrast, the actions presented to children are most of the times familiar to them (they reflect scenarios children are likely to encounter in their daily lives, e.g., a child wants to

⁷ The term 'patient' is used here in contrast to the term 'agent', that is, the one who performs the action. It refers to the recipient of the action, whether the outcomes are positive (e.g., the patient has been helped, comforted, etc.) or negative (e.g., the patient has been prevented from achieving its goal, it has been harmed, etc.).

throw the ball to her friend and accidentally hits her in the head) and are also more variable across the experiments. They fall within the domain of property damage or violation (e.g., Baird & Astington, 2004; Killen et al., 2011; Margoni & Surian, 2020; Piaget, 1932; Vaish et al., 2010) or physical harm (e.g., Gvozdic et al., 2016; Nelson, 1980; Nelson-Le Gall, 1985; Nobes et al., 2009; Zelazo et al., 1996)⁸. Also, in some studies the impact of the action on the patient's welfare is explicitly shown (e.g., Gvozdic et al., 2016; Margoni & Surian, 2020; Nelson, 1980; Nelson-Le Gall, 1985; Zelazo et al., 1996) although this is not systematic (e.g., Baird & Astington, 2004; Killen et al., 2011; Vaish et al., 2010).

From this rudimentary description of the way in which outcomes are displayed in the stimuli used with infants (and children), it appears that – at least in the literature we are here considering – infants' arousal potential related to the processing of moral outcomes is probably at a minimal level or at least at a lower level than that of children. Please note that we do not want to argue that infants do not have any emotional reactions to the moral stimuli they are presented with; they probably do (see e.g., Steckler et al., 2018). However, none of the experiments conducted on infants' intent-based socio-moral evaluations that we have reviewed so far have examined actions directly performed on patients, such as when an agent physically harms a victim (as opposed to a victim being ignored by an unequal distributor, for instance). Additionally, none have explicitly displayed the victim's suffering. In Kanakogi et al. (2017), for example, an agent collides with a patient, but the patient does not appear to be

⁸ As pointed out by Buon et al. (2014), a question that remains wide-open in the field of moral psychology is whether the ability to evaluate (un)cooperative, (un)fair and (un)helpful interactions or transgressions stems from a generic capacity to evaluate agents as a function of the valence of their actions (Premack & Premack, 1997) or whether it rests on a collection of domain-specific social evaluation systems (Haidt, 2007). Though this is not a question we address here, if this latter hypothesis is correct, one may wonder whether infants and children are tested on the same social evaluation system(s).

negatively and emotionally impacted by the agent's action, at least explicitly. By contrast, children's affective reaction when faced with the outcomes of the presented stimuli is arguably higher. This is so because the outcomes (for the patient's suffering) are often explicit and clearly negatively valenced. This might be also because children's emerging emotional competences (section 3.3) along with their growing social and emotional experience (section 3.4) with the presented moral actions, result in an increased affective reaction to the impact of the agents' actions on the victim's welfare. Coupled with their limited executive functioning abilities (e.g., Best & Miller, 2010; Caporaso et al., 2021), this might sway moral judgment in favor of an outcome bias.

Importantly, our claim is not even that infants are unable to encode the negative value of the outcomes of the actions they are presented with. They probably do, given that evidence shows that infants interpret actions as directed toward goals (e.g., Gergely et al., 1995; Woodward, 1998; for a review, see Baillargeon et al., 2016) and that the process of assigning a negative value to an ill-intentioned agent relies on the value of the intended outcome. For instance, the negative value attached to an agent A who attempts to hinder a patient B from achieving a goal rests on the negative value inherent in the act or outcome of hindering someone. Thus, both infants and children may be able to encode the negative value of the outcomes of a moral transgression but, based on the arguments we outline below, outcomes may appear more salient (and aversive) for the latter population, resulting in an outcome bias that would be difficult to overcome unless children have developed enough executive resources.

3.3. The Potential Role of the Development of Emotion Understanding and Affective Concern Towards Others in Distress

When we focus on the development of infants' processing of distress in others, it appears that infants are capable quite early to distinguish a wide range of emotional expressions (for review, see Grossman, 2010; Ruba & Pollack, 2020), as well as emotionally resonating with the suffering of others. Several studies have reported contagious crying (i.e., infants cry when hearing another infant crying) in newborns (Dondi et al., 1999; Sagi & Hoffman, 1976; Simner, 1971) and older infants (Geangu et al., 2011). Importantly, such emotional resonance seems to depend on the intensity and duration of the arousing stimuli (Davidov et al., 2021; Hay et al., 1981; Liddle et al., 2015; Roth-Hanania et al., 2011). This has led some researchers to propose that within their first year of life, infants do not systematically respond to others' emotional distress with personal distress, as claimed by more traditional accounts of empathy development (e.g., Hoffman, 2020), but may be capable of true empathic concern toward others in distress. Indeed, some studies have measured young infants' emotional expressions of concern for victims (i.e., facial expressions, vocalizations, and gestures) and reported evidence of empathic concern (Davidov et al., 2021; Liddle et al., 2015; Roth-Hanania et al., 2011) during the first year of life (for review, see Davidov et al., 2013; but see also Nichols et al., 2015). However, most of these studies report that concern for others in distress increases gradually during the second year (see also Hay et al., 1981) and later (e.g., Knafo et al., 2008). This suggests that with age but also experience as well as the increasing development of different top-down processes (e.g., self-other differentiation, emotional regulation, perspective taking), children develop a more sophisticated form of empathic understanding and concern (Decety & Holvoet, 2021; Vaish, 2016).

Relatedly and importantly for the current topic, one may wonder how infants respond to valanced events that are devoid of emotional content, such as those typically presented in studies investigating infants' socio-moral preferences (e.g., Hamlin, 2013). A recent series of experiments has started to tackle this issue by investigating infants' expectations following events that are supposed to produce emotional reactions (for review, see Ruba et al., 2019). In these studies, which used event-emotion matching paradigms, infants' attentional reactions are measured when someone expresses an emotional reaction that is either congruent or incongruent with the event she just experienced. Whereas some studies have reported that around their first birthday infants may be able to generate such expectancies (Reschke et al., 2017; Ruba et al., 2020; Skerry & Spelke, 2014), at this age this capacity appears to be very fragile, restricted to some types of events (Skerry & Spelke, 2014; Reschke et al., 2017; Ruba et al., 2019, 2020), and rarely replicated (Ruba et al., 2019, 2020). Other studies with older infants and toddlers suggest that this ability develops substantially during the second year of life, showing more robust and less event-dependent effects at around 1.5 year. Interestingly, a few studies have also explored infant's or toddler's concern for a victim who did not show overt emotional cues of distress following the negative event she experienced. They report such affective perspective taking capacities at the age of 1.5 years (Chiarelli & Poulin-Dubois, 2013; Kassecker et al., 2023; Vaish et al., 2009; but see Chiarella & Poulin-Dubois, 2015), with only null results reported before this age (Chiarelli & Poulin-Dubois, 2013; Vaish, 2016).

In summary, current evidence from the field of early emotion understanding indicates that the second year of life can be considered pivotal for truly being concerned about the victim of an immoral event, especially when the victim does not overtly express their feelings, as it is always the case in the stimuli used in infant experiments. Consequently, on one hand,

preverbal infants, whose affective understanding and empathic concern abilities are still relatively limited, may be less sensitive to information about the outcomes of moral transgressions on victims, making intent-based socio-moral evaluations more likely to occur. On the other hand, these newly emerging emotional abilities may allow toddlers and young preschoolers to substantially increase their consideration of the impact of moral transgressions on others' well-being, resulting in an *outcome bias* as long as their ability to regulate emotions and thoughts remains insufficiently developed (Decety & Holvoet, 2021).

3.4. *The Potential Role of Social and Moral Experiences*

Preschoolers' higher sensitivity to outcomes of agents' prosocial or antisocial actions may rely on yet another crucial factor, that is, their growing experience with moral interactions and *moral outcomes*. Although this factor has often been evoked as potentially critical for the development of moral judgment competencies (e.g., Hamlin, 2013b; Dahl & Killen, 2018; Mammen & Paulus, 2023; Margoni & Surian, 2016; Rogoff et al., 2018), the way in which it could impact the development of moral cognition has surprisingly received little attention in recent literature in relation to the present topic. However, between infancy and toddlerhood, the manner in which young children interact with their physical and social environments undergoes drastic changes, partly due to growing locomotor abilities. Whereas at the beginning of their lives virtually all infants interact with their environment in a positive way⁹ (for a relevant discussion, see Tasimi, 2020), already by the end of the first year and beyond, they start to explore their physical environment more widely leading to many more

⁹ Many caregiver-infant relationships around the world are dominated by positive social interactions and emotional exchanges involving smiling, cooing, making eye contact, and showing signs of pleasure and excitement in response to stimuli (Rubin & Chung, 2013). Some data also reveal that maternal expressions are limited to positive experiences, especially toward younger infants (Malatesta & Haviland, 1982).

parental prohibitions as well as parent-child conflicts (Gralinski & Kopp, 1993; Karasik et al., 2014; Campos et al., 1992). Toddlers also begin to actively interact with other toddlers leading to a substantial increase in conflicts, aggressive behaviors, and harmful peer interactions (e.g., Côté et al., 2007; Hay et al., 2011; Tremblay, 2018). According to Tremblay and collaborators (1999), by 17 months of age, most toddlers are physically aggressive toward siblings, peers and adults (see also Thomsen, 2020). For Hay (2017), who studied parental reports, the maximum level of use of force with peers is achieved at around 18 months of age, then it progressively decreases thereafter. Through these novel social experiences, children are thus exposed to a significant number of moral transgressions, either as persecutors, victims, or mere observers, which might in turn affect their reaction, representation and/or evaluation of moral outcomes (Ruba et al., 2019, 2020; Ruba & Pollak, 2020; Skerry & Spelke, 2014).

3.4.1. Direct Experience of Morally Valenced Outcomes.

When toddlers and young children take part in socio-moral interactions as persecutors or mere observers, they start to be confronted with the expressions of the victim's distress, distress to which they are clearly not indifferent (e.g., Davidov et al., 2021; Geangu et al., 2011; Hepach, 2017; Hepach et al., 2019; Liddle et al., 2015; Roth-Hanania et al., 2011; Smetana, 1984; Zahn-waxler et al., 1979; for a review, see Decety & Holvoet, 2021). When they become the target of such transgressions, they can directly experience the aversive consequences of being physically harmed, having their property taken away, broken, or stolen, or being the victim of unequal distributions of valued resources. These different experiences as observers or victims may help children develop social scripts that allow them to understand the impact of moral transgressions on the well-being of others, and to imagine how the other people might feel when in a similar situation (Ruba & Pollak, 2020; Vaish, 2016; Vaish et al., 2009). As a result, when children are tested with cases in which the transgression is (or not) followed by

an overt reaction of distress from the victim, they will be better able to properly infer and enrich their representation of the negative outcomes these acts may have had on the victim. When toddlers and children are instead faced with moral transgressions that are typically and systematically associated with overt distress in others (i.e., physical harm), a process of negative conditioning might progressively be at work, enabling children to consider these *actions* as aversive, even though outcomes are not explicitly shown (Blair, 1995, 2007; Cushman et al., 2012; Weisberg & Leslie, 2012).

3.4.2. Social Feedback is Outcome-based.

In addition to directly experiencing or witnessing distress from the victim when faced with certain types of interactions, the experience of caregivers' prohibitions could also be an important factor influencing the evaluation of negative outcomes. From their second year of life, children are subjected to parental prohibitions on a daily basis (Dahl, 2016): when they exhibit aggressive behaviors toward a peer or a caregiver (e.g., hitting, biting, kicking), when they engage in activities that are dangerous to themselves (e.g., climbing on furniture), or when they cause trouble (e.g., creating disorder). Interestingly, when it comes to moral transgressions, caregivers appear to be more insistent and less compliant than with other types of potentially prohibited behaviors (Dahl & Campos, 2013; Dahl et al., 2011; Smith et al., 2017). This occurs despite the fact that a number of times children's aggression is not performed with the initial intention to cause harm but to achieve another goal (e.g., get a desired toy; Hay, 2017; Tremblay, 2018). Critically, when caregivers provide verbal explanations for their prohibitions and punishment of the child's behavior, which they tend to do more as children grow up (Dahl, 2016), they focus on the *outcomes* of the children's actions by using either a neutral tone ('Tom is *crying* because you pushed him') or an emotional one ('You made Doug *cry*. It's not nice to bite', 'Can't you see Tom's *hurt*? Don't push him!') (Zahn-

Waxler et al., 1979; see also Dahl, 2016; Nucci & Turiel, 1978; Smith et al., 2017). These explanations have been shown to impact the child's subsequent behaviors (Zahn-Waxler et al., 1979). In sum, the experience with moral interactions that children start to acquire from their second year of life exposes them to a range of emotions and social reactions that may increase their focus on the moral outcomes, which may consequently impact their cognitive and affective reaction to this factor and result in an outcome bias in moral judgment.

4. Proposition and Directions for Future Research

4.1. Proposition

In the previous section, instead of focusing on the factors that could impact children's ability to reason about intentions, we focused on the factors that could impact children's consideration of *outcomes* in their moral judgments. We argue that the type of stimuli used with children (versus those used with infants), combined with their growing emotional understanding and empathic concern abilities and their experience with moral situations they encounter (including the way parents react to transgressions), could *increase* children's moral sensitivity to information about outcomes. In the context of the dual-process models of morality (Buon et al., 2016; Cushman, 2008; Cushman et al., 2007), which posit that being able to generate intent-based judgment relies on the ability to manage competing inputs (intentions and outcomes), this enhancement would lead to a 'reversal of the balance', making outcome-based evaluations stronger than intent-based evaluations. In the case of conflict between the relevant inputs (i.e., accidental harm), and until children's inhibitory control (e.g., Buon et al., 2016; Gvozdic et al., 2016) and/or theory of mind abilities (e.g., Baird & Astington, 2004; Killen et al., 2011; Young & Saxe, 2009) are insufficiently developed, children would be unable to override the negative evaluation triggered by the perception of someone who caused harm and would be unable to retrieve the ability to generate intent-based moral

judgments. Whereas infants seem to be equipped with an initial (and built-in) capacity to both understand socio-moral actions in mentalistic terms (Rosset & Rottman, 2014) and attach negative and positive value to a number of moral outcomes (Ting et al., 2020), further cognitive development and social experience, as well as testing materials, influence the way in which children process the downstream consequences of a moral transgression for a victim and determine an outcome bias in moral judgment.

4.2. What Needs to Be Investigated in Infants

Research is needed to investigate to what extent and in which contexts infants' socio-moral expectations and evaluations are sensitive to outcomes. The studies we have reviewed above that have been conducted to test if infants are sensitive to intentions vs. outcomes have almost exclusively used stimuli reflecting either helping vs. hindering or equal vs. unequal distributions situations (e.g., Hamlin, 2013a; Strid & Meristo, 2020), limiting the generalizability of these findings to other types of moral transgressions and outcomes. The exploration of infants' socio-moral intent- vs. outcome-based preferences and expectations would benefit from encompassing different types of harmful situation (e.g., physical harm) while systematically varying the visual and emotional cues associated with the victim's well-being (e.g., bully conditions in Margoni et al., 2018; Rhodes et al., 2015; Ting & Baillargeon, 2021). If our hypothesis is correct, we anticipate that infants' sensitivity to agents' intentions expressed through their (manual reaching or approaching) preferences will diminish as the availability of overt information regarding the consequences of the action for the well-being of the victim increases. Simultaneously, it would be interesting to assess infants' affective reactions to the stimuli using alternative measures to standard preference and violation-of-expectation measures (e.g., pupil dilation; Hepach & Westermann, 2013; Kassecker et al., 2023). The question of how preverbal infants by comparison to toddlers and preschoolers

represent and affectively react to the impact of a moral action on others' well-being *with and especially without obvious visual cues* would also help figuring out the processes underlying infants' socio-moral evaluations. On one hand, they may be able to represent the impact of moral transgressions on others' well-being even when visual cues are not explicit but might disregard this information to generate their socio-moral evaluations (for an indirect argument in favor of this hypothesis, see Uzevovsky et al., 2020). On the other hand, infants may prioritize information about intentions because their ability to properly represent and affectively react to the impact of a moral transgression on others' well-being, especially without obvious visual cues, is still insufficiently developed (e.g., Chiarella & Poulin-Dubois, 2013; Ruba & Pollak, 2020) and thus cannot sway their evaluations.

4.3. What Needs to Be Investigated in Toddlers and Young Preschoolers

Toddlerhood is a developmental period considered pivotal for acquiring relevant experiences in moral interactions (e.g., Hay, 2017) as well as for improving empathic responding together with affective perspective skills (Decety & Holvoet, 2021; Vaish, 2016) that could influence one's sensitivity to moral outcomes. Experimental data could be enriched by a triangulation with observational data obtained from participants in their natural social environments (Dahl, 2016) and/or data obtained from parents' reports of their children's behaviors and activities (e.g., Dahl & Campos, 2013). Examining the impact of additional contextual variables, such as whether participants grew up with elder siblings (Nichols et al., 2015; Perner et al., 1994) or what type of childcare arrangements they are/were in (NICHD, 2001), on their sensitivity to moral outcomes would also be valuable since these factors too can be the source of some variability in the input children receive regarding moral interactions.

It could also prove useful to investigate if the type of parental style impacts children's sensitivity to moral outcomes, as we posit that parental feedback may play a role in shaping children's moral sensitivity (e.g., Perra et al., 2021). It is important to note that our analysis intentionally focused on parental prohibitions regarding harmful acts occurring between infancy and preschool age since our aim was to assess the discrepancies in evaluations of harmful actions that exist between these two developmental periods. However, it should be acknowledged that the content of caregivers' prohibitions, along with the explanations that follow them, is likely to undergo substantial changes as children grow older (e.g., Smetana, 1984; Zahn-waxler et al., 1979; for a related discussion regarding the development of prosocial behaviors, see Dahl & Brownell, 2019). These changes may also extend to the type of transgressions for which prohibitions are most frequently employed. For instance, we may imagine that caregivers primarily use prohibitions for situations involving physical harm with their toddlers and then extend them to other types of transgressions – such as not sharing, not helping or hindering – as their children get older (Dahl et al., 2017). Analyzing the content and rate of caregiver's prohibitions as a function of both children's developmental period and type of moral situation may thus be relevant to better characterize the contribution of the factors that we have already mentioned (intentions and outcomes) as well as other factors (e.g., negligence, foresight) that, even if omitted from our analysis, are also likely pertinent to discuss the development of moral judgment abilities (e.g., Gönültaş et al., 2021; Kneer & Skoczeń, 2023; Margoni & Surian, 2022; Nobes & Martin, 2022; Nobes et al., 2009; Woo et al., 2017).

Last, in addition to elucidate how all these factors as well as methodological differences in how infants and preschoolers are tested impact on young preschoolers' propensity to generate outcome- vs. intent-based moral judgments, from our proposal it follows that it will

be crucial to assess how limitations in the executive functioning skills interact with an enhanced sensitivity for moral outcomes in preschoolers to determine an outcome bias in moral judgment.

5. Final Remarks

5.1. Is There Only One Candidate for Explaining the Discrepancy Between Infants' and Children's Moral Performances?

In this paper, we provide an alternative reading of the discrepancy between infants' and children's moral competencies to enrich and complement both current developmental accounts of the outcome-to-intent shift (Hamlin, 2013a; Hilton & Kuhlmeier, 2019; Margoni & Surian, 2016, 2020; Proft & Rakoczy, 2019) and recent dual-process models (Buon et al., 2016; Cushman, 2008; Martin, Leddy, et al., 2021). Our review of the available accounts arising from developmental literature led us to recognize that the existing disparity between infants' and children's moral competencies is likely, at least in part, due to differences in the tasks used to test these two populations. These differences are numerous, and we believe it is unlikely that only one of them can explain the entire developmental pattern. For example, the demands placed on response generation may be the cause of 3-year-olds' struggles in generating intent-based moral judgments in classical moral judgment tasks, but other factors should explain why 4- and 5-year-olds sometimes also fail to generate intent-based judgments in classical moral tasks, given that they can generate intent-based judgments *even when* response generation demands are not reduced (Margoni & Surian, 2020). Although recent research has started to investigate the impact of potential candidates for explaining the outcome-to-intent shift (e.g., Margoni & Surian, 2020; Proft & Rakoczy, 2019; Van de Vondervoort & Hamlin, 2018), the conclusions we can draw from it are still limited due to the fact that in most of these studies multiple variables were manipulated at the same time. In other words, because each task that

was employed in this recent literature differed from classical tasks in several aspects (e.g., the format in which stimuli are presented, the salience of intentional cues, the type of intentional cues, the reduction of processing demands, etc.), it is difficult to understand which aspect(s) of the task(s) is critical for participants to be able to generate intent-based moral judgments. Based on our analysis of the intent-to-outcome and outcome-to-intent developmental shifts, we emphasize the need to control – and systematically vary – the type of morally harmful actions presented, as well as the way in which outcomes are displayed in the stimuli. We additionally raise the hypothesis that children's initial experiences with moral interactions and awareness of the downstream negative consequences of harmful actions for a victim, as well as their emotional reaction to and emotional understanding of moral outcomes, would determine an outcome bias which will interfere with the expression of the ability to generate intent-based moral judgments.

5.2. *Conceptual Continuity vs. Conceptual Change Accounts*

One might object to the present paper that it does not take a clear position regarding the debate on the nature of the (intent-to-)outcome-to-intent developmental pattern, that is, whether these shifts are better and solely explained by changes occurring outside the moral domain (*conceptual continuity view*; Margoni & Surian, 2016, 2020) or instead they reflect conceptual acquisitions within the moral domain (*conceptual change view*; Cushman et al., 2013). We did this intentionally. Beyond the fact that each of us holds a slightly different perspective on this topic, we both believe that whereas the current state of research is more in line with the conceptual continuity view, it does not yet provide us with sufficient unambiguous evidence to take a definitive stance in favor of either of these positions.

On one hand, being able to improve young preschoolers' ability to generate intent-based moral judgments with the help of diverse experimental procedures (e.g., Margoni &

Surian, 2020; Nobes et al., 2016, 2017; Proft & Rakoczy, 2019) clearly argues in favor of the conceptual continuity view. In particular, crucial is the evidence that *3-year-olds*, who – according to the conceptual change proponents – are believed to possess only outcome-based concepts of moral goodness and badness, are instead able to express a mature conceptual repertoire, if only the moral judgment task is simplified enough (Margoni & Surian, 2020).

We have proposed that infants might prove to be sensitive to moral outcomes if the stimuli they were presented with contained all the relevant cues, especially those related to the impact of the agent's action on the patient's well-being. If this turns out to be the case, one could argue that infants, toddlers, and children are all sensitive to information about the agents' intentions and the caused outcomes. Prioritizing information about intentions would then depend on both the availability of cues and the depth to which discarding or downplaying the information about outcomes would require the use of past experience and various psychological abilities such as theory of mind skills, affective perspective skills, executive functioning skills, and inferential reasoning abilities. Our proposal that due to the stimuli used, the development of empathic response and the experience with moral interactions and parental feedback, preschoolers exhibit an outcome bias which coupled with limited executive functioning abilities temporarily disrupts their ability to properly detect and integrate information about intentions in their moral judgment, would then be in line with the conceptual continuity view.

On the other hand, we discussed a model according to which different types of moral judgments would rely differently on information about intentions and outcomes (see the Cushman's dual-process model mentioned above; Cushman, 2008, 2013; Cushman et al., 2013). In line with this model, Martin, Leddy et al. (2021) have reported that whereas preschoolers were already able to generate intent-based responses when asked to choose a

playmate (partner choice), a strong outcome-to-intent shift was noticeable when children were asked to assign punishment. Evidence of the same dissociation between developmental trends in naughtiness judgments and punishment judgments has also been reported by Cushman et al. (2013) (see also, Martin et al., 2021, for results on adults under cognitive load). In particular, this latter study reported evidence that during development children start to make intent-based punishment judgments only after (and because) they start to make intent-based naughtiness judgments. Also, at the processing level, children were tested for both judgments (counterbalancing the order of test questions across participants), and punishment judgments were found to be influenced by prior (intent-based) naughtiness judgments, but not vice versa. Since there is no reason that resources outside the moral domain (theory of mind, executive functioning) would be more required for one question type than another, conceptual change proponents would conclude that these results are hardly reconcilable with the conceptual continuity view. However, the findings reported by Cushman et al. (2013) we have just reviewed were not replicated by the only currently existing attempt to directly replicate them (see Study 1 in Margoni and Surian, 2017). A valuable direction for future work will be to pool together these data to estimate through a meta-analysis the strength of the evidence (or lack thereof) for the conceptual change view (Margoni & Shepperd, 2020).

Let us now return to the question of the discrepancy between infants' and preschoolers' performance in generating intent-based moral judgments. If we keep in mind that the measures used in infant studies are based on the choice of one agent over another, which might be considered as a partner choice (Martin, Leddy, et al., 2021), an intriguing possibility is that conceptual continuity between infants and preschoolers would exist for some types of judgments (partner choice and perhaps other types of moral judgments) but not for others (punishment judgments). Along this line of reasoning, it could be hypothesized that

punishment judgments, which mainly take information about the causal responsibility for the negative outcomes as input (Cushman, 2008; Kneer & Machery, 2019), would develop independently from the other judgments and at least partly thanks to children's experiences and in-depth processing of the impact of actions on others' well-being. However, in addition to being intriguing, this hypothesis remains speculative. Thus, a valuable direction for future work will be to clearly establish to what extent the measures currently used with infants reflect the measures used with preschool- and school-age children (e.g., moral judgment, partner choice judgment, punishment judgment). Additionally, research will benefit from exploring infants' socio-moral abilities by using other measures that are more likely to reflect punishment judgments (for novel procedures designed to assess infants' punishment decisions, see Kanakogi et al., 2022; Ziv et al., 2021).

5.3. Conclusion

This paper focuses on a long-standing debate that has recently garnered renewed attention from the scientific community. By examining the differences between testing procedures used with infants and children, recent efforts to elucidate why infants seem to outperform young preschoolers in generating intent-based socio-moral evaluations have begun to identify several promising candidates. Although further experimental research is needed to precisely isolate the impact of these factors on individuals' moral competences and the underlying processes, they likely play a role in explaining at least a part of this intricate developmental pattern we refer as intent-to-outcome and outcome-to-intent shifts. However, we believe that one crucial factor has been overlooked in the current debate: *outcomes*. By emphasizing how consequences may be less salient in the procedures used with infants than in those used with preschoolers and how preschoolers might be over-sensitive to outcomes,

we aim to contribute a meaningful piece to this developmental puzzle that might have many days ahead of it.

References

- Abelard, P. (1971). *Ethics* (trans. Luscombe, D. E.). Oxford: Oxford University Press.
- Amir, D., Ahl, R. E., Parsons, W. S., & McAuliffe, K. (2021). Children are more forgiving of accidental harms across development. *Journal of Experimental Child Psychology*, 205, 105081. <https://doi.org/10.1016/j.jecp.2020.105081>
- Armsby, R. E. (1971). A reexamination of the development of moral judgments in children. *Child Development*, 1241-1248. <https://doi.org/10.1111/j.1467-8624.1971.tb02017.x>
- Baillargeon, R., Scott, R. M., & Bian, L. (2016). Psychological reasoning in infancy. *Annual Review of Psychology*, 67, 159-186. <https://doi.org/10.1146/annurev-psych-010213-115033>
- Baillargeon, R., Scott, R. M., & He, Z. (2010). False-belief understanding in infants. *Trends in Cognitive Sciences*, 14, 110-118. <https://doi.org/10.1016/j.tics.2009.12.006>
- Baird, J. A., & Astington, J. W. (2004). The role of mental state understanding in the development of moral cognition and moral action. *New Directions for Child and Adolescent Development*, 103, 37-49. <https://doi.org/10.1002/cd.96>
- Baker, E. R., D'Esterre, A. P., & Weaver, J. P. (2021). Executive function and Theory of Mind in explaining young children's moral reasoning: A Test of the Hierarchical Competing Systems Model. *Cognitive Development*, 58, 101035. <https://doi.org/10.1016/j.cogdev.2021.101035>
- Barrett, H. C., Bolyanatz, A., Crittenden, A. N., Fessler, D. M., Fitzpatrick, S., Gurven, M., ... & Laurence, S. (2016). Small-scale societies exhibit fundamental variation in the role of intentions in moral judgment. *Proceedings of the National Academy of Sciences*, 113, 4688-4693. <https://doi.org/10.1073/pnas.1522070113>

Behne, T., Carpenter, M., Call, J., & Tomasello, M. (2005). Unwilling versus unable: Infants' understanding of intentional action. *Developmental Psychology*, 41, 328–337.

<https://doi.org/10.1037/0012-1649.41.2.328>

Best, J. R., & Miller, P. H. (2010). A developmental perspective on executive function. *Child Development*, 81, 1641-1660. <https://doi.org/10.1111/j.1467-8624.2010.01499.x>

Bian, L., Sloane, S., & Baillargeon, R. (2018). Infants expect ingroup support to override fairness when resources are limited. *Proceedings of the National Academy of Sciences*, 115, 2705-2710. <https://doi.org/10.1073/pnas.1719445115>

Blair, R. J. R. (1995). A cognitive developmental approach to morality: investigating the psychopath. *Cognition*, 57, 1–29. [https://doi.org/10.1016/0010-0277\(95\)00676-P](https://doi.org/10.1016/0010-0277(95)00676-P)

Blair, R. J. R. (2007). The amygdala and ventromedial prefrontal cortex in morality and psychopathy. *Trends in Cognitive Sciences*, 11, 7–12.

<https://doi.org/10.1016/j.tics.2007.07.003>

Blair, R. J. R. (2017). Emotion-based learning systems and the development of morality. *Cognition*, 167, 38–45. <https://doi.org/10.1016/j.cognition.2017.03.013>

Bloom, P. (2017). Empathy and its discontents. *Trends in Cognitive Sciences*, 21, 24-31. <https://doi.org/10.1016/j.tics.2016.11.004>

Brandone, A. C., & Wellman, H. M. (2009). You can't always get what you want: Infants understand failed goal-directed actions. *Psychological Science*, 20, 85-91.

<https://doi.org/10.1111/j.1467-9280.2008.02246.x>

Buon, M., Dupoux, E., Jacob, P., Chaste, P., & Zalla, T. (2013). The role of causal and intentional judgments in moral reasoning in individuals with high functioning autism. *Journal of Autism and Developmental Disorders*, 43, 458–470.

<https://doi.org/10.1007/s10803-012-1588-7>

Buon, M., Jacob, P., Loissel, E., & Dupoux, E. (2013). A non-mentalistic cause-based heuristic in human social evaluations. *Cognition*, 126, 149–155.

<https://doi.org/10.1016/j.cognition.2012.09.006>

Buon, M., Jacob, P., Margules, S., Brunet, I., Dutat, M., Cabrol, D., & Dupoux, E. (2014). Friend or foe? Early social evaluation of human interactions. *PLoS ONE*, 9, e88612.

<https://doi.org/10.1371/journal.pone.0088612>

Buon, M., Seara-cardoso, A., & Viding, E. (2016). Why (and how) should we study the interplay between emotional arousal, Theory of Mind, and inhibitory control to understand moral cognition? *Psychonomic Bulletin & Review*, 23, 1660-1680.

<https://doi.org/10.3758/s13423-016-1042-5>

Buttelmann, D., Carpenter, M., & Tomasello, M. (2009). Eighteen-month-old infants show false belief understanding in an active helping paradigm. *Cognition*, 112, 337-342.

<https://doi.org/10.1016/j.cognition.2009.05.006>

Campos, J. J., Mumme, D., Kermoian, R., & Campos, R. G. (1994). A functionalist perspective on the nature of emotion. *Japanese Journal of Research on Emotions*, 2, 1-20.

<https://doi.org/10.4092/jsre.2.1>

Carlson, S. M., & Moses, L. J. (2001). Individual differences in inhibitory control and children's theory of mind. *Child Development*, 72, 1032-1053.

<https://doi.org/10.1111/1467-8624.00333>

Caporaso, J. S., Marcovitch, S., & Boseovski, J. J. (2021). Executive function and the development of social information processing during the preschool years. *Cognitive Development*, 58, 101018. <https://doi.org/10.1016/j.cogdev.2021.101018>

Chandler, M. J., Greenspan, S., Barenboim, C., Chandler, M. J., & Greenspan, S. (1973).

Judgments of intentionality in response to videotaped and verbally presented moral

dilemmas: the medium is the message. *Child Development*, 44, 315–320.

<https://doi.org/10.1111/j.1467-8624.1973.tb02157.x>

Chernyak, N., & Sobel, D. M. (2016). “But he didn’t mean to do it”: Preschoolers correct punishments imposed on accidental transgressors. *Cognitive Development*, 39, 13–20.

<https://doi.org/10.1016/j.cogdev.2016.03.002>

Chiarella, S. S., & Poulin-Dubois, D. (2013). Cry babies and pollyannas: Infants can detect unjustified emotional reactions. *Infancy*, 18, 81-96. <https://doi.org/10.1111/infa.12028>

Chiarella, S. S., & Poulin-Dubois, D. (2015). “Aren’t you supposed to be sad?” Infants do not treat a stoic person as an unreliable emoter. *Infant Behavior and Development*, 38, 57-66. <https://doi.org/10.1016/j.infbeh.2014.12.007>

Cho, I., Song, H. Joo, Kim, H., & Sul, S. (2020). Older adults consider others’ intentions less but allocentric outcomes more than young adults during an ultimatum game.

Psychology and Aging, 35, 974-980. <https://doi.org/10.1037/pag0000577>

Choi, Y., & Luo, Y. (2023). Understanding preferences in infancy. *WIREs Cognitive Science*, 14, e1643. <https://doi.org/10.1002/wcs.1643>

Costanzo, P. R., Coie, J. D., Grumet, J. F., & Farnill, D. (1973). A reexamination of the effects of intent and consequence on children’s moral judgments. *Child Development*, 44, 154–161. <https://doi.org/10.1111/j.1467-8624.1973.tb02127.x>

Côté, S. M., Vaillancourt, T., Barker, E. D., Nagin, D., & Tremblay, R. E. (2007). The joint development of physical and indirect aggression : Predictors of continuity and change during childhood. *Development and Psychopathology*, 19, 37-55.

<https://doi.org/10.1017/s0954579407070034>

Cushman, F. (2008). Crime and punishment : Distinguishing the roles of causal and intentional analyses in moral judgment. *Cognition*, 108, 353–380.

<https://doi.org/10.1016/j.cognition.2008.03.006>

Cushman, F. (2013). Action, outcome, and value: A dual-system framework for morality. *Personality and Social Psychology Review*, 17, 273-292.

<https://doi.org/10.1177/1088868313495594>

Cushman, F., Dreber, A., Wang, Y., & Costa, J. (2009). Accidental outcomes guide punishment in a “trembling hand” game. *PloS One*, 4, e6699.

<https://doi.org/10.1371/journal.pone.0006699>

Cushman, F., Gray, K., Gaffey, A., & Mendes, W. B. (2012). Simulating murder: The aversion to harmful action. *Emotion*, 12, 2–7. <https://doi.org/10.1037/a0025071>

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

Cushman, F., Young, L., & Greene, J. D. (2007). Our multisystem moral psychology: Towards a consensus view. In J. M. Doris (Ed.), *Oxford handbook of moral psychology*. Oxford: Oxford University Press.

Dahl, A. (2016). Mothers’ insistence when prohibiting infants from harming others in everyday interactions. *Frontiers in Psychology*, 7, 215992.

<https://doi.org/10.3389/fpsyg.2016.01448>

Dahl, A., & Brownell, C. A. (2019). The social origins of human prosociality. *Current Directions in Psychological Science*, 28, 274-279. <https://doi.org/10.1177/0963721419830386>

Dahl, A., & Campos, J. J. (2013). Domain differences in early social interactions. *Child Development*, 84, 817–825. <https://doi.org/10.1111/cdev.12002>

Dahl, A., Campos, J. J., & Witherington, D. C. (2011). Emotional action and communication in early moral development. *Emotion Review*, 3, 147-157.

<https://doi.org/10.1177/1754073910387948>

- Dahl, A., & Killen, M. (2018). Moral reasoning: Theory and research in developmental science. In Wixted J., Ghetti S. (Eds.), *Stevens' handbook of experimental psychology and cognitive neuroscience. Vol. 4: Developmental and social psychology* (4th ed., pp. 323-356). Wiley. <https://doi.org/10.1002/9781119170174.epcn410>
- Dahl, A., Satlof-Bedrick, E. S., Hammond, S. I., Drummond, J. K., Waugh, W. E., & Brownell, C. A. (2017). Explicit scaffolding increases simple helping in younger infants. *Developmental Psychology*, 53, 407-416. <https://doi.org/10.1037/dev0000244>
- Davidov, M., Paz, Y., Roth-Hanania, R., Uzefovsky, F., Orlitsky, T., Mankuta, D., & Zahn-Waxler, C. (2021). Caring babies: Concern for others in distress during infancy. *Developmental Science*, 24, e13016. <https://doi.org/10.1111/desc.13016>
- Davidov, M., Zahn-Waxler, C., Roth-Hanania, R., & Knafo, A. (2013). Concern for others in the first year of life: Theory, evidence, and avenues for research. *Child Development Perspectives*, 7, 126-131. <https://doi.org/10.1111/cdep.12028>
- Decety, J. (2010). The neurodevelopment of empathy in humans. *Developmental Neuroscience*, 32, 257–267. <https://doi.org/10.1159/000317771>
- Decety, J., & Holvoet, C. (2021). The emergence of empathy: A developmental neuroscience perspective. *Developmental Review*, 62, 100999. <https://doi.org/10.1016/j.dr.2021.100999>
- Dondi, M., Simion, F., & Caltran, G. (1999). Can newborns discriminate between their own cry and the cry of another newborn infant? *Developmental Psychology*, 35, 418–426. <https://doi.org/10.1037/0012-1649.35.2.418>
- Dunfield, K. A., & Kuhlmeier, V. A. (2010). Intention-mediated selective helping in infancy. *Psychological Science*, 21, 523-527. <https://doi.org/10.1177/0956797610364119>

- Eisenberg, N. (2000). Emotion, regulation, and moral development. *Annual Review of Psychology*, 51, 665-697. <https://doi.org/10.1146/annurev.psych.51.1.665>
- Farnill, D. (1974). The effects of social-judgment set on children's use of intent information. *Journal of Personality*, 42, 276-289. <https://doi.org/10.1111/j.1467-6494.1974.tb00674.x>
- Fu, G., Xiao, W. S., Killen, M., & Lee, K. (2014). Moral judgment and its relation to second-order theory of mind. *Developmental Psychology*, 50, 2085-2092. <https://doi.org/10.1037/a0037077>
- Geangu, E., Benga, O., Stahl, D., & Striano, T. (2011). Individual differences in infants' emotional resonance to a peer in distress: Self–other awareness and emotion regulation. *Social Development*, 20, 450–470. <https://doi.org/10.1111/j.1467-9507.2010.00596.x>
- Geraci, A., Simion, F., & Surian, L. (2022). Infants' intention-based evaluations of distributive actions. *Journal of Experimental Child Psychology*, 220, 105429. <https://doi.org/10.1016/j.jecp.2022.105429>
- Gergely, G., Nádasdy, Z., Csibra, G., & Bíró, S. (1995). Taking the intentional stance at 12 months of age. *Cognition*, 56, 165-193. [https://doi.org/10.1016/0010-0277\(95\)00661-h](https://doi.org/10.1016/0010-0277(95)00661-h)
- Gönültaş, S., Richardson, C. B., & Mulvey, K. L. (2021). But they weren't being careful! Role of theory of mind in moral judgments about victim and transgressor negligence. *Journal of Experimental Child Psychology*, 212, 105234. <https://doi.org/10.1016/j.jecp.2021.105234>
- Gralinski, J. H., & Kopp, C. B. (1993). Everyday rules for behavior: Mothers' requests to young children. *Developmental Psychology*, 29, 573-584. <https://doi.org/10.1037/0012-1649.29.3.573>

- Gray, K., Young, L., & Waytz, A. (2012). Mind perception is the essence of morality. *Psychological Inquiry*, 23, 101-124. <https://doi.org/10.1080/1047840x.2012.651387>
- Greene, J. D., Nystrom, L. E., Engell, A. D., Darley, J. M., & Cohen, J. D. (2004). The neural bases of cognitive conflict and control in moral judgment. *Neuron*, 44, 389-400. <https://doi.org/10.1016/j.neuron.2004.09.027>
- Grossmann, T. (2010). The development of emotion perception in face and voice during infancy. *Restorative Neurology and Neuroscience*, 28, 219-236. <https://doi.org/10.3233/RNN-2010-0499>
- Gvozdic, K., Moutier, S., Dupoux, E., & Buon, M. (2016). Priming children's use of intentions in moral judgement with metacognitive training. *Frontiers in Psychology*, 7, 172662. <https://doi.org/10.3389/fpsyg.2016.00190>
- Hackel, L. M., Doll, B. B., & Amodio, D. M. (2015). Instrumental learning of traits versus rewards: dissociable neural correlates and effects on choice. *Nature Neuroscience*, 18, 1233-1235. <https://doi.org/10.1038/nn.4080>
- Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review*, 108, 814-834. <https://doi.org/10.1037/0033-295X.108.4.814>
- Haidt, J. (2007). The new synthesis in moral psychology. *Science*, 316, 998-1002. <https://doi.org/10.1126/science.1137651>
- Hamlin, J. K. (2013a). Failed attempts to help and harm: Intention versus outcome in preverbal infants' social evaluations. *Cognition*, 128, 451-474. <https://doi.org/10.1016/j.cognition.2013.04.004>
- Hamlin, J. K. (2013b). Moral judgment and action in preverbal infants and toddlers: Evidence for an innate moral core. *Current Directions in Psychological Science*, 22, 186-193.

<https://doi.org/10.1177/0963721412470687>

Hamlin, K. J., Ullman, T., Tenenbaum, J., Goodman, N., & Baker, C. (2013). The mentalistic basis of core social cognition: Experiments in preverbal infants and a computational model. *Developmental Science*, 16, 209–226. <https://doi.org/10.1111/desc.12017>

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

Hay, D. F. (2017). The early development of human aggression. *Child Development Perspectives*, 11, 102-106. <https://doi.org/10.1111/cdep.12220>

Hay, D. F., Hurst, S. L., Waters, C. S., & Chadwick, A. (2011). Infants' use of force to defend toys: The origins of instrumental aggression. *Infancy*, 16, 471-489. <https://doi.org/10.1111/j.1532-7078.2011.00069.x>

Hay, D. F., Nash, A., & Pedersen, J. (1981). Responses of six-month-olds to the distress of their peers. *Child Development*, 52, 1071-1075. <https://doi.org/10.1111/j.1467-8624.1981.tb03151.x>

Hebble, P. W. (1971). The development of elementary school children's judgment of intent. *Child Development*, 42, 1203-1215. <https://doi.org/10.1111/j.1467-8624.1971.tb02014.x>

Helwig, C. C., Hildebrandt, C., & Turiel, E. (1995). Children's judgments about psychological harm in social context. *Child Development*, 66, 1680-1693. <https://doi.org/10.1111/j.1467-8624.1995.tb00958.x>

Hepach, R. (2017). Prosocial arousal in children. *Child Development Perspectives*, 11, 50-55. <https://doi.org/10.1111/cdep.12209>

Hepach, R., Vaish, A., Müller, K., & Tomasello, M. (2019). The relation between young children's physiological arousal and their motivation to help others. *Neuropsychologia*,

- 126, 113-119. <https://doi.org/10.1016/j.neuropsychologia.2017.10.010>
- Hepach, R., & Westermann, G. (2013). Infants' sensitivity to the congruence of others' emotions and actions. *Journal of Experimental Child Psychology*, 115, 16-29. <https://doi.org/10.1016/j.jecp.2012.12.013>
- Hilton, B. C., & Kuhlmeier, V. A. (2019). Intention attribution and the development of moral evaluation. *Frontiers in Psychology*, 9, 423561. <https://doi.org/10.3389/fpsyg.2018.02663>
- Huebner, B., Dwyer, S., & Hauser, M. (2009). The role of emotion in moral psychology. *Trends in Cognitive Sciences*, 13, 1-6. <https://doi.org/10.1016/j.tics.2008.09.006>
- Hoffman, M. L. (2000). *Empathy and moral development: Implications for caring and justice*. Cambridge, UK: Cambridge University Press.
- Imamoglu, E. O. (1975). Children's awareness and usage of intention cues. *Child Development*, 46, 39-45. <https://doi.org/10.2307/1128831>
- JASP Team (2022). JASP (Version 0.16.3) [Computer software]. <https://jasp-stats.org/>
- Jin, K. S., & Baillargeon, R. (2017). Infants possess an abstract expectation of ingroup support. *Proceedings of the National Academy of Sciences*, 114, 8199-8204. <https://doi.org/10.1073/pnas.1706286114>
- Kanakogi, Y., Inoue, Y., Matsuda, G., Butler, D., Hiraki, K., & Myowa-Yamakoshi, M. (2017). Preverbal infants affirm third-party interventions that protect victims from aggressors. *Nature Human Behaviour*, 1, 0037. <https://doi.org/10.1038/s41562-016-0037>
- Kanakogi, Y., Miyazaki, M., Takahashi, H., Kobayashi, T., & Hiraki, K. (2022). Third-party punishment by preverbal infants. *Nature Human Behaviour*, 6, 1234-1242. <https://doi.org/10.1038/s41562-022-01354-2>

- Karasik, L. B., Tamis-LeMonda, C. S., Adolph, K. E. (2014). Crawling and walking infants elicit different verbal responses from mothers. *Developmental Science*, 17, 388–395.
<https://doi.org/10.1111/desc.12129>
- Karniol, R. (1978). Children's use of intention cues in evaluating behavior. *Psychological Bulletin*, 85, 76-85. <https://doi.org/10.1037/0033-2909.85.1.76>
- Kassecker, A., Verschoor, S. A., & Schmidt, M. F. (2023). Human infants are aroused and concerned by moral transgressions. *Proceedings of the National Academy of Sciences*, 120, e2306344120. <https://doi.org/10.1073/pnas.2306344120>
- Killen, M., Lynn Mulvey, K., Richardson, C., Jampol, N., & Woodward, A. (2011). The accidental transgressor: Morally-relevant theory of mind. *Cognition*, 119, 197–215.
<https://doi.org/10.1016/j.cognition.2011.01.006>
- Killen, M., & Smetana, J. G. (2015). Origins and development of morality. In M. E. Lamb & R. M. Lerner (Eds.), *Handbook of child psychology and developmental science: Socioemotional processes* (7th ed., pp. 701–749). Wiley.
<https://doi.org/10.1002/9781118963418.childpsy317>
- King, M. (1971). The development of some intention concepts in young children. *Child Development*, 42, 1145-1152. <https://doi.org/10.1111/j.1467-8624.1971.tb02009.x>
- Knafo, A., Zahn-Waxler, C., Van Hulle, C., Robinson, J. L., & Rhee, S. H. (2008). The developmental origins of a disposition toward empathy: Genetic and environmental contributions. *Emotion*, 8, 737-752. <https://doi.org/10.1037/a0014179>
- Kneer, M., & Machery, E. (2019). No luck for moral luck. *Cognition*, 182, 331-348.
<https://doi.org/10.1016/j.cognition.2018.09.003>
- Kneer, M., & Skoczeń, I. (2023). Outcome effects, moral luck and the hindsight bias. *Cognition*, 232, 105258. <https://doi.org/10.1016/j.cognition.2022.105258>

- Lee, Y. E., Ellie Yun, J. E., Kim, E. Y., & Song, H. J. (2015). The development of infants' sensitivity to behavioral intentions when inferring others' social preferences. *PLoS ONE*, *10*, e0135588. <https://doi.org/10.1371/journal.pone.0135588>
- Li, J., Hou, W., Zhu, L., & Tomasello, M. (2020). The development of intent-based moral judgment and moral behavior in the context of indirect reciprocity: A cross-cultural study. *International Journal of Behavioral Development*, *44*, 525-533. <https://doi.org/10.1177/0165025420935636>
- Li, J., & Tomasello, M. (2018). The development of intention-based sociomoral judgment and distribution behavior from a thirdparty stance. *Journal of Experimental Child Psychology*, *167*, 78–92. <https://doi.org/10.1016/j.jecp.2017.09.021>
- Liddle, M. J. E., Bradley, B. S., & Mcgrath, A. (2015). Baby empathy: Infant distress and peer prosocial responses. *Infant Mental Health Journal*, *36*, 446-458. <https://doi.org/10.1002/imhj.21519>
- Lucca, K., Capelier-Mourguy, A., Cirelli, L., Byers-Heinlein, K., Ben, R. D., et al. (2021). *Infants' social evaluation of helpers and hinderers: a large-scale, multi-lab, coordinated replication study*. Pre-print retrieved at <https://doi.org/10.31234/osf.io/qhxxm>
- Lucca, K., Pospisil, J., & Sommerville, J. A. (2018). Fairness informs social decision making in infancy. *PloS one*, *13*, e0192848. <https://doi.org/10.1371/journal.pone.0192848>
- Malatesta, C. Z., & Haviland, J. M. (1982). Learning display rules: the socialization of emotion expression in infancy. *Child Development*, *53*, 991-1003. <https://doi.org/10.2307/1129139>
- Malle, B. F. (2021). Moral judgments. *Annual Review of Psychology*, *72*, 293-318. <https://doi.org/10.1146/annurev-psych-072220-104358>
- Malle, B. F. (2024). Blame and punishment. In B. F. Malle & P. Robbins (Eds.), *Cambridge*

Handbook of Moral Psychology. Cambridge University Press.

Mammen, M., & Paulus, M. (2023). The communicative nature of moral development: A theoretical framework on the emergence of moral reasoning in social interactions. *Cognitive Development*, 66, 101336.

<https://doi.org/10.1016/j.cogdev.2023.101336>

Margolis, E., & Laurence, S. (2013). In defense of nativism. *Philosophical Studies*, 165, 693-718. <https://doi.org/10.1007/s11098-012-9972-x>

Margoni, F., Baillargeon, R., & Surian, L. (2018). Infants distinguish between leaders and bullies. *Proceedings of the National Academy of Sciences*, 115, E8835-E8843.

<https://doi.org/10.1073/pnas.1801677115>

Margoni, F., Cho, I., & Gutchess, A. (2023). Intent-based moral judgment in old age. *The Journals of Gerontology: Series B*, 78, 1136-1141.

<https://doi.org/10.1093/geronb/gbac114>

Margoni, F., Geipel, J., Hadjichristidis, C., Bakiav, R., & Surian, L. (2023). Age-related differences in moral judgment: The role of probability judgments. *Cognitive Science*, 47, e13345. <https://doi.org/10.1111/cogs.13345>

Margoni, F., Geipel, J., Hadjichristidis, C., & Surian, L. (2018). Moral judgment in old age: Evidence for an intent-to-outcome shift. *Experimental Psychology*, 65, 105–114.

<https://doi.org/10.1027/1618-3169/a000395>

Margoni, F., Geipel, J., Hadjichristidis, C., & Surian, L. (2019). The influence of agents' negligence in shaping younger and older adults' moral judgment. *Cognitive Development*, 49, 116–126. <https://doi.org/10.1016/j.cogdev.2018.12.002>

Margoni, F., Geipel, J., Hadjichristidis, C., & Surian, L. (2021). Inequity aversion in old age: An outcome bias in older adults' socio-economic decisions. *Cognitive Development*, 58,

101016. <https://doi.org/10.1016/j.cogdev.2021.101016>
- Margoni, F., Guglielmetti, G., & Surian, L. (2019). Brief report: Young children with autism can generate intent-based moral judgments. *Journal of Autism and Developmental Disorders*, 49, 5078-5085. <https://doi.org/10.1007/s10803-019-04212-9>
- Margoni, F., & Shepperd, M. (2020). Changing the logic of replication: A case from infant studies. *Infant Behavior and Development*, 61, 101483. <https://doi.org/10.1016/j.infbeh.2020.101483>
- Margoni, F., & Surian, L. (2016). Explaining the U-shaped development of intent-based moral judgments. *Frontiers in Psychology*, 7, 171613. <https://doi.org/10.3389/fpsyg.2016.00219>
- Margoni, F., & Surian, L. (2017). Children's intention-based moral judgments of helping agents. *Cognitive Development*, 41, 46-64. <https://doi.org/10.1016/j.cogdev.2016.12.001>
- Margoni, F., & Surian, L. (2020). Conceptual continuity in the development of intent-based moral judgment. *Journal of Experimental Child Psychology*, 194, 104812. <https://doi.org/10.1016/j.jecp.2020.104812>
- Margoni, F., & Surian, L. (2022). Judging accidental harm: Due care and foreseeability of side effects. *Current Psychology*, 41, 8774–8783. <https://doi.org/10.1007/s12144-020-01334-7>
- Margoni, F., Surian, L., & Baillargeon, R. (2024). The violation-of-expectation paradigm: A conceptual overview. *Psychological Review*, 131, 716-748. <https://doi.org/10.1037/rev0000450>
- Margoni, F., & Thomsen, L. (2023). *How infants predict respect-based power*. Pre-print retrieved at <https://doi.org/10.31234/osf.io/m426d>

- Martin, J. W., Buon, M., & Cushman, F. (2021). The effect of cognitive load on intent-based moral judgment. *Cognitive Science*, 45, e12965. <https://doi.org/10.1111/cogs.12965>
- Martin, J. W., Leddy, K., Young, L., & McAuliffe, K. (2021). An earlier role for intent in children's partner choice versus punishment. *Journal of Experimental Psychology: General*, 151, 597–612. <https://doi.org/10.1037/xge0001093>
- Mascaro, O., & Csibra, G. (2012). Representation of stable social dominance relations by human infants. *Proceedings of the National Academy of Sciences*, 109, 6862-6867. <https://doi.org/10.1073/pnas.1113194109>
- McNamara, R. A., Willard, A. K., Norenzayan, A., & Henrich, J. (2019). Weighing outcome vs. intent across societies: How cultural models of mind shape moral reasoning. *Cognition*, 182, 95-108. <https://doi.org/10.1016/j.cognition.2018.09.008>
- Meristo, M., & Surian, L. (2013). Do infants detect indirect reciprocity? *Cognition*, 129, 102-113. <https://doi.org/10.1016/j.cognition.2013.06.006>
- Mikhail, J. (2007). Universal moral grammar : theory , evidence and the future. *Trends in Cognitive Sciences*, 11, 143-152. <https://doi.org/10.1016/j.tics.2006.12.007>
- Minton, A. R., Snyder, J. S., Young, N. A., Graupmann, V., & Mikels, J. A. (2024). Motives matter more with age: Adult age differences in response to sociomoral violations. *Journal of Experimental Psychology: General*. Advance online publication. <https://dx.doi.org/10.1037/xge0001578>
- Moran, J. M., Young, L. L., Saxe, R., Lee, S. M., O'Young, D., Mavros, P. L., & Gabrieli, J. D. (2011). Impaired theory of mind for moral judgment in high-functioning autism. *Proceedings of the National Academy of Sciences of the United States of America*, 108, 2688-2692. <https://doi.org/10.1073/pnas.1011734108>
- Monroe, A. E., & Malle, B. F. (2017). Two paths to blame: Intentionality directs moral

- information processing along two distinct tracks. *Journal of Experimental Psychology: General*, 146, 123-133. <https://doi.org/10.1037/xge0000234>
- Moran, J. M., Jolly, E., & Mitchell, J. P. (2012). Social-cognitive deficits in normal aging. *Journal of Neuroscience*, 32, 5553–5561. <https://doi.org/10.1523/jneurosci.5511-11.2012>
- Moran, J. D., & O'brien, G. (1983). The development of intention-based moral judgments in three-and four-year-old children. *The Journal of Genetic Psychology*, 143, 175-179. <https://doi.org/10.1080/00221325.1983.10533549>
- Nelson-Le Gall, S. A. (1985). Motive-outcome matching and outcome foreseeability: Effects on attribution of intentionality and moral judgments. *Developmental Psychology*, 21, 332–337. <https://doi.org/10.1037/0012-1649.21.2.332>
- Nelson, S. A. (1980). Factors influencing young children's use of motives and outcomes as moral criteria. *Child Development*, 51, 823-829. <https://www.jstor.org/stable/1129470>
- NICHD Early Child Care Research Network. (2001). Child care and children's peer interaction at 24 and 36 months: The NICHD study of early child care. *Child Development*, 72, 1478-1500. <https://doi.org/10.1111/1467-8624.00361>
- Nichols, S. R., Svetlova, M., & Brownell, C. A. (2015). Toddlers' responses to infants' negative emotions. *Infancy*, 20, 70-97. <https://doi.org/10.1111/infa.12066>
- Nobes, G., & Martin, J. W. (2022). They should have known better: The roles of negligence and outcome in moral judgements of accidental actions. *British Journal of Psychology*, 113, 370-395. <https://doi.org/10.1111/bjop.12536>
- Nobes, G., Panagiotaki, G., & Bartholomew, K. J. (2016). The influence of intention, outcome and question-wording on children's and adults' moral judgments. *Cognition*, 157, 190-204. <https://doi.org/10.1016/j.cognition.2016.08.019>

- Nobes, G., Panagiotaki, G., & Engelhardt, P. E. (2017). The development of intention-based morality: The influence of intention salience and recency, negligence, and outcome on children's and adults' judgments. *Developmental Psychology, 53*, 1895-1911.
<https://doi.org/10.1037/dev0000380>
- Nobes, G., Panagiotaki, G., & Pawson, C. (2009). The influence of negligence, intention, and outcome on children's moral judgments. *Journal of Experimental Child Psychology, 104*, 382-397. <https://doi.org/10.1016/j.jecp.2009.08.001>
- Nucci, L. P., & Turiel, E. (1978). Social interactions and the development of social concepts in preschool children. *Child Development, 49*, 400-407.
<http://www.jstor.org/stable/11287>
- Ochoa, K. D., Rodini, J. F., & Moses, L. J. (2022). False belief understanding and moral judgment in young children. *Developmental Psychology, 58*, 2022-2035.
<https://doi.org/10.1037/dev0001411>
- Onishi, K. H., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science, 308*, 255-258. <https://doi.org/10.1126/science.1107621>
- Perner, J., & Roessler, J. (2012). From infants' to children's appreciation of belief. *Trends in Cognitive Sciences, 16*, 519-525. <https://doi.org/10.1016/j.tics.2012.08.004>
- Perner, J., Ruffman, T., & Leekam, S. R. (1994). Theory of mind is contagious: you catch it from your sibs. *Child Development, 65*, 1228-1238. <https://doi.org/10.1111/j.1467-8624.1994.tb00814.x>
- Perra, O., Paine, A. L., & Hay, D. F. (2021). Continuity and change in anger and aggressiveness from infancy to childhood: The protective effects of positive parenting. *Development and Psychopathology, 33*, 937-956. <https://doi.org/10.1017/s0954579420000243>
- Piaget, J. (1932). *Le jugement moral chez l'enfant*. Paris: Presses Universitaires de France.

- Piaget, J. (1941). Le mécanisme du développement mental et les lois du groupement des opérations. *Archives de Psychologie*, 28, 215-285.
- Premack, D., & Premack, A. J. (1997). Infants attribute value± to the goal-directed actions of self-propelled objects. *Journal of Cognitive Neuroscience*, 9, 848-856.
<https://doi.org/10.1162/jocn.1997.9.6.848>
- Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, 1, 515-526.
<https://doi.org/10.1017/S0140525X00076512>
- Proft, M., & Rakoczy, H. (2019). The ontogeny of intent-based normative judgments. *Developmental Science*, 22, e12728. <https://doi.org/10.1111/desc.12728>
- Pun, A., Birch, S. A., & Baron, A. S. (2021). The power of allies: Infants' expectations of social obligations during intergroup conflict. *Cognition*, 211, 104630.
<https://doi.org/10.1016/j.cognition.2021.104630>
- Raihani, N. J., & Barclay, P. (2016). Exploring the trade-off between quality and fairness in human partner choice. *Royal Society Open Science*, 3, 160510.
<https://doi.org/10.1098/rsos.160510>
- Rand, D. G. (2016). Cooperation, fast and slow: Meta-analytic evidence for a theory of social heuristics and self-interested deliberation. *Psychological Science*, 27, 1192-1206.
<https://doi.org/10.1177/0956797616654455>
- Reschke, P., Walle, E., Flom, R., & Guenther, D. (2017). Twelve-monthold infants' sensitivity to others' emotions following positive and negative events. *Infancy*, 22, 874-881.
<http://dx.doi.org/10.1111/inf.12193>
- Rhodes, M., Hetherington, C., Brink, K., & Wellman, H. M. (2015). Infants' use of social partnerships to predict behavior. *Developmental Science*, 18, 909-916.

<https://doi.org/10.1111/desc.12267>

Rogoff, B., Dahl, A., & Callanan, M. (2018). The importance of understanding children's lived experience. *Developmental Review, 50*, 5-15. <https://doi.org/10.1016/j.dr.2018.05.006>

Rosset, E., & Rottman, J. (2014). The big 'whoops!' in the study of intentional behavior: An appeal for a new framework in understanding human actions. *Journal of Cognition and Culture, 14*, 27-39. <https://doi.org/10.1163/15685373-12342108>

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. <https://doi.org/10.1016/j.infbeh.2011.04.007>

Ruba, A. L., Meltzoff, A. N., & Repacholi, B. M. (2019). How do you feel? Preverbal infants match negative emotions to events. *Developmental Psychology, 55*, 1138. <http://dx.doi.org/10.1037/dev0000711>

Ruba, A. L., Meltzoff, A. N., & Repacholi, B. M. (2020). The development of negative event-emotion matching in infancy: Implications for theories in affective science. *Affective Science, 1*, 4-19. <https://doi.org/10.1007/s42761-020-00005-x>

Ruba, A. L., & Pollak, S. D. (2020). The development of emotion reasoning in infancy and early childhood. *Annual Review of Developmental Psychology, 2*, 503-531. <https://doi.org/10.1146/annurev-devpsych-060320-102556>

Rubin, K. H., & Chung, O. B. (2013). *Parenting beliefs, behaviors, and parent-child relations: A cross-cultural perspective*. Psychology Press.

Sagi, A., & Hoffman, M. L. (1976). Empathic distress in the newborn. *Developmental Psychology, 12*, 175-176. <https://doi.org/10.1037/0012-1649.12.2.175>

Salthouse, T. A., Atkinson, T. M., & Berish, D. E. (2003). Executive functioning as a potential mediator of age-related cognitive decline in normal adults. *Journal of Experimental*

- Psychology: General*, 132, 566-594. <https://doi.org/10.1037/0096-3445.132.4.566>
- Scott, R. M., & Baillargeon, R. (2017). Early false-belief understanding. *Trends in Cognitive Sciences*, 21, 237-249. <https://doi.org/10.1016/j.tics.2017.01.012>
- Sebastian, C. L., Fontaine, N. M., Bird, G., Blakemore, S. J., De Brito, S. A., McCrory, E. J., & Viding, E. (2012). Neural processing associated with cognitive and affective Theory of Mind in adolescents and adults. *Social Cognitive and Affective Neuroscience*, 7, 53-63. <https://doi.org/10.1093/scan/nsr023>
- Setoh, P., Scott, R. M., & Baillargeon, R. (2016). Two-and-a-half-year-olds succeed at a traditional false-belief task with reduced processing demands. *Proceedings of the National Academy of Sciences*, 113, 13360-13365. <https://doi.org/10.1073/pnas.1609203113>
- Shamay-Tsoory, S. G. (2011). The neural bases for empathy. *Neuroscientist*, 17, 18–24. <https://doi.org/10.1177/1073858410379268>
- Shamay-tsoory, S. G., & Aharon-peretz, J. (2007). Dissociable prefrontal networks for cognitive and affective theory of mind: A lesion study. *Neuropsychologia*, 45, 3054-3067. <https://doi.org/10.1016/j.neuropsychologia.2007.05.021>
- Shamay-Tsoory, S. G., Aharon-Peretz, J., & Perry, D. (2009). Two systems for empathy: A double dissociation between emotional and cognitive empathy in inferior frontal gyrus versus ventromedial prefrontal lesions. *Brain*, 132, 617-627. <https://doi.org/10.1093/brain/awn279>
- Skerry, A. E., & Spelke, E. S. (2014). Preverbal infants identify emotional reactions that are incongruent with goal outcomes. *Cognition*, 130, 204-216. <https://doi.org/10.1016/j.cognition.2013.11.002>
- Simner, M. L. (1971). Newborn's response to the cry of another infant. *Developmental*

- Psychology*, 5, 136–150. <https://doi.org/10.1037/h0031066>
- Sloane, S., Baillargeon, R., & Premack, D. (2012). Do infants have a sense of fairness? *Psychological Science*, 23, 196-204. <https://doi.org/10.1177/0956797611422072>
- Smetana, J. G. (1984). Toddlers' social interactions regarding moral and conventional transgressions. *Child Development*, 55, 1767-1776.
<http://www.jstor.org/stable/1129924>
- Smith, A. (1759/1982). *The theory of moral sentiments*. Liberty Classics.
- Smith, C. E., Noh, J. Y., Rizzo, M. T., & Harris, P. L. (2017). When and why parents prompt their children to apologize: The roles of transgression type and parenting style. *Journal of Family Studies*, 23, 38-61. <https://doi.org/10.1080/13229400.2016.1176588>
- Sousa, P., & Piazza, J. (2014). Harmful transgressions qua moral transgressions: A deflationary view. *Thinking & Reasoning*, 20, 99-128.
<https://doi.org/10.1080/13546783.2013.834845>
- Southgate, V., Chevallier, C., & Csibra, G. (2010). Seventeen-month-olds appeal to false beliefs to interpret others' referential communication. *Developmental Science*, 13, 907-912. <https://doi.org/10.1111/j.1467-7687.2009.00946.x>
- Stavans, M., & Baillargeon, R. (2019). Infants expect leaders to right wrongs. *Proceedings of the National Academy of Sciences*, 116, 16292-16301.
<https://doi.org/10.1073/pnas.1820091116>
- Steckler, C. M., Liberman, Z., Van de Vondervoort, J. W., Slevinsky, J., Le, D. T., & Hamlin, J. K. (2018). Feeling out a link between feeling and infant sociomoral evaluation. *British Journal of Developmental Psychology*, 36, 482-500.
<https://doi.org/10.1111/bjdp.12232>
- Stratton-Lake, P. (2014). Intuitionism in ethics. In E. Zalta (Ed.), *Stanford Encyclopedia of*

- Philosophy*. <https://plato.stanford.edu/entries/intuitionismethics/>
- Strid, K., & Meristo, M. (2020). Infants consider the distributor's intentions in resource allocation. *Frontiers in Psychology, 11*, 596213.
<https://doi.org/10.3389/fpsyg.2020.596213>
- Surian, L., & Margoni, F. (2020). First steps toward an understanding of procedural fairness. *Developmental Science, 23*, e12939. <https://doi.org/10.1111/desc.12939>
- Tasimi, A. (2020). Connecting the dots on the origins of social knowledge. *Perspectives on Psychological Science, 15*, 397-410. <https://doi.org/10.1177/1745691619885861>
- Tatone, D., Schlingloff-Nemecz, L., & Pomiechowska, B. (2023). Infants do not use payoff information to infer individual goals in joint-action events. *Cognitive Development, 66*, 101329. <https://doi.org/10.1016/j.cogdev.2023.101329>
- Thomas, A. J. (2024). *Cognitive representations of social relationships and their developmental origins*. Pre-print retrieved at <https://doi.org/10.31234/osf.io/xhrfu>
- Thomas, A. J., Thomsen, L., Lukowski, A. F., Abramyan, M., & Sarnecka, B. W. (2018). Toddlers prefer those who win but not when they win by force. *Nature Human Behaviour, 2*, 662-669. <https://doi.org/10.1038/s41562-018-0415-3>
- Thomsen, L. (2020). The developmental origins of social hierarchy: How infants and young children mentally represent and respond to power and status. *Current Opinion in Psychology, 33*, 201-208. <https://doi.org/10.1016/j.copsyc.2019.07.044>
- Thomsen, L., & Carey, S. (2013). Core cognition of relational models. In M. Banajii & S. Gelman (Eds.), *Navigating the social world: What infants, children and other species can teach us* (pp. 17-22). New York: Oxford University Press.
<https://doi.org/10.1093/acprof:oso/9780199890712.003.0004>
- Thomsen, L., Frankenhuys, W. E., Ingold-Smith, M., & Carey, S. (2011). Big and mighty:

Preverbal infants mentally represent social dominance. *Science*, 331, 477-480.

<https://doi.org/10.1126/science.1199198>

Ting, F., & Baillargeon, R. (2021). Toddlers draw broad negative inferences from wrongdoers' moral violations. *Proceedings of the National Academy of Sciences*, 118, e2109045118.

<https://doi.org/10.1073/pnas.2109045118>

Ting, F., Dawkins, M. B., Stavans, M., & Baillargeon, R. (2020). Principles and concepts in early moral cognition. In J. Decety (Ed.), *The social brain: A developmental perspective* (pp. 41–65). MIT Press.

Tousignant, B., Eugène, F., & Jackson, P. L. (2017). A developmental perspective on the neural bases of human empathy. *Infant Behavior and Development*, 48, 5-12.

<https://doi.org/10.1016/j.infbeh.2015.11.006>

Tremblay, R. E. (2018). Développement des agressions physiques de la petite enfance à l'âge adulte. *Enfance*, 3, 407-419. <https://doi.org/10.3917/enf2.183.0407>

Tremblay, R. E., Japel, C., Perusse, D., McDuff, P., Boivin, M., Zoccolillo, M., & Montplaisir, J. (1999). The search for the age of 'onset' of physical aggression: Rousseau and Bandura revisited. *Criminal Behaviour and Mental Health*, 9, 8-23.

<https://doi.org/10.1002/cbm.288>

Uzefovsky, F., Paz, Y., & Davidov, M. (2020). Young infants are pro-victims, but it depends on the context. *British Journal of Psychology*, 111, 322-334.

<https://doi.org/10.1111/bjop.12402>

Vaish, A. (2016). Flexible concern: The development of multidetermined and context-dependent empathic responding. *Child Development Perspectives*, 10, 149-154.

<https://doi.org/10.1111/cdep.12178>

Vaish, A., Carpenter, M., & Tomasello, M. (2009). Sympathy through affective perspective

- taking and its relation to prosocial behavior in toddlers. *Developmental Psychology*, 45, 534-543. <https://doi.org/10.1037/a0014322>
- Vaish, A., Carpenter, M., & Tomasello, M. (2010). Young children selectively avoid helping people with harmful intentions. *Child Development*, 81, 1661-1669. <https://doi.org/10.1111/j.1467-8624.2010.01500.x>
- Van de Vondervoort, J. W., & Hamlin, J. K. (2018). Preschoolers focus on others' intentions when forming sociomoral judgments. *Frontiers in Psychology*, 9, 404965. <https://doi.org/10.3389/fpsyg.2018.01851>
- Warneken, F., & Tomasello, M. (2006). Altruistic helping in human infants and young chimpanzees. *Science*, 311, 1301-1303. <https://doi.org/10.1126/science.1121448>
- Weisberg, D. S., & Leslie, A. M. (2012). The role of victims' emotions in preschoolers' moral judgments. *Review of Philosophy and Psychology*, 3, 439-455. <https://doi.org/10.1007/s13164-012-0101-8>
- Wellman, H. M., & Liu, D. (2004). Scaling of theory-of-mind tasks. *Child Development*, 75, 523-541. <https://doi.org/10.1111/j.1467-8624.2004.00691.x>
- Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, 13, 103-128. [https://doi.org/10.1016/0010-0277\(83\)90004-5](https://doi.org/10.1016/0010-0277(83)90004-5)
- Woo, B. M., & Spelke, E. S. (2023). Toddlers' social evaluations of agents who act on false beliefs. *Developmental Science*, 26, e13314. <https://doi.org/10.1111/desc.13314>
- Woo, B. M., Steckler, C. M., Le, D. T., & Hamlin, J. K. (2017). Social evaluation of intentional, truly accidental, and negligently accidental helpers and harmers by 10-month-old infants. *Cognition*, 168, 154-163. <https://doi.org/10.1016/j.cognition.2017.06.029>
- Woo, B. M., Tan, E., & Hamlin, J. K. (2022). Human morality is based on an early-emerging

moral core. *Annual Review of Developmental Psychology*, 4, 41-61.

<https://doi.org/10.1146/annurev-devpsych-121020-023312>

Woo, B. M., Tan, E., Yuen, F. L., & Hamlin, J. K. (2023). Socially evaluative contexts facilitate mentalizing. *Trends in Cognitive Sciences*, 27, 17-29.

<https://doi.org/10.1016/j.tics.2022.10.003>

Woodward, A. L. (1998). Infants selectively encode the goal object of an actor's reach.

Cognition, 69, 1-34. [https://doi.org/10.1016/s0010-0277\(98\)00058-4](https://doi.org/10.1016/s0010-0277(98)00058-4)

Young, L., Cushman, F., Hauser, M., & Saxe, R. (2007). The neural basis of the interaction between theory of mind and moral judgment. *Proceedings of the National Academy of Sciences*, 104, 8235-8240. <https://doi.org/10.1073/pnas.0701408104>

Young, L., & Saxe, R. (2009a). Innocent intentions: A correlation between forgiveness for accidental harm and neural activity. *Neuropsychologia*, 47, 2065-2072.

<https://doi.org/10.1016/j.neuropsychologia.2009.03.020>

Young, L., & Saxe, R. (2009b). An fMRI investigation of spontaneous mental state inference for moral judgment. *Journal of Cognitive Neuroscience*, 21, 1396-1405.

<https://doi.org/10.1162/jocn.2009.21137>

Young, L., Scholz, J., & Saxe, R. (2011). Neural evidence for “intuitive prosecution”: The use of mental state information for negative moral verdicts. *Social Neuroscience*, 6, 302-315. <https://doi.org/10.1080/17470919.2010.529712>

Young, L., Koenigs, M., Kruepke, M., & Newman, J. P. (2012). Psychopathy increases perceived moral permissibility of accidents. *Journal of Abnormal Psychology*, 121, 659-667. <https://doi.org/10.1037/a0027489>

- Yuill, N., & Perner, J. (1988). Intentionality and knowledge in children's judgments of actor's responsibility and recipient's emotional reaction. *Developmental Psychology, 24*, 358-365. <https://doi.org/10.1037//0012-1649.24.3.358>
- Zahn-Waxler, C., Radke-Yarrow, M., & King, R. A. (1979). Child rearing and children's prosocial initiations toward victims of distress. *Child Development, 50*, 319-330. <https://doi.org/10.2307/1129406>
- Zelazo, P. D., Helwig, C. C., & Lau, A. (1996). Intention, act, and outcome in behavioral prediction and moral judgment. *Child Development, 67*, 2478-2492. <https://doi.org/10.1111/j.1467-8624.1996.tb01869.x>
- Ziv, T., Whiteman, J. D., & Sommerville, J. A. (2021). Toddlers' interventions toward fair and unfair individuals. *Cognition, 214*, 104781. <https://doi.org/10.1016/j.cognition.2021.104781>