UiT The Arctic University of Norway

Faculty of Health Sciences – Department of Psychology (IPS)

Teleological Stance: A Systematic Review and Meta-Analysis

Infants’ expectation of efficiency in goal-directed actions.

Håkon Wik

Thesis in Cand.Psychol., PSY-2901, December 2021.

Teleological Stance: A Systematic Review and Meta-Analysis

Infants' expectation of efficiency in goal-directed action

Teleologisk holdning: En systematisk litteraturgjennomgang og metaanalyse

Spedbarns forventing om effektivitet av målrettet handling

Author: Håkon Wik

Supervisor: Mikołaj Hernik

PSY-2901

Thesis in Cand.Psychol.

Faculty of Health Sciences

Department of Psychology (IPS)

UiT The Arctic University of Norway

December 2021

Preface

The scientific field of developmental psychology has fascinated me with the use of simple studies which says more than a thousand words. More specifically the studies with infants' participants who can't speak for themselves. Yet the use of a simple methodological techniques (such as looking time studies) allows us to begin to understand what infants are communicating to us. By the inspiration of my supervisor, this led me to the studies of infants' expectation about efficiency of action. I remembered Mikołaj Hernik lecturing on this topic at my seventh semester, back in Autumn 2019. The teleological stance theory interested me as a potential precursor for the emerging mentalistic ability, known as theory of mind.

The supervisor, Mikołaj Hernik formulated the topic for this thesis. He has taught me how to implement a systematic review and code data for the meta-analysis. Through his guidance this journey has become a manageable experience. He has made himself available by giving simple and detailed answers to the complications of the scientific process. I am grateful for the effort which he put into the quality of the thesis. Through this eleventh semester he has read through a manifold of drafts to give feedback on the ongoing work.

The systematic review and meta-analysis of the thesis was written independently, but with guidance and feedback from the supervisor. I was responsible for the collection, reading and writing about the papers for the review. The supervisor who has knowledge of the literature checked for any omissions. The coding of data was carried out independently but checked carefully by the supervisor for errors or misconceptions. The supervisor proposed which analyzes were appropriate, but they were conducted, presented, and discussed by me.

Abstract

Background

Teleological stance is also known as infants' "naïve theory of rational action" (Csibra et al., 1999). It has been studied in a violation-of-expectation paradigm assessing difference in infants' looking time to events consistent and inconsistent with the expectation that agents act efficiently. The aim of this thesis is to estimate the size of this effect in experimental and control conditions of the paradigm through meta-analysis, and to interpret the findings in the light of the systematic review.

Methods

Search was carried out at Google Scholar database in May 2021. Studies included habituation or familiarization phase of efficient or inefficient agent, followed by a test phase measuring looking time at efficient vs. inefficient action in new situation.

Findings

Total of 15 papers involving 3- to 15-month-old infants (n = 1020) was included. The 36 experimental conditions yielded a small effect size ($d = 0.43$, $p = <.001$), variance explained by age ($z = 2.10$, $p = .036$), but remaining excess variance ($Q\ (25) = 78.6$, $p = <.001$). Null finding was found for the 17 control conditions ($d = -0.074$, $p = .326$).

Interpretation

The meta-analysis confirms infants' expectation of efficiency in agents. Infants are also likely to abandon this expectation if agent is previously inefficient. Exploratory analysis supports a presence of a developmental milestone in the last part of the first year of life. Evidence of publication bias challenge a non-linear trend, and inclusion of new or unpublished studies is needed to assess it with greater certainty.

Keywords: teleological stance, goal attribution, infancy, systematic review, meta-analysis

Introduction

“Social cognition refers to the ability to understand other people” (Striano & Reid, 2006). A developmental milestone in social cognition is children sophisticated use of mentalization, which emerges by the fifth year of life (Wellman et al., 2001). This implies the ability to represent others mental states such as intention, desires, and beliefs (Wimmer & Perner, 1983). A common assumption has been that younger children may think about others intention and desire but are not able to think in terms of their beliefs (Keil, 2013). Wimmer and Perner (1983) found that when children at 3.5 years of age acquire knowledge from what only they saw, they fail to understand that others might not know this as well. However, the understanding that others have different beliefs have been found in 15-month-old when using a non-verbal experiment (Onishi & Baillargeon, 2005). The discussion on how and when a mentalistic understanding emerges (often referred to as Theory of Mind) remains an open discussion in the literature.

This review will discuss accounts of social cognition skills that precede the emerging mentalization ability. The first is a mentalistic account which suggest that the interpretation of others mental states develop continually. While a well-develop mentalization ability allows for sophisticated reasoning, an earlier stage in development could initially be more restricted to the concept of intentionality. The literature used in this review has labelled such an account as the intentional stance theory. Tomasello (1995) propose that infants undergo a major developmental change already before the age of one, which he dubbed the “social-cognitive revolution”. At around 9-month of age, infants go from a primarily face-to-face (dyadic) interaction with their caregiver, to then start engaging in a person-object-person (triadic) interaction with the outside world (Striano & Reid, 2006). They engage in shared attention and acquire an understanding of communicative gestures such as pointing. Joint attention has been offered as the first plausible sign of a mentalistic attribution (Johnson, 2000). Tomasello

(1995) suggest that this early ability reflects an emerging understanding of agents as intentional. This account implies that infants acquire a simpler mentalization ability before developing a more sophisticated social cognitive understanding

The second is a non-mentalistic account, which the reviewed literature has labelled as the teleological stance theory. Gergely et al. (1997) propose it as a more general (teleological) interpretation system, were infants' reason according to representational elements accessible before mentalization. Infants already have experience with action, goal, and constraint before learning about others intention (to act), desire (for goal), and belief (about constraint). Teleological stance theory suggest that infants can arrive at the same interpretation by using their own non-mentalistic thinking (Griffin & Baron-Cohen, 2002). Mentalistic attribution should not be necessary for all forms of action prediction because most beliefs tend to be true, and behavior is usually performed efficiently. Instead, infants' reasoning might reflect a naïve theory of rational action, that only later in development becomes associated with intentional action (Csibra et al., 1999).

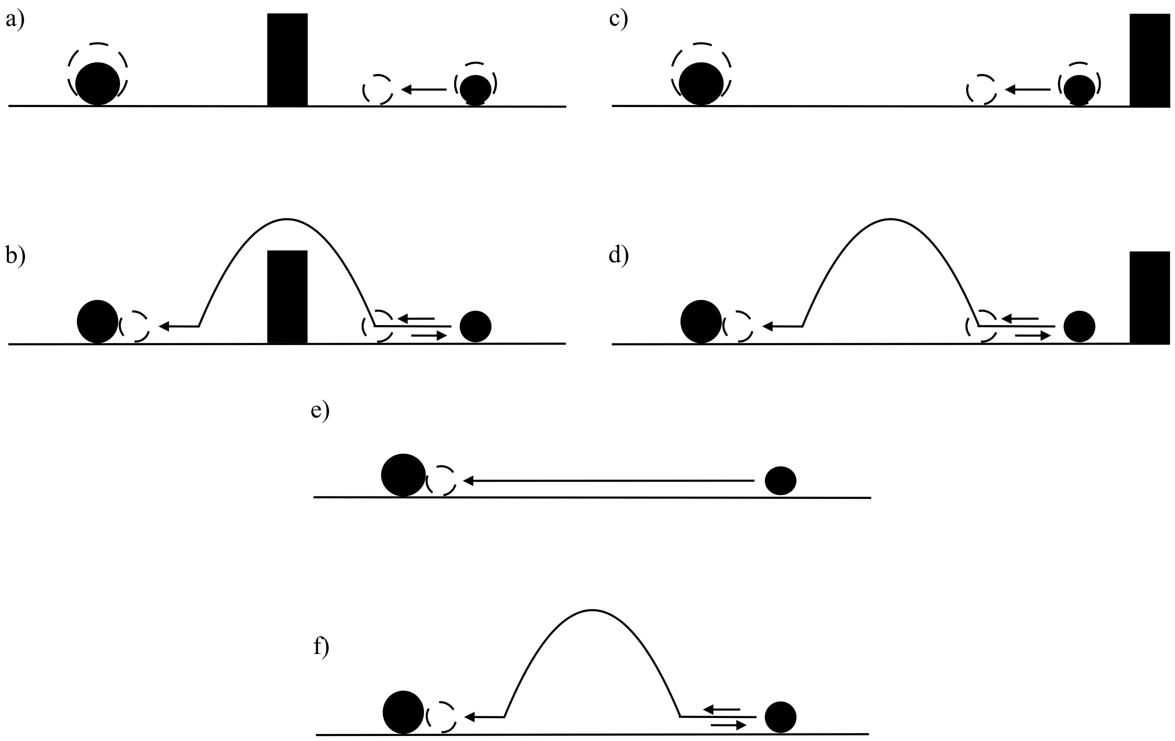
In this review I will overview the literature of teleological stance theory, as well as its relation to the intentional stance theory. The overview will offer four major topic which will be addressed: 1) Does goal attribution require beliefs about mental states or a non-mentalistic interpretation? 2) What cues are used for goal attribution? 3) When does goal attribution emerge and what is the foundation for it? 4) How are agents' inefficient action interpretated? I will start of by presenting the experiment which created an arena for these discussions.

Interpreting goal-directed action

Imagine seeing a couple of ball shaped creatures (discs). The larger ball suddenly expands, and then contracts to its original size. The same expansion-contraction is performed by the little ball. The little one moves towards the larger ball, stops next to an obstacle (rectangle), and return to its original position. Finally, the little ball jumps over the obstacle and continues forward until it contacts the larger ball. They ones again repeat the expansion-contraction sequence. Heider and Simmel (1944) found that adult typically attribute mentalistic interpretation to geometric figures demonstrating animate characteristics. Gergely et al. (1995) observed that adults did the same with the presented scenario of the ball shaped creatures. Perhaps the little ball is hesitant to pass the obstacle, and the larger ball is encouraging it to jump over? How will an infant interpret the same scenario?

Figure 1.

Illustrations of Gergely et al. (1995) rational and non-rational approach conditions.



Note. Experimental condition (rational) habituation phase is showing a little ball moving towards a barrier (a). The ball returns to its original position, moves again to jump over the barrier, and then contacts the larger ball (b). The (non-rational) habituation phase of the control condition shows a similar action (c & d), but the barrier doesn't form an obstacle as it is placed behind the ball. Test phase for both condition shows the little ball taking a new straight path (e) or the familiar jumping action (f) in the absence of any barriers.

Presenting the experiment

Gergely et al. (1995) developed this experiment to figure out how infants interpret goal-directed action. The habituation study exposed 12-month-old infants to the stimulus of the ball-shaped creatures. The study included an experimental (rational approach) condition and a control (non-rational approach) condition. They only differed in that the habituation phase of the control condition did not place the rectangle as an obstacle between the two balls. In the following test phase, the rectangle was not present in either condition. Both conditions either saw the little ball continuing the jumping action or changing to taking a straight path. Which of these stimuli will infants dishabituate more to? Infants generally dishabituate more to a novel stimulus (Oakes, 2010). However, this study found that infants looked longer at the familiar movement. Gergely et al. (1995) suggest that the looking time data reflects a surprise in the infants for the old action, as it became inefficient ones the obstacle was removed. The finding is used as evidence for infants' ability to reason about the efficiency of goal-directed action in agents. Gergely et al. (1995) hypothesized that infants have a theory of agency which assumes efficiency of goal-directed action. In comparison, the results from the control condition showed no differentiation in looking time between the stimuli. The null finding is surprising, as infants usually dishabituate more to the novel stimulus. The lack of effect could possibly be caused by the small sample size. However, Gergely et al. (1995) suggest that infants in the control group abandon action prediction as it did not meet the requirement of efficiency. Infants reasoning about the efficiency could explain why a novelty effect was not found. The finding supports the hypothesis that infants apply a rationality principle to observed actions of efficient agents.

Intentional interpretation

At the time, Gergely et al. (1995) were in favor of the intentional stance theory. The intentional stance suggests that infants attribute intentionality to the agent. Dennett (1989) explains it as a strategy for behavior prediction, where you treat the object of interest as a rational agent with beliefs and desires. Gergely et al. (1995) conclusion implies that infants acquire a simpler form of a mentalization ability, even earlier than suggested by Onishi and Baillargeon (2005). However, the interpretation of the experiment has been criticized for not reflecting a mentalistic ability. The results indicate an expectancy about motion path, but this is not the same as attributing intention (Griffin & Baron-Cohen, 2002). Either way, the finding opened a new discussion on how infants attribute goals to agents.

The revised experiment

Gergely and Csibra (1997) acknowledge the criticism for suggesting that 12-month-old infants are taking the intentional stance, and thus clarified that it could equally be explained by a non-mentalistic interpretation of goal-directed action. Gergely et al. (1997) and Csibra et al. (1999) developed a revised experiment to test if infants' goal-attribution could initially be based on a lower-level interpretations. Infants at 9- and 12-month of age were exposed to a similar habituation procedure as those in the study by Gergely et al. (1995), but now in the absence of cues that could indicate agency. The study included an experimental (efficient) and control (inefficient) condition. Infants in both conditions were exposed to the little ball flying over a rectangle towards the larger ball. Infants could not establish whether this curvilinear pathway was caused by self-propelled movement or whether the observed object was set in motion by an external force (e.g., thrown). In the test phase, the rectangular obstacle was removed. Infants in the experimental group dishabituated more to the familiar curved pattern, compared to when taking a new action on a straight path. This gives further support for infants applying a rationality principle for goal attribution. In comparison, the control group was habituated to the same action, but the rectangle allowed for the ball to pass under it as well. Infants did not form a specific expectation for the inefficient ball in the new situation. This experiment replicates Gergely et al. (1995), but this time in even younger infants and without the reliance on agency cues. The finding suggests a more generalized interpretation of goal-directed action, which is not restricted to objects perceived as animate.

Teleological interpretation

The teleological stance theory proposes that infants' reason teleologically by establishing a relation among the three representational elements: the action (jumping), the goal (larger ball), and the constraint (rectangle) (Gergely et al., 1997). Infants' psychological framework relies on the rationality principle to understand goal-directed action, which assumes efficiency of action given the situational constraint. Since infants have an early understanding that objects are incapable of passing through rigid elements (Baillargeon, 1987), a rational approach should be to maneuver around the obstacle. The action performed follows an equifinal outcome, which should provide the infants with the goal of the action. By acquiring the relation between the goal and the current constraint, infants infer a likely new action by applying the rationality principle. However, if the action performed is deemed irrational, infants do not form an expectation in the new situation. The theory suggest that the fundamentals of goal-directed reasoning is based on a non-mentalistic interpretation, only to later develop into the attribution of intentional action (Csibra et al. 1999).

The interferential principle

Csibra et al. (2003) note that the rationality principle also acts as an interferential principle. A teleological representation makes is possible to infer the likely content of any of the three elements: the action, the goal, or the constraint. In the original experiment, Gergely et al. (1995) demonstrated that infants expect a new action based on the known goal and the changed constraints. In a modified version of this experiment, Csibra et al. (2003) included an occluder that covered the infants view of a potential obstacle which the agent jumped over. Infants at 12-month of age dishabituated more to the jumping action when the removal of the occlude revealed no obstacle. This indicates that infants rationalized the jumping action by inferring a likely constraint to the goal-directed action. In another habituation study by Csibra et al. (2003), 12-month-old infants also applied a teleological representation without seeing the end goal. This indicates that infants established a relation between the action and the constraint to infer a likely goal. What is interesting is that the 9-month-old infants in both experimental groups did not display a similar looking pattern. Csibra et al. (2003) suggest that the youngest infants might lack the hypothesis formation processes for reasoning without a well-formed teleological representation (action, goal, constraint).

A relevant question is whether infants can infer goals to failed goal-directed action. Csibra et al. (2003) found that infants have an expectation for goal attribution to end in contact. So how will infants interpret goal-directed action when it ends short of achieving the goal? Brandone and Wellman (2009) investigated this by habituating 8-, 10-, and 12-month-old infants to a human reaching over a barrier to grasp a ball. Infants at 10- and 12-month dishabituated more to the familiar indirect action without the barrier, both in a condition where the agent was successful or failed to achieve the goal. This indicates that infants as young as 10-month of age understand failed goal-directed action, possibly by attributing intention in this incomplete teleological context. Infants at 8-month of age attributed goals as well, but only to the previously successful agent. The results indicates that infants are not initially reliant on the attributing intention to understand goal-directed action. Rather, Brandone and Wellman (2009) suggest that this emerging ability for mentalization might be based on a fundamental non-mentalistic interpretation of goal-directed action. The finding is consistent with the idea of a transition between the teleological and intentional stance theory.

Cues for goal attribution

The first conducted studies on infants' goal attribution have been using two-dimensional computer animation. A question is whether infants generalize this effect in more diverse scenarios? The finding by Gergely et al. (1995; 1997) turns out to replicate well in 12-

month-old infants when using puppets, real humans (Sodian et al., 2004), or hands (Phillips & Wellman, 2005). Perhaps some featural or behavioral cues could positively affect infants' likelihood for attribution goals, such as familiarity with human action. In contrast, Csibra (2008) have questioned whether younger infants fail to attribute goals in two-dimensional studies due to the lack of rich features. This claim is based on Kamewari et al. (2005) modified experiment which used three different conditions involving: human, humanoid robot, and a moving box. Infants at 6.5-month of age interpreted human and humanoid robot as goal-directed, but not for the moving box. This is an unexpected finding, as Csibra et al. (1999) previously failed to demonstrate goal attribution in 6-month-olds. The youngest infants either fails to establish a teleological relation in a 2D computer animation, or teleological reasoning could initially be restricted to human-like agents.

In a replication study, Csibra (2008) found that 6.5-month-old infants also were able to attribute goals to a moving box when it made both right and left detours. Csibra (2008) suggest that behavioral variability or some degree of unpredictability enables infants to attribute goals to unfamiliar objects. In comparison, the stimuli in Gergely et al. (1995; 1997) experiment varied in horizonal direction or height of jump required to pass the obstacle. But movement variation was not present in Kamewari et al. (2005) humanoid robot condition. This could mean that another factor allowed the infants to attribute goals. It is possible that infants attribute goals in the absence of behavioral variability when the agent resemble a familiar agent (i.e., human). This finding shows that 6.5-month-old infants' form teleological inference but are to a greater extent influenced by cues of agency.

Biro et al. (2007) proposed three possible functional relation between agency and goal-directedness. The "mandatory link" hypothesis suggests a required identification of agency for interpreting goal-directed action. This hypothesis is unlikely as infants seem not to rely on previous cues of agency, such as self-propelled movement (Gergely et al., 1997). The "exclusive link" hypothesis precludes goal-directed interpretation when there is evidence for the object as non-agent. Biro et al. (2007) experiment tested this by including a third ball, which caused the little ball to launch from contact. Infants at 12-month of age still attributed goals after receiving counter evidence of the movement as self-propelled. This finding makes the "exclusive link" hypothesis unlikely as well. The "probabilistic link" hypothesis assumes that agent categorization biases towards goal-directed action interpretation. Biro et al. (2007) experiment found that infants are less likely to attribute goals when there is a delayed launching movement after contact with the third ball. This indicates that infants differentiate between action that is self-propelled or externally caused. The findings from the two experiment show that agency cues (e.g., self-propelled movement) are indeed not necessary but are likely to bias infants towards attributing goals. The next topic goes into more details on how the effect of familiarity relates to teleological reasoning.

Foundation of the emerging goal attribution

A more recent topic in the literature has been dealing with the question whether teleological reasoning is based on first-person experience, or prior knowledge of others goal-directed action (Skerry et al., 2013). This involves the ability to attribute goals to unfamiliar action, or whether infants own experience is a prerequisite for this. Skerry et al. (2013) and Liu et al. (2019) investigated this in 3.5-month-old (pre-grasping) infants' by exposing them to the scenario of a human grasping a ball over a barrier. Skerry et al. (2013) experiment included a pre phase where infants were equipped with Velcro-gloves, regular gloves, or received no training picking up a ball. This allowed only the first group to be able to successfully lift the ball (also covered in Velcro). In the following habituation and test phase, only the infants who received first-person experience lifting the ball, dishabituated more to the agents' indirect reach in the absence of the barrier. The finding is surprising as it shows evidence for goal-attribution in infants as young as 3.5-month of age, but also suggest a greater influence of first-person experience. Lui et al. (2019) experiment tested whether the 3.5-month-old infants could have been impacted by the agent wearing gloves (as in Skerry et al., 2013). By exposing infants to a similar experiment without the pre phase, Lui et al. (2019) found that the infants would attribute goals to the agent, but not those wearing gloves.

This finding shows that like older infants (Brandone & Wellman, 2009; Phillips & Wellman, 2005), 3.5-month-olds dishabituate more to the indirect reach when it became inefficient in the absence of the barrier. The study supports previous findings which demonstrates that infants attribute goals to action which they themselves cannot perform yet, such as walking before 6.5-month (Kamewari et al., 2005) and jumping before 12-month of age (Sodian et al., 2004). However, the effect in 3.5-month-old infants is rather weak and inconsistent (Liu et al., 2019). Lui et al. (2019) found that infants where not only more likely to attribute goals to bare hand (rather than gloved), but also when the hand was touching the goal object (rather than grasping it). This suggests that infants goal attribution is biased towards human resemblance and familiarity of action (including first-person experience).

However, Skerry et al. (2013) claims it as unlikely that infants learn about efficiency and constraints based on first-person motoric experience alone. Infants at this age do not receive much (if any) experience with indirect reaching, due to very poor motor control and inability to grasp objects. Rather, Skerry et al. (2013) concludes that infants apply a general assumption of efficiency to action which they interpret as goal directed. Liu et al. (2019) support the claim of a general prerequisite but note that the origin of this prior knowledge remains unknown. It is possible that this ability emerges in fetal development or that infants learn to reason teleologically over the first postnatal months. Whether infants' have an innate ability for goal attribution remains an unanswered topic.

Interpretation of inefficiency

So far, I have stated that the control condition reflects infants' inability to form movement expectancy for inefficient agents in a new situation. This is based on Gergely et al. (1995) hypothesis that infants abandon action prediction due to not meeting the requirement of efficiency. However, it is not well understood how exactly infants interpret the control condition. Note that the infants in the control conditions are not exposed to the well-formed teleological representation as in the experimental condition. The main difference is that the third representation element in a teleological relation (the situation constraints) is not present, as the rectangle figure do not form an obstacle. Thus, the lack of effect could potentially be explained by the incomplete teleological context, rather than inefficiency of action. If infants interpret the context (rather than the efficiency), then they are likely to form an expectancy when an obstacle is added.

Liu and Spelke (2017) constructed a modified control condition to figure out how infants interpret the inefficient agent. In their habituation study, 6-month-old infants were habituated to an inefficient agent performing a high jump in the absence of an obstacle (barrier placed behind the agent). In the test phase of this control condition, infants either saw the same high jump or a slightly less inefficient low jump towards the goal (without the obstacle). The result supported previous finding that infants do not interpret the action of an inefficient agent. In a modified version of this experiment, a barrier was placed as an obstacle first in the test phase. However, the barrier was too small for the high jump to be efficient. Results shows that infants dishabituated more to the agent continuing the high jump, compared to when taking a more efficient low jump. This suggest that infants expect agent to adjust their action for efficiency, even if they so far only saw it act inefficiently. Liu and Spelke (2017) propose that infants' expectation for efficiency is based on an over-hypothesis that all agents minimize the costs of action. This suggest that infants (at least at 6 months of age) have a general expectation for efficiency of goal-directed action, which seems not to make exception for inefficient agents. Based on Liu and Spelke (2017) study, a new question emerges on how infants process the action of inefficient agents? I will further discuss this topic in the following meta-analysis.

Presenting the meta-analysis

In the review I have been through four major topics about infants' goal attribution: 1) Does goal attribution require beliefs about mental states or a non-mentalistic interpretation? 2) What cues are used for goal attribution? 3) When does goal attribution emerge and what is the foundation for it? 4) How are agents' inefficient action interpretated? A meta-analysis on the included studies in this review will present a new way to look at some of these questions.

The papers included in this review have recognized the possibility that infants' goal attribution could initially be based on a non-mentalistic interpretation. They are also open to the idea that goal-directed reasoning gradually integrates mentalistic interpretation, such as intentionality of action. There is not a clear answer to how or when these processes emerge. By analyzing how the effect size changes with age, we could get a clearer picture about infants' developing goal attribution. We can establish when goal attribution is likely to emerge, and whether the age trend is continuous or demonstrate a rapid change through development. This could potentially reveal evidence for a social cognitive milestone in infancy or demonstrate a more gradual integration of cues for more sophisticated reasoning.

The multiple reported studies reviewed in the intro show reliable effects for the results indicating expectation of efficiency in infancy. The main goal of this meta-analysis is to provide meta-analytic estimate of that size of the effect. The use of habituation of looking time generally shows that infants prefer novel or complex stimuli (Oakes, 2010). This implies that infants tend to habituate visually, and thus spend less time looking at the old familiar stimuli. In this review, the habituation studies assess infants' expectation of efficient action by using a violation-of-expectation paradigm (Aslin, 2007). Unlike a novelty effect, the findings show that infants dishabituate more to the familiar motion path as it became inefficient. This effect was not found in the control condition, suggesting that the effect could not be explained by preference for the complex stimuli. The occurring dishabituation for the familiar stimuli (compared to the novel) is used as evidence for goal attribution. This suggests an expectation for efficiency of agents' action, which we attend to find the overall effect size for.

The expected effect for the control condition is less clear. Infants will usually look longer at a novel stimulus (Aslin, 2007). Instead, the null findings in typical control conditions of teleological stance experiments indicate that some other mental processes could take place. Infants might attempt to interpret the control condition but are not able to form a clear action expectancy in the new situation. The analysis should in this case reflect an equal dishabituation for both stimuli (no preferences). Alternatively, the null finding might be caused by the small samples used in the individual studies. If this is the case, then I expect the meta-analysis to reveal a small effect size demonstrating that infants dishabituate more to the novel action. A meta-analysis for the control may help deciding between these alternatives. Based on Liu and Spelke (2017) study using a modified control condition, infants do not seem to rely on previous experience with efficiency (or lack thereof) to interpret goal-directed action. This means that infants still expect efficiency of movement, even if they only saw the agent previously acting inefficient. Based on this finding I have considered three different hypotheses for how infants might interpret the control condition:

1. Infants do not interpret the habituation phase of the control condition in the sense of attribution of goals. Recall how the control group is not exposed to the well-formed teleological context as the third teleological element of constraint is missing. Csibra et al. (2003) found that 9-month-old infants do not form an expectancy for a constraint when a wall covered the view of what the agent jumped over. He suggests that infants younger than 9-month of age might lack the hypothesis formation processes for reasoning without a well-

formed teleological representation (action, goal, constraint). It is unlikely that infants will form a movement expectancy for the control condition when it is not obvious that it can be interpreted teleologically. Rather infants might only perceive the difference between the stimuli without attempting to interpret them in any ways related to goal attribution and action efficiency. If this hypothesis is correct, meta-analysis may reveal a novelty effect as infants do not attribute goals to the incomplete teleological context of the control condition.

2. Infants abandon action prediction when it does not meet the requirement of efficiency (Gergely et al., 1995). If the rationality principle is not met, then infants will not be able to predict a rational action in a new situation. Thus, infants will not from a movement expectancy in the following test phase. The finding from Lui and Spelke (2017) is not easily consistent with this interpretation. The inefficient habituation phase might bias against goal attribution in the test phase without excluding it all together. For instance, how large must a detour be around a constraint for it to be viewed as not goal-directed? It is likely that some criteria are used for abandoning goal attribution due to the action lacking efficiency. If this hypothesis is correct, meta-analysis may reveal a null finding as infants abandon goal attribution due to the inefficient stimuli of the control condition.

3. Infants differentiate their interpretation of the control condition as they develop. Note that older infants in the experimental conditions acquire the ability to reason in more sophisticated ways, and thus do not rely on a well-formed teleological context (action, goal, constraint). While older infants cannot arrive at a clear action expectation because the agent is inefficient (null finding), younger infants may not attempt to analyze the control condition in term of goal attribution (novelty effect). If this hypothesis is correct, a correlation analysis should reveal an age trend effect which gradually disappear with age.

In this meta-analysis I will primarily look at the (1) effect size in the experimental condition, 2) effect size in the control condition, and (3) how the effect sizes changes with age for both conditions. I will relate the findings to the topics discussed in the presented review.

Method

The Meta-Analysis followed the guidelines of the "Preferred Reporting Items for Systematic Review and Meta-Analysis" (PRISMA, Page et al., 2021a, 2021b), with the method sections corresponding to the PRISMA 2020 checklist (see table 2).

Eligibility criteria

Relevant studies were identified based on the following inclusion criteria for population, indicator, comparison, outcome and study design (PICOS): (P) infants' participants below 24 months of age, (I) habituation or familiarization phase of efficient or inefficient stimuli, (C) test phase comparing dishabituation to efficient and inefficient stimuli in new situation, (O) measuring looking time at central fixation for the efficient and inefficient stimuli, (S) study using a violation-of-expectation paradigm.

Information sources

The search for the systematic review and meta-analysis was conducted on the Google Scholar database at the start of May 2021. All results were screened over three days. The University website (CEU People) for the two most frequent authors on the topic, Csibra and Gergely, was screened for additional papers on May 24. and 29., respectively.

Search strategy

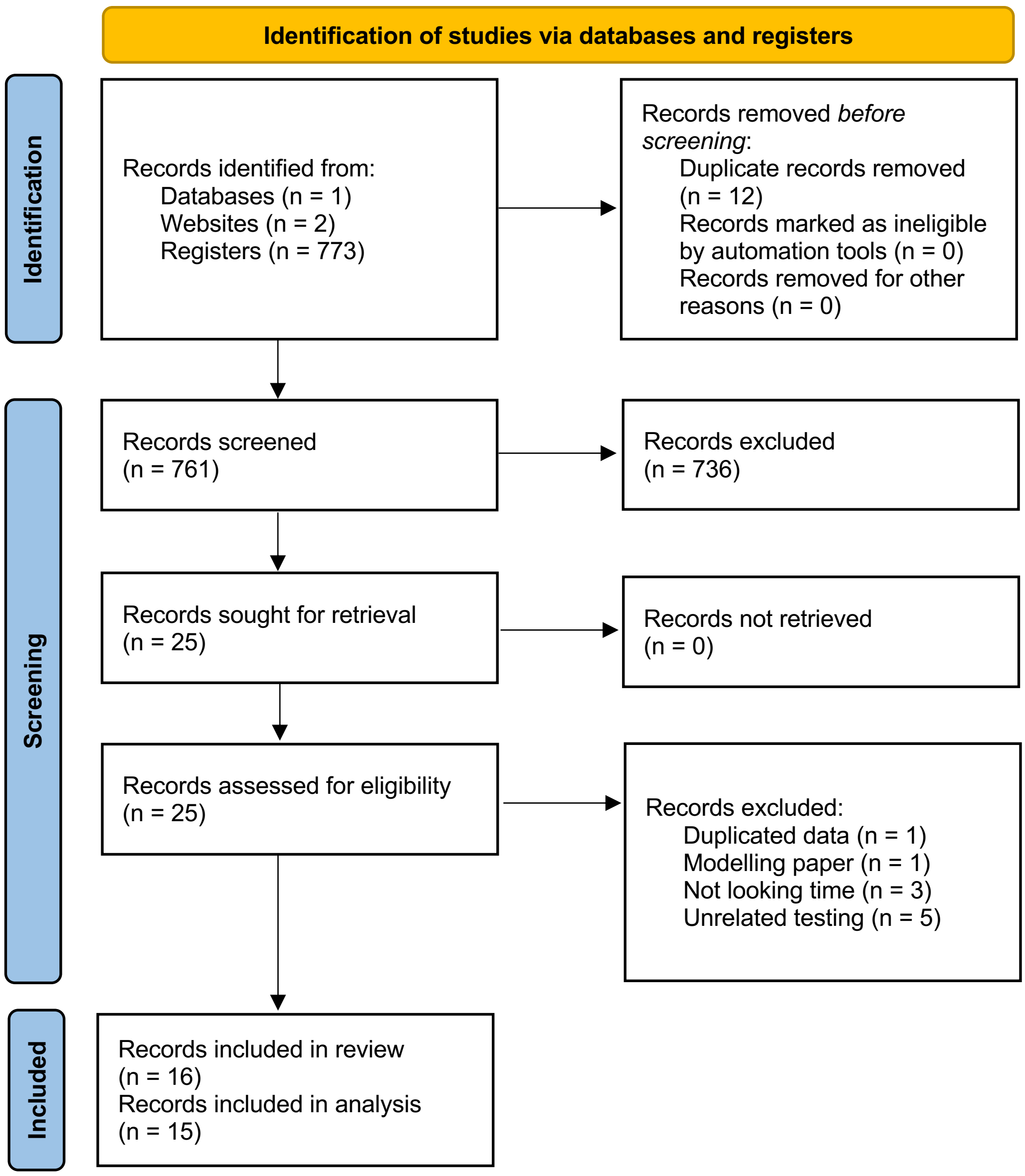
To capture the existing literature, the following search terms was used: "teleological" OR "teleology" "looking time" OR "looking times" "infant" OR "infants". The search yielded 760 results on the Google Scholar database. A missing paper was brough to attention by the supervisor. This paper was found available at the University website of Gergely and Csibra.

Selection process

The papers were sorted by relevance and the results were screened independently by reading the title and the part of the text where the keywords occurred. Potentially relevant papers to the topic were further screened by reading the abstract or accessing the paper. From the total of 773 paper screened, 25 papers were found to be relevant to the topic. Exclusion of 9 of the papers was further performed. One paper was excluded for using data already included in a related paper, one was a modelling paper, three used alternative measurements to looking time, and five papers were testing similar but different aspects. All the 16 included papers where peer reviewed. The sample was double checked by the supervisor who is familiar to the literature of the topic. Figure 2 illustrate the screening process.

Figure 2.

Prisma flowchart



Note. See "Synthesis method" for details on the inclusions criteria for the meta-analysis.

Data Collection Process

Data extraction was performed in accordance with guidelines for MetaLab (metalab.standford.edu), a platform for conducting and accessing meta-analyses (Gasparini et al., 2021). The data was coded in MetaLab template spreadsheet with pre-structured field specifications. This included information of paper descriptive, experimental descriptive and variables to compute effect sizes, and any other information that could be useful as potential moderators. If means or SD were not reported in text, the data was extracted from figures by using a web-based tool called WebPlotDigitizer (Rohatgi, 2021). The data set was checked for similarities between studies to find potential duplicates. Two duplicates were found, and it was decided to keep the data from the first published paper. The data collection for the meta-analysis was carried out independently, but double checked for accuracy by the supervisor.

Data items

The following study characteristics were coded: number of participants, mean age, means of dependent variable, standard deviations, and t-value. Age range, gender, and average number of habituation trials were coded as potential moderator variables. Age range was not included in the current analysis.

Effect measures

Cohen's d and correlation coefficients was calculated directly for 11 experimental and 4 control condition from data made available by the authors. Correlation coefficient was further calculated for 2 experimental and 1 control condition from SD and t-values in papers. The following formula was used (Csibra et al., 2016):

$$r = \frac{SD_1^2 + SD_2^2 - ((M_1 - M_2)^2 \times \frac{n}{t^2})}{2 \times SD_1 \times SD_2}$$

The average values of the correlation coefficient r from 13 experimental (r = 0.66) and 5 control conditions (r = 0.43) was used as estimates for the remaining 21 experimental and

10 control conditions. The estimated correlation coefficient was used to calculate Cohen's d for 25 experimental and 11 control conditions with formula (9) by Lakens (2013):

$$Cohen's\ d_{rm} = \frac{M_{Diff}}{\sqrt{SD_1^2 + SD_2^2 - 2 \times r \times SD^2 \times SD^2}} \times \sqrt{2(1-r)}$$

Since Gergely et al. (1997) reported no SD values, Cohen's d was calculated with t-value for the 2 experimental and 2 control conditions with formula (7) by Lakens (2013):

$$Cohen's\ d_z = \frac{t}{\sqrt{n}}$$

Cohen's d SE was calculated for all 53 conditions with formula by Hedges and Olkin (1985):

$$SE = \sqrt{\frac{1}{n} + \frac{d^2}{2n}}$$

Data were coded so that positive effect size indicate longer looking time for familiar/ efficient stimuli, and negative effect size indicate looking time for novel/efficient stimuli.

Synthesis method

A broad inclusion criterion was used to decide which experimental condition to use for the meta-analysis. Experiment condition had to include a (1) habituation or familiarization phase of an efficient goal-directed action, with the following (2) test phase of an efficient vs. inefficient action in a new situation. This includes conditions which were originally designed or interpreted as testing for the role of factors counter-effective to interpreting action as goal-directed (i.e., not-self-propelled, failed-reaching, no state-change). Although this might lower the overall effect size, the broad inclusion offers a diverse data set that could better differentiate between age. The control condition included studies with a (1) habituation or familiarization phase of an inefficient goal-directed action, with a following (2) test phase of an efficient vs. inefficient action. This includes Lui and Spelke (2017) third condition which test if infants differentiate between small (efficient) and high (inefficient) jumping height over a small barrier introduced in the test phase. Phillips and Wellman (2005) "reaching without an object" and Brandone et al. (2009) "no goal object" did not fit the criteria of presenting a goal-directed action and were therefore defined as alternative conditions. The data for these studies was coded, but eventually not included in the current analysis. The paper by Gergely (2003) was excluded for insufficient reporting of information to calculate effect size. The same applies for Csibra et al. (1999) experimental condition with 12-month-old infants.

The meta-analysis was conducted using JASP Metaanalysis module (version 0.16) (JASP Team, 2021) with the default method of random effects analysis with a restricted maximum likelihood estimator (REML). All presented figures were generated using this version of JASP. A sensitivity analysis was planned to check robustness of finding by using the average correlation coefficient of a meta-analysis of looking time study from Csibra et al. (2016), which compares within-subject congruent and incongruent outcomes. Statistical heterogeneity was investigated with Cochran's Q-test, estimate heterogeneity variance (τ^2) and inconsistency (I^2). A meta-regression analysis was conducted to access the influence of coefficient variables. Subgroup and exploratory analysis were carried out to explore variance of results across studies (statistical heterogeneity).

Reporting bias assessment

Funnel plots for the meta-analysis was generated to assess small-study effects. Potential asymmetry detected by visual inspection where formally tested with a non-parametric rank test or parametric regression test (Egger's test).

Results

Final sample

The sample ended up at 15 papers (53 conditions) published between 1995 and 2019. Experimental conditions included a total of 737 infants between 91 and 450 days (*M* = 233), and the control conditions including 283 infants in the same age range (*M* = 247). The meta-analysis included 36 experimental conditions with the Cohen's d effect sizes ranging from -0.42 to 2.75, and the 17 control conditions ranging from -3.25 to 0.56.

Risk of bias in studies

Figure 3.

Funnel plot for experimental condition

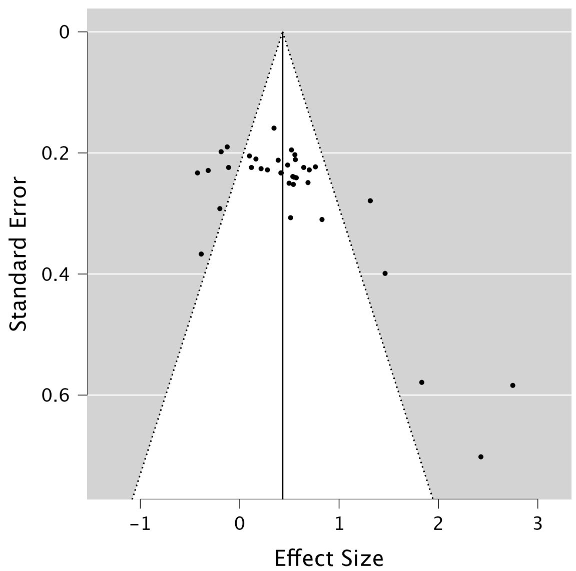
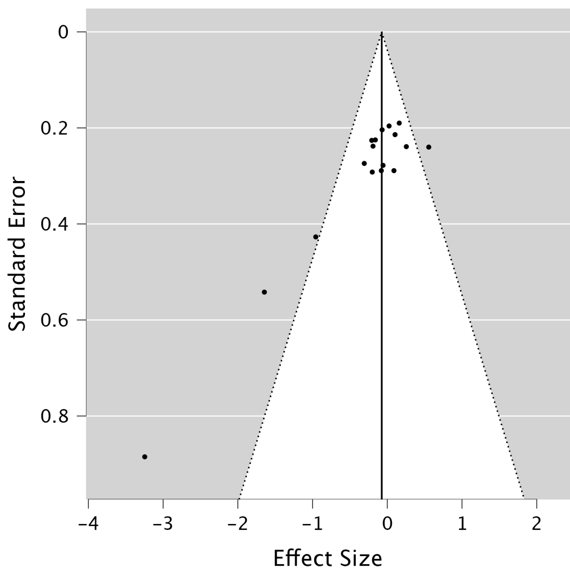


Figure 4.

Funnel plot for control condition



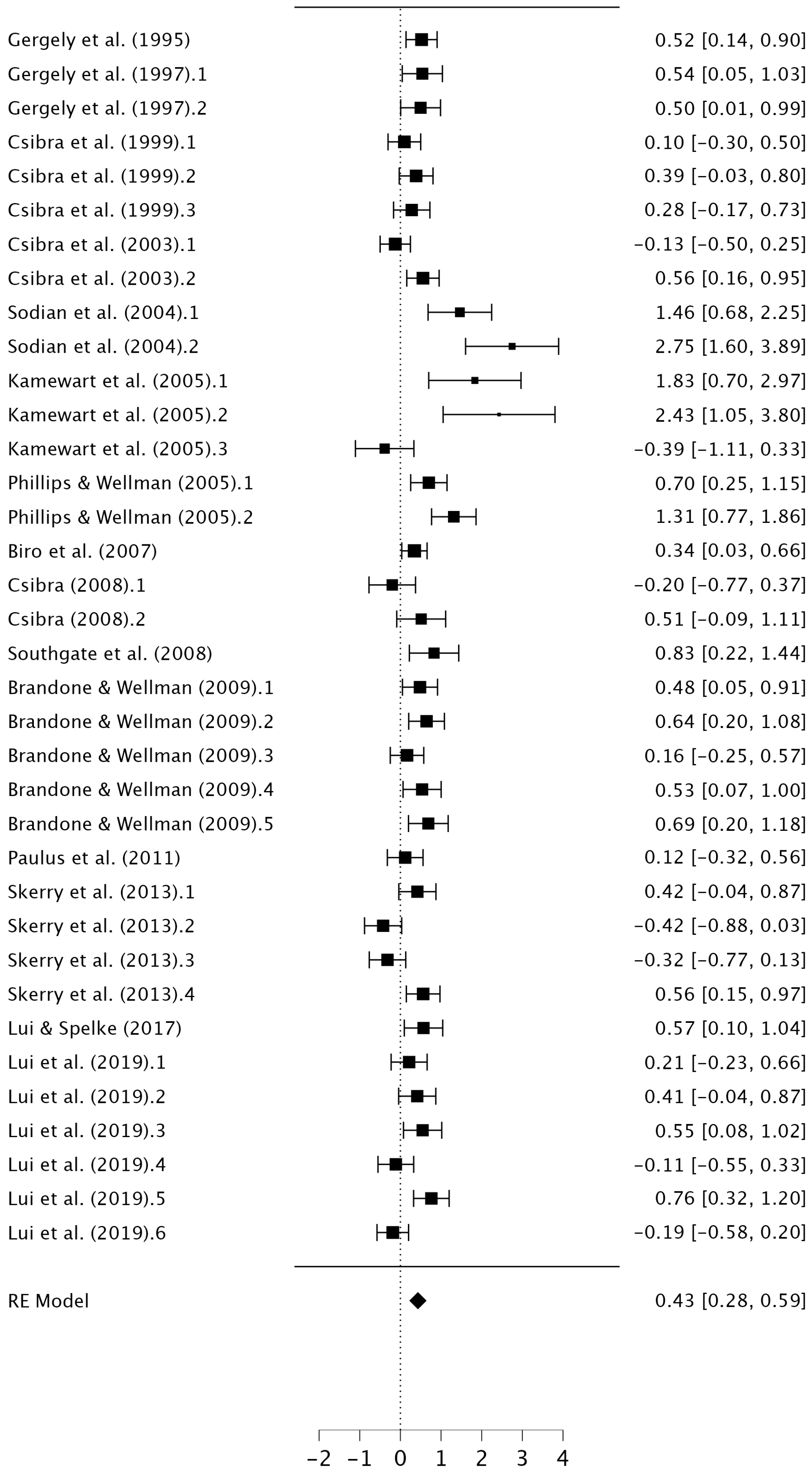
Note. The black dots show effect size against study precision. The white funnel illustrates a 95% confidence interval around the overall effect size. Dots in the gray field may result from heterogeneity.

The funnel plots (Figure 3 & 4) assess the risk of bias in the meta-data by plotting effect size against study precision. Data points should cluster within the funnel if there are no evidence for publication bias or between study heterogeneity (Sterne et al., 2011). Visual inspection of the plots indicated potential asymmetry in both conditions. This was confirmed as statistically significant with Egger's test for funnel plot asymmetry for the experimental (z = 4.94, p = <.001) and control conditions (z = -4.97, p = <.001). Rank correlation test for funnel plot asymmetry suggest evidence of publication bias in the experimental (Kendall's tau = 0.32, p = .006) and control conditions (Kendall's tau = -0.45, p = .012).

Results of syntheses

Figure 5.

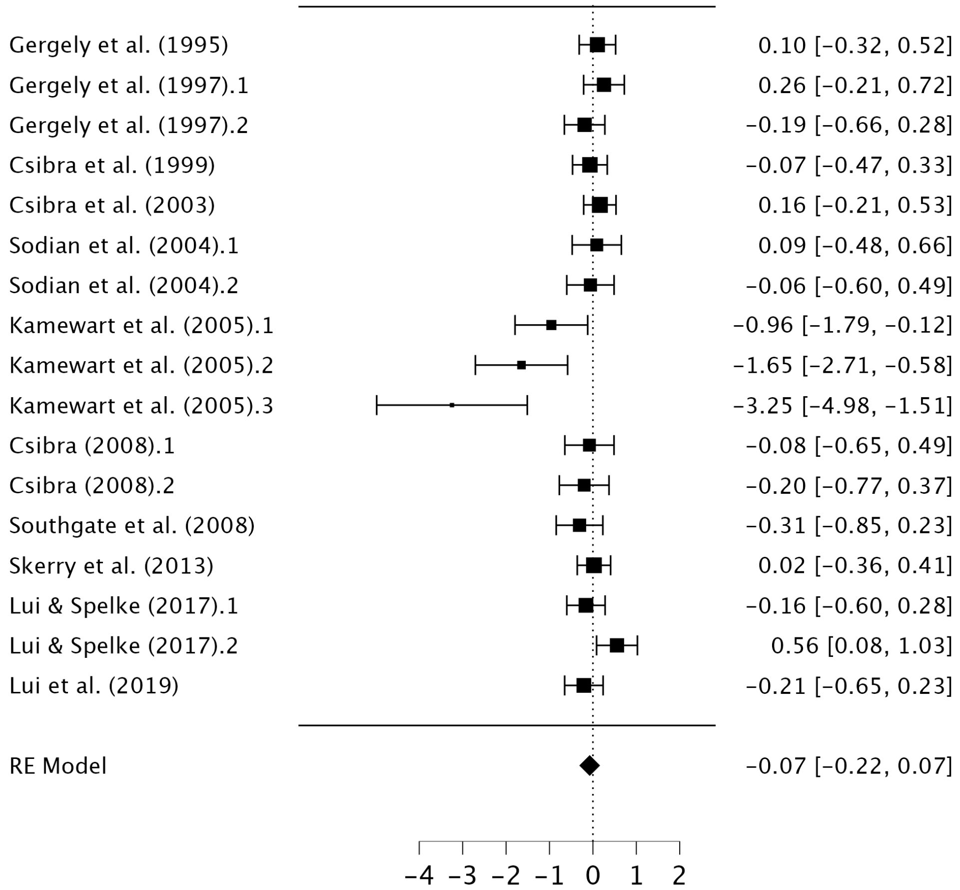
Forrest plot for experimental condition – random effect meta-analysis (REML)



Note. Cohen's d and SE values shown to the right, with overall effect size beneath. Effects visualized in the middle.

Figure 6.

Forrest plot for control condition – random effect meta-analysis (REML)



Note. Cohen's d and SE values shown to the right, with overall effect size beneath. Effects visualized in the middle.

An omnibus meta-analysis confirmed a statistically significant different between the experimental and control conditions (z = 4.20, p = <.001). Meta-analysis for the experimental condition revealed a small positive significant effect size (d = 0.43, p = <.001) and a null finding for the control conditions (d = -0.074, p = .324). A separate meta-analysis was conducted assuming the correlation coefficient (r = .42) from Csibra et al. (2016) in cases where the correlation coefficient could not be estimated from the available data. It resulted in approximately the same effect size for experimental (d = 0.44, p = <.001) and control conditions (d = -0.074, p = .326). Cochran's Q-test indicated significant residual heterogeneity for the experimental (Q (35) = 118.9, p = <.001) and control conditions (Q (16) = 38.56, p = .001). This refuted the hypothesis of homogeneity, which suggest that the effect sizes differed across studies. Residual heterogeneity estimate showed excess variance in the experimental ($\tau^2 = 0.15$, $I^2(\%) = 72.9$) compared to the control conditions ($\tau^2 = 0.027$, $I^2(\%) = 29.4$). Thus, there is clear evidence for between study heterogeneity in the experimental conditions, which justified performing further analysis to investigate causes of variance.

A meta-regression analysis was conducted due to the excess heterogeneity variance in the experimental conditions. To check whether moderator variables could have played a role, the variables of mean age, gender and number of habituation trials were included as coefficient factors. Mean age significantly explained variance of effect size in the experimental conditions ($z = 2.10$, $p = .036$), but not gender ($z = 1.04$, $p = .717$) or number of habituation trials ($z = -0.097$, $p = .923$). For the control condition, no significant moderators were found with mean age ($z = 1.07$, $p = .283$), gender ($z = 0.42$, $p = .673$), and number of habituation trails ($z = -0.24$, $p = .809$). While the test for residual heterogeneity indicated that the age moderator explained difference between studies in the experimental conditions, there were still unexplained excess variance (Q (25) = 78.6, p = <.001). Subgroup analyses of age cluster were carried out to investigate the age trend further.

Table 1.

Overall effect size per age groups – Random effects model

3.5-mo.-old		6-7-mo.-old		8-10-mo.-old		12-14-mo.-old	
d	*p*	*d*	*p*	*d*	*p*	*d*	*p*
0.19	.158	0.57	.042	0.32	<.001	0.83	<.001

Note. Sorted based on age clusters in scatterplot

Subgroup analyses revealed inconsistency in the size of effect between the age groups in experimental conditions (Table 1). Asymmetry in funnel plot indicate influence of underpowered studies effect sizes. A scatterplot was assembled by using the mean age in months (x-axis) against the effect sizes weighted by sample size (y-axis). When assuming non-linear age trend, the scatterplot demonstrated a more consistent developmental pattern, with the size of effect increasing more rapidly after around 9-month of age (Figure 7).

Figure 7.

Scatterplot assuming non-linear age trend

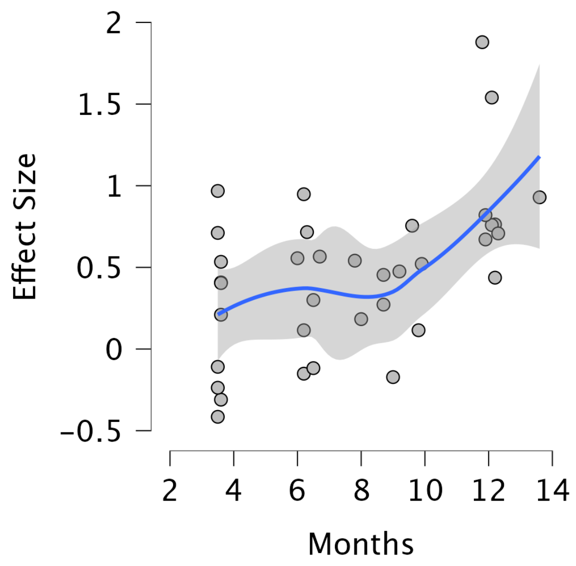
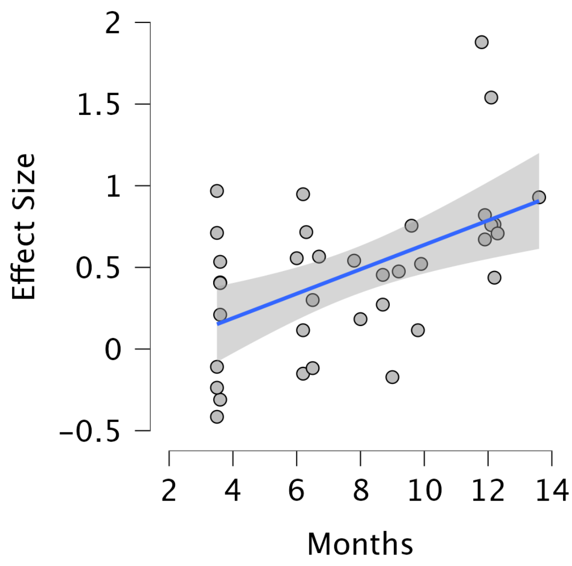


Figure 8.

Scatterplot assuming linear age trend



Note. Figures explores inconsistency in subgroup analysis. The gray dots show Cohen's d effect size (weighted by sample size) against mean age. The gray field shows 95% confidence interval around the regression line in blue.

Figure 9.

Correlation analysis excluding 12-14-mo.-old group

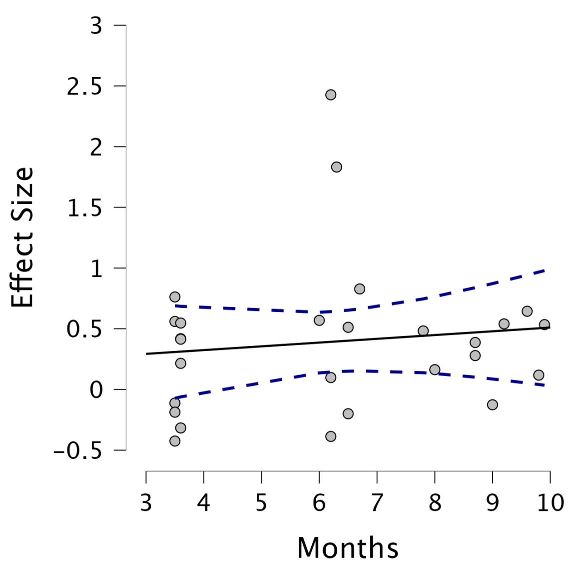
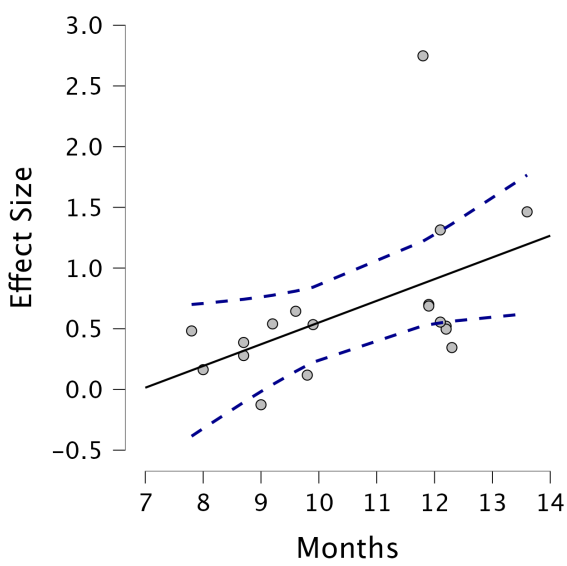


Figure 10.

Correlation analysis excluding 3.5- & 6-7-mo.-old group



Note. Figures demonstrate the exploratory correlation analysis for the age cluster subgroups. The gray dots show Cohen's d effect size against mean age. The blue lines show 95% confidence interval around the regression line in black.

Another meta-regression analysis was conducted to investigate a potential developmental milestone occurring after 9-month of age (as suggested in the literature). Since the infants' oldest infants accounted for the largest effect size, a factor was created separating

12- to 14-month-old group from the youngest age groups. This coefficient factor significantly explained variance in the effect sizes (z = 3.038, p = .002). An exploratory partial correlation analysis revealed no significant correlation between effect size and mean age in the youngest age groups (r = .119, p = .562) after controlling for sample size (Figure 9). The strongest correlation was found when excluding infants from the 3.5- and 6-7-month-old group (r = .585, p = .014) (Figure 10).

Discussion

The main goal of the meta-analysis was to provide meta-analytic estimate of the overall effect size for infants reasoning about goal-directed action in efficient (experimental) and inefficient agents (control conditions). The topics presented in the review will further be addressed in a new way with the findings from the meta-analysis.

Main effect

The overall effect size of the meta-analysis revealed a small statistically significant effect for the experimental and no significant effect for the control condition. The finding the experimental conditions confirmed that infants attribute goals to efficient agents. The null finding of the control conditions was not consistent with the hypothesis of a novelty effect for the inefficient stimuli. This suggest that infants respond to the stimulus of the inefficient agent as more than merely a visual stimulation. Both findings supported infants' appliance of a rationality principle to predict the observed action of efficient agents. The finding from the control conditions will be interpreted in more details later.

Age trend effect

The meta-regression analysis showed evidence for an age trend effect in the experimental conditions. This indicate that with age, infants are more likely to attribute goals to efficient agents. This is to be expected, as the reviewed literature showed that older infants attribute goals in more diverse situations (e.g., Csibra et al., 2003). For the control condition, infants seem not to change how they process the inefficient stimuli throughout development. This allowed for rejecting the hypothesis that infants interpret the control conditions in a more sophisticated manner as they age.

The subgroup analysis revealed inconsistency in the size of effect between the age clusters. This might be accounted for by the liberal inclusion criteria, or the influence of underpowered studies. Since the 12- to 14-month-old group accounted for the largest effect size, this could possibly indicate a non-linear age trend. Meta-regression analysis found that mean age accounted significantly for the variance in effect size, but not gender or the number of habituation trails. Further meta-regression analysis found statistical evidence for the oldest infants showing significantly higher effects than younger infants. The exploratory correlation analysis revealed a more consistent (non-significant) age trend in the youngest infants, with a moderate correlation between effect size and age for the 8-10- and 12-14-month-old group. This is not to say that no changes arise beforehand. Difference in significance indicate that the effect variate more in 3.5- and 6-7-month-old group, which could reflect a lesser ability to generalize goal attribution to more diverse situations. The analysis gives some support for Tomasello (1995) supposition of a developmental milestone occurring in the last part of the first year of life. However, the scatterplot showed two values of effect sizes that could contribute to lift the effect in the older infants. Inclusion of more studies are required to get a clearer picture of whether the age trend rather develops more linear (Figure 9).

Mentalistic vs. non-mentalistic attribution

If the analysis reflects a non-linear developmental pattern in infants' goal attribution, then this could offer a potential link between non-mentalistic and mentalistic attribution of goals. Csibra et al. (2003) found that 12-month-old (but not 9-month-olds) infants are able to attribute goals in incomplete teleological contexts. Meaning that when two of the teleological elements are made salient (action, goal, or constraint), older infants often successfully infer the likely third element. This assumes that the teleological relation could be fulfilled but is yet to be confirmed to do so by the observer (unlike the control condition). More sophisticated reasoning would allow older infants to imagine the completion of the teleological relation. A potential mechanism could be the emergence of a mentalistic ability. This would allow infants to attempt to understand the agent by hypothesizing their intention (to act), desire (for goal), or belief (about constraint). Brandone et al. (2009) found results which support some mentalistic interpretation occurring in 10- and 12-month-old infants.

Emergence of goal attribution

The subgroup analysis revealed a non-significant effect size for the 3.5-month-old infants. The evidence for goal attribution in 3.5-month-old infants either is not very clear, or the effect is masked by big variance of the effect in the included studies. Since a broad inclusion criterion was used for the analysis, the effect variance might reflect an initial restricted (rather than no) ability for goal attribution. Emergence of goal attribution could be confined to agents and actions familiar to the infants. While older infants generalize the effect more successfully to action themselves cannot perform (Sodian et al., 2004; Kamewari et al., 2005), 3.5-month-olds show to be less flexible in their ability to do the same (Skerry et al., 2013; Lui et al., 2019). While the reviewed literature supports such hypothesis, replication studies are needed to be more certain of the findings from the individual studies.

Interpretation of the control condition

I proposed three different hypotheses to explain possible finding from the control conditions: (1) infants do not interpret the stimuli as goal directed, (2) infants abandon action prediction due to inefficiency, or (3) infants change their interpretation of the control condition as they develop. As mentioned earlier, the finding was not consistent with a novelty nor an age trend effect. This makes both the first and latter hypothesis unlikely to explain how infants interpret the control conditions. The null finding may indicate that infants respond to inefficient stimuli as more than merely a visual stimulation, while the lack of age-effect suggest that this response is consistent throughout development. The meta-analysis gives support for Gergely et al. (1995) hypothesis that infants abandon goal attribution due to not meeting the requirement of efficiency.

This calls into question the findings from Lui et al. (2017) experiment, which demonstrated goal attribution to previous inefficient agents (when introducing a constraint). I proposed that Lui et al. (2017) experiment could perhaps test something slightly different from infants' interpretation of inefficient agents. Lui et al. (2017) experiment could be testing differentiation between level of efficiencies, but not infants' interpretation of truly inefficient agents. Meaning that infants will differentiate between high and low jumps but perceive both jumping heights as more or less efficient. Thus, the finding show that the criteria for abandoning action prediction turns out to be more conservative (i.e., only when irrational).

An alternative account by Lui et al. (2017) proposed that infants suspend expectation of efficiency only in the narrow context of the control conditions without any constraint. The difference between the habituation and test phase of the control condition might not be salient for the infants. The main difference was the removal of a barrier that did not form an obstacle as it was placed behind the agent. Since the barrier was not related to goal attribution, the removal might not have indicated the test phase as a new situation. This account refines Gergely et al. (1995) hypothesis that infants abandon action prediction due to inefficiency. The null finding from the meta-analysis would only be relevant to contexts perceived as the same, meaning that the inefficient habituation phase won't generalize well for other scenarios.

Which of these alternatives are more likely to explain the null finding from the control conditions? The null finding itself only indicate that some mental processes likely arise, which cannot be accounted for by visual dishabituation. Gergely et al. (1995) proposed that the response reflect goal attribution to the inefficient agent. While it seems like infants assimilate the inefficient stimuli, it is less clear whether accommodation occurs. It should be unlikely that learning that occurs during the habituation phase will exceed well-established knowledge that agents minimize the costs of action (i.e., over-hypothesis). This gives support for Lui et al. (2017) interpretation that infants only suspend action prediction due to the narrow context. I wish to further challenge this idea by proposing an alternative account that infants don't completely abandon action prediction, but rather are biased against it due to not meeting the efficiency principle. This allows for a more flexible goal attribution which makes exception for when new cues are introduced (i.e., barrier), while abandoning action perceived as irrational (jumping without a barrier) or not as salient (straight path).

Limitation of evidence included in the review

The studies included in the review used relatively small sample sizes for the experimental (*M* = 20.5) and control conditions (*M* = 16.6). Egger's test for funnel plot asymmetry was found to be significant for both conditions, which indicate evidence for small-study bias. For instance, the study with the largest effect sizes in experimental conditions used a small sample size (n = 14). For the control conditions, the paper with the lowest sample size (n = 8) showed the largest negative effect sizes for all three conditions. Rank correlation test for funnel plot asymmetry suggested evidence of publication bias in the experimental and control conditions. I suspect that here could be a file-drawer-effect, where small sample studies are less likely to be published due to small or non-significant findings. Nevertheless, I believe the overall effect size should be a good estimate due to the broad inclusion criteria used in the meta-analysis.

Implication of studies

The multiple reported studies reviewed show reliable effects for the results indicating infants' expectation of efficiency. The meta-analysis has established that we can expect a small effect size for the experimental, and a null finding for the control conditions. The meta-regression analysis demonstrated an age trend in the experimental condition, which according to the subgroup and exploratory analysis indicate a more rapid increase in the last part of the first year of life. This could be a useful cue in the pursuit of establishing when early mentalistic interpretation is likely to emerge in infancy. The ability is likely to be restricted to the concept of intentionality (Tomasello, 1995) or goal-directed action (Brandone et al., 2009), but supports Onishi and Baillargeon (2005) account for an earlier emerging mentalistic ability towards the developing Theory of Mind.

The null finding for the control condition is consistent with the hypothesis that infants interpret the inefficient stimuli in terms of teleological reasoning. This supports Gergely et al. (1995) hypothesis that infants abandon action prediction due to not meeting the requirement of efficiency. To correspond with the finding by Lui et al. (2017) which showed that infants attribute goals to previous inefficient agent, I wish to refine Gergely et al. (1995) original interpretation of the null finding. I propose that infants don’t completely abandon action prediction, but rather are biased against it due to not meeting the efficiency principle. This allows for more flexibility in teleological reasoning by making exception for when new cues are introduced, while abandoning action prediction in less salient scenarios. While the meta-analysis doesn’t deny the possibility that infants abandon action prediction only in the narrow context of the control condition, I hope for this account to further challenge new studies on how infants perceive inefficient goal-directed agents.

Other information

The data collection process for the Meta-Analysis was carried out in MetaLab template spreadsheet. This gives us the opportunity for the analysis to be made publicly accessible on the Metalab website (metalab.standford.edu) and allows for further updates with new or unpublished studies (Gasparini et al., 2021).

References

Aslin, R. N. (2007). What's in a look? *Developmental science, 10*(1), 48-53. https://doi.org/10.1111/j.1467-7687.2007.00563.x

Baillargeon, R. (1987). Object permanence in 3½-and 4½-month-old infants. *Developmental psychology, 23*(5), 655. https://doi.org/10.1037/0012-1649.23.5.655

Biro, S., Csibra, G., & Gergely, G. (2007). The role of behavioral cues in understanding goal-directed actions in infancy. *Progress in brain research, 164*, 303-322. https://doi.org/10.1016/S0079-6123(07)64017-5

Brandone, A. C., & Wellman, H. M. (2009). You can't always get what you want: Infants understand failed goal-directed actions. *Psychological Science, 20*(1), 85-91. https://doi.org/10.1111/j.1467-9280.2008.02246.x

Csibra, G. (2008). Goal attribution to inanimate agents by 6.5-month-old infants. *Cognition, 107*(2), 705-717. https://doi.org/10.1016/j.cognition.2007.08.001

Csibra, G., Bíró, S., Koós, O., & Gergely, G. (2003). One-year-old infants use teleological representations of actions productively. *Cognitive Science, 27*(1), 111-133. https://doi.org/10.1016/S0364-0213(02)00112-X

Csibra, G., Gergely, G., Bíró, S., Koos, O., & Brockbank, M. (1999). Goal attribution without agency cues: the perception of 'pure reason' in infancy. *Cognition, 72*(3), 237-267. https://doi.org/10.1016/S0010-0277(99)00039-6

Csibra, G., Hernik, M., Mascaro, O., Tatone, D., & Lengyel, M. (2016). Statistical treatment of looking-time data. *Developmental psychology, 52*(4), 521. https://doi.org/f8hghz

Dennett, D. C. (1989). *The intentional stance.* MIT press.

Gasparini, L., Tsuji, S., & Bergmann, C. (2021). 10 easy steps to conducting transparent, reproducible meta-analyses for infant researchers. http://doi.org/10.31219/osf.io/wp3fs

Gergely, G., & Csibra, G. (1997). Teleological reasoning in infancy: The infant's naive theory of rational action: A reply to Premack and Premack. *Cognition, 63*(2), 227-233. https://doi.org/10.1016/S0010-0277(97)00004-8

Gergely, G., Csibra, G., Biro, S., & Koós, O. (1997). Understanding rational action in infancy. *Psychology of Language and Communication, 1*(2), 29-37.

Gergely, G., Nádasdy, Z., Csibra, G., & Bíró, S. (1995). Taking the intentional stance at 12 months of age. *Cognition, 56*(2), 165-193. https://doi.org/fvjkg5

Griffin, R., & Baron-Cohen, S. (2002). The intentional stance: Developmental and neurocognitive perspectives. In Andrew Brook & Don Ross (eds.)*, Daniel Dennett* (pp. 83-116). Cambridge University Press

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Academic press.

Heider, F., & Simmel, M. (1944). An experimental study of apparent behavior. *The American journal of psychology, 57*(2), 243-259. https://doi.org/10.2307/1416950

JASP Team (2021). *JASP* (Version 0.16) [Computer software]. https://jasp-stats.org

Johnson, S. C. (2000). The recognition of mentalistic agents in infancy. *Trends in cognitive sciences, 4*(1), 22-28. https://doi.org/10.1016/s1364-6613(99)01414-x

Kamewari, K., Kato, M., Kanda, T., Ishiguro, H., & Hiraki, K. (2005). Six-and-a-half-month-old children positively attribute goals to human action and to humanoid-robot motion. *Cognitive Development, 20*(2), 303-320. https://doi.org/10.1016/j.cogdev.2005.04.004

Keil, F. (2013). *Developmental psychology: the growth of mind and behavior.* WW Norton & Company.

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Frontiers in psychology, 4*, 863. https://doi.org/10.3389/fpsyg.2013.00863

Liu, S., Brooks, N. B., & Spelke, E. S. (2019). Origins of the concepts cause, cost, and goal in prereaching infants. *Proceedings of the National Academy of Sciences, 116*(36), 17747-17752. https://doi.org/10.1073/pnas.1904410116

Liu, S., & Spelke, E. S. (2017). Six-month-old infants expect agents to minimize the cost of their actions. *Cognition, 160, 35-42.* https://doi.org/10.1016/j.cognition.2016.12.007

Oakes, L. M. (2010). Using habituation of looking time to assess mental processes in infancy. *Journal of Cognition and Development, 11*(3), 255-268. https://doi.org/cdwb4s

Onishi, K. H., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *science, 308*(5719), 255-258. https://doi.org/10.1126/science.1107621

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., ... Moher, D. (2021a). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ,* n71. https://doi.org/10.1136/bmj.n71

Page, M. J., Moher, D., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R. C., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li. T., Loder, E. W., Mayo-Wilson, E., McDonald, S., ... & McKenzie, J. E. (2021b). PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews. *BMJ,* n160. https://doi.org/10.1136/bmj.n160

Paulus, M., Hunnius, S., van Wijngaarden, C., Vrins, S., van Rooij, I., & Bekkering, H. (2011). The role of frequency information and teleological reasoning in infants' and adults' action prediction. *Developmental psychology, 47*(4), 976. https://doi.org/fq5gq6

Phillips, A. T., & Wellman, H. M. (2005). Infants' understanding of object-directed action. *Cognition, 98*(2), 137-155. https://doi.org/10.1016/j.cognition.2004.11.005

Rohatgi, A. (2021). *WebPlotDigitizer* (Version 4.5) [Computer software]. https://automeris.io/WebPlotDigitizer

Skerry, A. E., Carey, S. E., & Spelke, E. S. (2013). First-person action experience reveals sensitivity to action efficiency in prereaching infants. *Proceedings of the National Academy of Sciences, 110*(46), 18728-18733. https://doi.org/f5grhj

Sodian, B., Schoeppner, B., & Metz, U. (2004). Do infants apply the principle of rational action to human agents? *Infant Behavior and Development, 27*(1), 31-41. https://doi.org/10.1016/j.infbeh.2003.05.006

Sterne, J. A., Sutton, A. J., Ioannidis, J. P., Terrin, N., Jones, D. R., Lau, J., Carpenter, J., Rücker, G., Harbord, R., Schmid, C. H., Tetzlaff, J., Deeks, J. J., Peters, J., Macaskill, P., Schwarzer, G., Duval. S., Altman, D. G., Moher, D., & Higgins, J. P. (2011). Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *Bmj*, 343. https://doi.org/10.1136/bmj.d4002

Striano, T., & Reid, V. M. (2006). Social cognition in the first year. *Trends in cognitive sciences, 10*(10), 471-476. https://doi.org/10.1016/j.tics.2006.08.006

Tomasello, M. (1995). Joint attention as social cognition. In C. Moore & P. J. Dunham (Eds.), *Joint attention: Its origins and role in development* (pp. 103-130). Lawrence Erlbaum Associates, Inc.

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: The truth about false belief. *Child development, 72*(3), 655-684. https://doi.org/10.1111/1467-8624.00304

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition, 13*(1), 103-128. https://doi.org/10.1016/0010-0277(83)90004-5

Appendix

Table 2.

PRISMA 2020 Checklist

Section and Topic	Item #	Checklist item	Location where item is reported
Title			
Title	1	Identify the report as a systematic review	Title page.
Abstract			
Abstract	2	See the PRISMA 2020 for Abstracts checklist.	Page 3.
Introduction			
Rationale	3	Describe the rationale for the review in the context of existing knowledge.	Page 4 – 5.
Objectives	4	Provide an explicit statement of the objective(s) or question(s) the review addresses	Page 15 – 17.
Methods			
Eligibility criteria	5	Specify the inclusion and exclusion criteria for the review and how studies were grouped for the synthesis.	Page 18.
Information sources	6	Specify all databases, registers, websites, organizations, reference lists and other sources searched or consulted to identify studies. Specify the date when each source was last searched or consulted.	Page 18
Search strategy	7	Present the full search strategies for all databases, registers, and websites, including any filters and limits used.	Page 18
Selection process	8	Specify the methods used to decide whether a study met the inclusion criteria of the review, including how many reviewers screened each record and each report retrieved, whether they worked independently, and if applicable, details of automation tools used in the process.	Page 19.
Data collection process	9	Specify the methods used to collect data from reports, including how many reviews collected data from each report, whether they worked independently, and processes for obtaining or confirming data from study investigators, and if applicable, details of automation tools used in the process.	Page 20.
Data items	10a	List and define all outcomes for which data were sought. Specify whether all results that were compatible with each outcome domain in each study were sought (e.g. for all measures, time points, analyses), and if not, the methods used to decide which results to collect.	Page 20.
	10b	List and define all other variables for which data were sought (e.g. participant and intervention characteristics, funding sources). Describe any assumptions made about any missing or unclear information.	Page 20.
Study risk of bias assessment	11	Specify the methods used to assess risk of bias in the included studies, including details of the tool(s) used, how many reviewers assessed each study and whether they worked independently, and if applicable, details of automation tools used in the process.	NA
Effect measures	12	Specify for each outcome the effect measure(s) (e.g. risk ratio, mean difference) used in the synthesis or presentation of results.	Page 20 – 21.
Synthesis methods	13a	Describe the processes used to decide which studies were eligible for each synthesis (e.g. tabulating the study intervention characteristics and comparing against the planned groups for each synthesis (item #5)).	Page 21 – 22.
	13b	Describe any methods required to prepare the data for presentation or synthesis, such as handling of missing summary statistics, or data conversions.	NA
	13c	Describe any methods used to tabulate or visually display results of individual studies and syntheses.	Table 3.
	13d	Describe any methods used to synthesize results and provide a rationale for the choice(s). If meta-analysis was performed, describe the model(s), method(s) to identify the presence and extent of statistical heterogeneity, and software package(s) used.	Page 21 – 22.
	13e	Describe any methods used to explore possible causes of heterogeneity among study results (e.g. subgroup analysis, meta-regression).	Page 21 – 22.
	13f	Describe any sensitivity analyses conducted to assess robustness of the synthesized results.	Page 21 – 22.
Reporting bias assessment	14	Describe any methods used to assess risk of bias due to missing results in a synthesis (arising from reporting biases).	Page 22.

Certainty assessment	15	Describe any methods used to assess certainty (or confidence) in the body of evidence for an outcome.	NA
Results			
Study selection	16a	Describe the results of the search and selection process, from the number of records identified in the search to the number of studies included in the review, ideally using a flow diagram.	Page 22 – 23.
	16b	Cite studies that might appear to meet the inclusion criteria, but which were excluded, and explain why they were excluded.	NA
Study characteristics	17	Cite each included study and present its characteristics.	Table 3.
Risk of bias in studies	18	Present assessments of risk of bias for each included study.	Page 23.
Results of individual studies	19	For all outcomes, present, for each study: (a) summary statistics for each group (where appropriate) and (b) an effect estimate and its precision (e.g. confidence/credible interval), ideally using structured tables or plots.	Page 24 – 25.
Results of syntheses	20a	For each synthesis, briefly summarize the characteristics and risk of bias among contributing studies.	NA
	20b	Present results of all statistical syntheses conducted. If meta-analysis was done, present for each the summary estimate and its precision (e.g. confidence/credible interval) and measures of statistical heterogeneity. If comparing groups, describe the direction of the effect.	Page 23 – 28.
	20c	Present results of all investigations of possible causes of heterogeneity among study results.	Page 25 – 28.
	20d	Present results of all sensitivity analyses conducted to assess the robustness of the synthesized results.	Page 25.
Reporting biases	21	Present assessments of risk of bias due to missing results (arising from reporting biases) for each synthesis assessed.	NA
Certainty of evidence	22	Present assessments of certainty (or confidence) in the body of evidence for each outcome assessed.	NA
Discussion			
Discussion	23a	Provide a general interpretation of the results in the context of other evidence.	Page 28 – 32.
	23b	Discuss any limitations of the evidence included in the review.	Page 32.
	23c	Discuss any limitations of the review processes used.	NA
	23d	Discuss implications of the results for practice, policy, and future research.	Page 32 – 33.
Other information			
Registration and protocol	24a	Provide registration information for the review, including register name and registration number, or state that the review was not registered.	NA
	24b	Indicate where the review protocol can be accessed, or state that a protocol was not prepared.	NA
	24c	Describe and explain any amendments to information provided at registration or in the protocol.	NA
Support	25	Describe sources of financial or non-financial support for the review, and the role of the funders or sponsors in the review.	NA
Competing interests	26	Declare any competing interests of review authors.	NA
Availability of data, code and other materials	27	Report which of the following are publicly available and where they can be found: template data collection forms; data extracted from included studies; data used for all analyses; analytic code; any other materials used in the review.	Page 33.

Table 3.

Meta-Analysis

Authors (year)	Paper		Demographics			Meta-Analysis		
	Condition	n	Age group	Mean age (Days)	Female (%)	Cohen's d	SE	r
Gergely et al. (1995)	Experiment	30	12	369.6	NA	0.521	0.195	0.656
	Control	22	12	369.6	NA	0.104	0.214	0.434
Gergely et al. (1997)	Experiment	18	9	279.3	0.305	0.540	0.252	NA
	Experiment	18	12	370.5	0.444	0.497	0.250	NA
	Control	18	9	179.3	0.305	0.255	0.239	NA
	Control	18	12	370.5	0.444	-0.191	0.238	NA

Csibra et al. (1999)	Experiment	24	6	188.3	0.666	0.098	0.205	0.656
	Experiment	24	9	264.6	0.458	0.387	0.212	0.656
	Control	24	9	264.6	0.458	-0.069	0.204	0.434
	Experiment	20	9	264.6	0.2	0.279	0.228	0.656
	Experiment	20	12	372.4	0.45	NA	NA	NA
Csibra et al. (2003)	Experiment	28	9	273.5	0.535	-0.126	0.190	0.656
	Experiment	28	12	369.2	0.571	0.555	0.203	0.141
	Control	28	12	369.2	0.571	0.169	0.190	0.006
Gergely (2003)	Experiment	20	12	369.0	NA	NA	NA	NA
	Control	20	12	369.0	NA	NA	NA	NA
Sodian et al. (2004)	Experiment	13	14	413.9	0.385	1.463	0.399	0.656
	Control	12	14	431.2	0.416	0.089	0.289	0.434
	Experiment	13	12	360.1	0.518	2.748	0.584	0.656
	Control	12	12	360.1	0.518	-0.057	0.278	0.434
Kamewari et al. (2005)	Experiment	8	6.5	192.5	0.375	1.832	0.579	0.656
	Control	8	6.5	192.5	0.375	-0.957	0.427	0.434
	Experiment	8	6.5	189.0	0.437	2.427	0.702	0.656
	Control	8	6.5	189.0	0.437	-1.645	0.542	0.434
	Experiment	8	6.5	189.0	0.5	-0.387	0.367	0.656
	Control	8	6.5	189.0	0.5	-3.246	0.885	0.434
Phillips & Wellman (2005)	Experiment	24	12	362.6	0.5	0.700	0.228	0.656
	Alternative	24	12	363.6	0.416	NA	NA	NA
	Experiment	24	12	367.6	0.458	1.314	0.279	0.656
	Alternative	24	12	363.6	0.541	NA	NA	NA
Bíró et al. (2007)	Experiment	42	12	372.7	0.357	0.345	0.159	0.656
Csibra (2008)	Experiment	12	6.5	198.1	0.458	-0.200	0.292	0.656
	Control	12	6.5	198.1	0.458	-0.080	0.289	0.434
	Experiment	12	6.5	198.1	0.458	0.512	0.307	0.656
	Control	12	6.5	198.1	0.458	-0.201	0.292	0.434
Southgate et al. (2008)	Experiment	14	6-8	203.8	0.5	0.828	0.310	0.656
	Control	14	6-8	212.9	0.5	-0.308	0.274	0.434
Brandone et al. (2009)	Experiment	23	8	236.9	0.507	0.482	0.220	0.656
	Experiment	24	10	290.5	0.507	0.644	0.224	0.656
	Experiment	23	8	244.2	0.507	0.163	0.210	0.656
	Experiment	20	10	302.3	0.507	0.534	0.239	0.656
	Experiment	20	12	362.2	0.507	0.687	0.249	0.656
	Alternative	24	10	NA	NA	NA	NA	NA
	Alternative	24	12	NA	NA	NA	NA	NA
Paulus et al. (2011)	Experiment	20	9	298.9	NA	0.118	0.224	0.656
Skerry et al. (2013)	Experiment	20	3	108.9	0.4	0.418	0.233	0.887
	Experiment	20	3	106.4	0.5	-0.425	0.233	0.725
	Experiment	20	3	108.1	0.5	-0.317	0.229	0.735
	Experiment	26	3	107.8	0.5	0.560	0.211	0.585
	Control	26	3	107.2	0.5	0.024	0.196	0.838
Lui & Spelke (2017)	Experiment	20	6	181.0	0.5	0.569	0.241	0.742
	Control	20	6	185.5	0.5	-0.160	0.225	0.131
	Control	20	6	177.6	0.5	0.555	0.240	0.562
Lui et al. (2019)	Experiment	20	3	108.0	0.55	0.215	0.226	0.779
	Experiment	20	3	108.0	0.6	0.414	0.223	0.771
	Experiment	20	3	108.0	0.575	0.547	0.240	0.726
	Control	20	3	108.0	0.575	-0.208	0.226	0.635
	Experiment	20	3	107.0	0.6	-0.112	0.224	0.446
	Experiment	26	3	107.0	0.403	0.762	0.223	0.690
	Experiment	26	3	107.0	0.403	-0.187	0.198	0.662