Best behaviors:

Young children's understanding of helping actions, its preconditions and consequences

Laura Schlingloff-Nemecz

Submitted to
Central European University
Department of Cognitive Science

In partial fulfillment of the requirements for the degree of Doctor of Philosophy in Cognitive Science

Supervisor: Gergely Csibra
Secondary supervisor: Christophe Heintz
Advisors: Barbara Pomiechowska, Denis Tatone

Budapest, Hungary & Vienna, Austria
2024

Declaration of authorship

I hereby declare that this submission is my own work and to the best of my knowledge it contains no materials previously published or written by another person, or which have been accepted for the award of any other degree or diploma at Central European University or any other educational institution, except where due acknowledgment is made in the form of bibliographical reference.

The present thesis includes work that appears in the following papers/manuscripts:

Schlingloff-Nemecz, L., Tatone, D. & Csibra, G. (2023). The Representation of Third-Party Helping Interactions in Infancy. *Annual Review of Developmental Psychology 5*, 67-88. https://doi.org/10.1146/annurev-devpsych-120321-033548

Schlingloff, L., Stavans, M. & Csibra, G. (2022). CoCollectors: An iPad game for developmental research on action understanding, partner choice, and cooperative behavior. PsychArchives. https://doi.org/10.23668/psycharchives.8176

Schlingloff, L., Tatone, D., Pomiechowska, B., & Csibra, G. (2020). Do Infants Think That Agents Choose What's Best? *Proceedings of the 42nd Annual Meeting of the Cognitive Science Society*.

Schlingloff, L., Csibra, G. & Tatone, D. (2020). Do 15-month-old infants prefer helpers? A replication of Hamlin *et al.* (2007). *Royal Society Open Science, 7*(4), 191795. https://doi.org/10.1098/rsos.191795

Other parts of the dissertation contain research that was performed collaboratively, and will be submitted for publication with the following co-authors:

Chapter 1.3 with Denis Tatone, Gergely Csibra, and Barbara Pomiechowska

Chapter 2.2 with Barbara Pomiechowska, Denis Tatone, and Gergely Csibra

Chapter 2.3 with Barbara Pomiechowska, Barbu Revencu, Denis Tatone, and Gergely Csibra

Section 3 with Maayan Stavans (joint first authors), Barbu Revencu, Kazuhide Hashiya, Hiromi Kobayashi, and Gergely Csibra

Laura Schlingloff-Nemecz

Acknowledgments

Gergely Csibra: Gergő, it's been a joy and a privilege to work with and learn from you over the last years. Your notion of good science imbues your supervision: from taking null results seriously and emphasizing theory rather than digging for effects, engaging earnestly and playfully with counterintuitive ideas and not being afraid to swim against the current, to never letting superficial, "intuitive" explanations slide. I am especially grateful for your guidance and support: Thanks for always having an open door and spending so much time just discussing ideas (& giving them time to stew), thanks for letting your students make our own way while nudging us towards what is worth pursuing. Somehow out of my biggest frustrations grew the most exciting projects. Last but not least, thank you for throwing and attending parties, sharing the good wine, creating community. I hope in the future I will be able to let your influence show.

Christophe Heintz: I very much appreciated our conversations and your helpful thoughts, comments, and recommendations over the years. Thank you for engaging with my ideas, challenging me, and making me see things from a different perspective.

Denis Tatone: Thanks for taking me under your wing when I was a baby-faced youngster, for the many hours spent chewing through theoretical issues and experimental design (sorry for making the kobos cross-eyed in the end…), for sharing incisive insights and reading recommendations, and for the writing bootcamp.

Barbara Pomiechowska: You were a wonderful mentor when it was most needed. Thanks for the many pep talks, patience, unwarranted degrees of confidence in my skills, and late evenings at the office. I admire your determined attitude, deep thinking, and thorough research practice—it was lovely to work with and alongside you.

Maayan Stavans: I had a great time co-collecting ideas with you, designing our coconut-cracking project, and collaborating through the pandemic mayhem! I learned a lot from your pragmatic approach and your meticulous attention to detail in developing experiments. I will always think of you when organizing information in a table!

The professors and (former) researchers of the department of Cognitive Science—especially Ágnes Kovács, Ernő Téglás, Dan Sperber, György Gergely, Mikołaj Hernik, Nicolò Cesana-Arlotti, Péter Rácz, Jonathan Kominsky, Azzurra Ruggeri: Thanks for providing critical & constructive feedback as well as practical help throughout the years.

Krisztina Andrási, Petra Kármán, Anna Kispál, Dorottya Mészégető, Marianna Nagy, Zsuzsanna Üllei Kovács, Iulia Savoş, Tímea Sugár, Borbála Széplaki-Köllőd, Flóra Tósaki, Orsolya Kovács (CDC) and Kaede Hattori & Hinako Shigenaga (Kyushu University): Thanks a ton for collecting all this data and making these studies come into existence. Special thanks to Dóri: You're just outstanding at what you do, and I'm so grateful for all your help during my maternity leave and after.

Ágnes Volein, Maria Tóth, Edit Vitrai, Christina Regorosa: Many many thanks for all the big and small ways you make the Babylab/Babakutató Labor/KiKo-Forschungszentrum an absolutely dreamy place to work. Somehow there always turned out to be a solution to whatever issue—I really appreciate it.

Réka Györgyné Finta: You are the heart of this department and I'm sure there are many more things to thank you for than I am even aware of. Thanks for tirelessly defending our interests, caring about us personally, and always having a kind word.

Everyone at the CDC and the department at large: I've very much enjoyed being your colleague, and I will miss it!

CEU and the ERC: Thanks for funding this research!

Katarina Begus, Dora Kampis, Arianna Curioni, Rachel Dudley: Thank you for beautiful friendships and sage advice on life, science, and everything else.

Bálint Varga, Francesca Bonalumi, Gábor Bródy, Georgina Török, Helena Miton, Ieva Lukošiūnaitė, Johannes Mahr, Liza Vizmathy, Maja Blesić, Mara Mavridaki, Martin Dockendorff, Nima Mussavifard, Paula Fischer, Simily Sabu, Thomas Wolf, Vjeran Kerić: You made the last years infinitely better. Daniela Kesper, Katharina Flüss, Lena Lebahn, Ranjana Rojem, Theresa Lienau: Thanks for helping keep me sane. Love y'all!!!!!!!

Barbu: I'm so very happy that we met. The list of things to be grateful for is too long. There's been lots of thicks and thins, and I couldn't have asked for a better bff through it all.

My parents, Holger and Angelika, and my sister Linda: Thanks for always encouraging and supporting me. I can't properly express how much I cherish being part of this family. Especially in the last year: thanks for all the babysitting and food deliveries…

Zsuzsi, Pál, and Anikó Nemecz: Thank you for yet more support, babysitting, and food, and for providing a home during my time in Hungary.

My grandmother, Johanna Kinzel: I will always be grateful for unconditional love & nurturing, for modelling resilience, and for all the library books when we were little.

My grandfather, Dieter Schlingloff: My interest in and appreciation for research comes from you. I will always remember how your eyes light up when you talk about the joy of making a discovery and being the first to know something new.

Nyuszi: Thanks for being the best home office mate (and the only one who purrs).

Zsuzsanna: Everything is better with you, the logistics of daily life, the intense challenges of being a PhD student parent, the biggest adventures. I don't know who or where I'd be without you, but it's not an exaggeration to say I wouldn't have made it to this point. You saved me endless times in the last years and frankly may deserve co-authorship (for morning coffees, late night co-working, listening to me rant, providing hugs, astute comments, and the right amount of real talk). I love you and I can't wait to see where life takes us next.

Finally, little Hilda: Thanks for keeping my mind sharp by making sure I'm not getting too much sleep, for making me pay attention to what really matters, and for being brave when I have to go to work. I hope that one day you, too, will be able to follow the path of your choosing—I'll be there to help whenever you want me to.

Abstract

To become competent social agents, young children must make sense of the frequently opaque behaviors of other people and draw appropriate conclusions from them. This dissertation is about how infants and children understand other agents' instrumental and social actions (specifically, helping) by using a naive utility calculus, and the inferences they make from observed interactions to character traits. It comprises three sections.

Section 1 addresses whether infants possess a concept of choice, and use it to generate the expectation that a goal-directed agent will choose the best of multiple available options, meaning the one that yields the highest rewards or requires the least cost to bring about. We argue that this capacity is a precondition for a mature understanding of helping, as the latter requires comparing the action options of the Helpee (contingent on whether or not she receives help) and the Helper (insofar as her options relate to the Helpee's outcome). To probe whether infants can compare alternatives of varying utility, we conducted a set of looking-time and eye-tracking experiments testing whether they think an agent should approach a relatively higher number of goal objects, or a goal that can be reached at relatively lower effort.

Section 2 explores infants' and children's understanding of helping actions. Specifically, we ask whether they possess a utility-based concept of helping whereby the goal of a Helper is to increase the utility the Helpee obtains in reaching her goal. To approach this question empirically, we ran a series of looking-time experiments with infants, as well as an experiment with preschoolers probing what they mean by the term "helping". We also report a replication attempt of Hamlin et al.'s (2007) finding that infants prefer Helpers, a paradigm often used to probe their understanding of helping actions.

Finally, Section 3 investigates whether children interpret third-party social interactions by spontaneously ascribing character traits to agents, and choose partners for their own cooperative endeavors accordingly. While it has been argued that young children, upon observing helping events, ascribe a stable prosocial disposition to a Helper, we maintain that it is unclear whether they do so spontaneously. We developed a tablet-based collaborative foraging game where the player first observes agents differing in helpfulness and skill, subsequently selects one of the previously seen agents as a partner, and plays together with the chosen partner. We used this game to study partner choice in 5- to 10-year-old children and adults across two cultural contexts (Hungary/Austria and Japan).

The research described in this dissertation thus aims to shed light on the mechanisms of early action understanding (i.e., whether infants consider alternative possible goals), test whether a hierarchical action representation and naive utility calculus underlie young children's reasoning about helping behaviors, and investigate to what extent the observation of cooperative interactions from a third-party perspective prompts children to infer traits and informs their own social decision-making.

Table of contents

General introduction

This thesis is about the cognitive processes and concepts by which infants and children make sense of the actions of others. As members of a deeply social species, this is a crucial challenge. Young children rely on their understanding of others' actions for tasks as diverse as acquiring a complex skill through social learning, coordinating with a partner in a joint action, managing conflicts, or mapping out their social network. However, the challenge is a difficult one, as any combination of agent-internal and external causes may contribute to a behavior: Someone may fetch ice from the freezer to cool down a glass of wine, or to reduce swelling on a bruise. In the last decades, great strides have been made to shed light on young children's action understanding. This work suggests that from a young age, infants possess a system of abstract, structured core knowledge about agents (Carey, 2009; Spelke, 2022). Even without a full-blown representational theory of mind, in which mental states such as beliefs and desires are invoked in explanations of behavior[1], infants can thus reason[2] about agents as causally efficacious beings whose efforts are directed at bringing about particular states of the world. This knowledge can support rich inferences and scaffold learning from early on. However, we will argue that some fundamental questions on how young children make sense of observed behaviors have not yet been addressed.

Broadly, this issue can be approached from two perspectives (Malle, 2022): First, one may seek an explanation for an action in terms of its underlying causes, that is, a proximate explanation of why an agent performed a certain behavior (e.g., a goal that an agent pursues). Second, one may generate an inference from observed actions to more general characteristics of the agent (e.g., character traits). The work reported here, structured in three sections (which are each preceded by a detailed introduction), is concerned with both: We investigate by which principles young children ascribe goals to agents, both in first-order instrumental and second-order social actions (helping), and whether children spontaneously interpret agents' behaviors in cooperative contexts as being indicative of cooperation-relevant traits, which subsequently inform children's own social decision-making.

A key assumption underlying the present research is that people take other agents to adhere to the principle of rationality (Dennett, 1987). This is a norm of practical reason,

[1] Whether or not such a theory of mind exists in preverbal infants is under debate (Poulin-Dubois, 2020; Rakoczy, 2022; Scott & Baillargeon, 2017). Here we do not commit to either position and will not discuss the issue further.
[2] The term „reasoning" is often used to refer specifically to a cognitive process that involves a representation of reasons and arguments. Here, we do not mean to imply that such a process is available to young infants (as parent of a toddler, it sure seems like it isn't), and instead use the term more broadly to describe a process of inference.

according to which agents who intend to bring about states of the world ought to select means that are suitable to achieve these goals (Wallace, 2003). One way to satisfy this norm is to act efficiently, i.e., to only invest as much effort as is needed to bring about an outcome. This principle has been argued to be a powerful tool for reasoning when it is applied in a probabilistic generative model of other people's actions, which can be inverted for action understanding (Baker et al., 2009). Relying on simple basic principles that can be flexibly applied to novel observed behaviors to derive goals, even if the input is sparse and ambiguous, is especially helpful for infants, who lack experience with the multitude of aims that those around them pursue. It has been argued that infants assume a "teleological stance" to infer goals from efficient actions (Gergely & Csibra, 2003). More recently, building on this framework, young children have been hypothesized to model agents as selecting action plans from the available options by which they can maximize expected utility (Jara-Ettinger et al., 2016). Here, we investigate the scope of such utility reasoning.

One of the themes we explore in this dissertation is the question to what extent infants and children, in making sense of observed actions, consider what the agent may also do, or could have done instead. Such considerations are essential in people's moral and legal reasoning, where it is assessed how an actually performed behavior compares to a normatively stipulated alternative. However, they may also feature in ordinary action understanding and goal ascription. If people (and infants) generally assume that agents choose maximally rewarding outcomes, and pursue them in an efficient way, this allows them to draw conclusions from observed behaviors to an agent's goal, and anticipate which goal she will select in the future.

In Section 1 of this dissertation, we ask whether this is indeed how infants and children make sense of the social world. For example, if I know you want to get a particular candy bar which is being sold in two adjoining vending machines, but in one of them is three times as expensive as in the other, I can expect that you will choose the cheaper one and will be puzzled if you don't. In such a situation, considering and comparing the utility of the options available to an agent allows the observer to generate action predictions and goal inferences (maybe you didn't really crave the candy, but wanted to get rid of as many coins as possible). We investigated whether infants expect agents whose goal they have been familiarized with (e.g., reach a banana) to choose, when multiple alternative options of varying expected utility become available, the optimal one (i.e., the one that yields higher rewards or can be reached at lower cost). In other words: We ask whether infants, in ascribing a goal, take the agent to express a choice and consider the non-chosen options, or merely expect action efficiency in a narrower sense.

In Section 2, we study whether similar reasoning underlies young children's inferences about actions with a social goal, specifically, helping. In social interactions, observers have to adjudicate whether an agent's behavior is motivated by self-interest or by concern for the welfare of another. If the latter, observers may recruit similar principles as they do for non-social instrumental goals. We ask whether an early concept of helping is that of an action with the goal of bringing about a relatively better (or even the best possible) outcome for another agent. Such a concept is grounded in a utility-reasoning framework of action understanding, and is a hierarchical representation whereby one goal (the Helpee's) is embedded in that of another (the Helper's) (Powell, 2022; Ullman et al., 2009). Here, too, an observer would have to compare the impact of the Helper's behavior on the Helpee's goal pursuit with a counterfactual scenario in which no aid was provided in order to assess whether the former generated a utility increase for the latter. We spell out this concept in more detail, and discuss two alternative, simpler candidates for an early helping concept. Moreover, we report a set of studies with infants and preschoolers in which we tested whether they possess such a concept and apply it to interpret observed third-party interactions. Also covered in this section is a replication of a famous paradigm which has been frequently used to establish an early understanding of helping: Hamlin and colleagues' (2007) manual choice task, with which they found that infants prefer Helpers over non-Helpers.

Finally, Section 3 is about the inferences children derive from observed helping actions. We probe whether children default on assuming that actions performed in a third-party social context are indicative of character traits (e.g., someone helps because she is a Helper), and will choose to interact in a cooperative task with agents who, as a result of these traits, are expected to generate more benefits for them. Human sociality relies on unique, sophisticated forms of cooperation, and it has been hypothesized that humans possess evolved cognitive mechanisms to help navigate cooperative endeavors; among them a propensity to detect cooperation-relevant characteristics of potential collaborators from behavior observation and choose interaction partners accordingly. Such a theoretical framework has been invoked to explain the early-emerging sensitivity of young children to observed helping behaviors, and their preference for Helpers (Kuhlmeier et al., 2014). However, we argue that it is far from obvious that children intuitively interpret actions as revealing character traits. To study this, we developed a tablet game in which participants can observe the behaviors of agents who vary in helpfulness and skill, choose who they themselves want to play with, and subsequently play with that partner; they receive minimal instruction, are not primed with explicit trait labels, and make unsupervised partner choices.

Overall, the thesis explores the processes that allow young children to understand and draw inferences from others' actions, particularly helping actions. We challenge common assumptions and highlight unresolved questions within the field, thus contributing to a more nuanced understanding of the origins and development of social cognition.

Section 1:

Do infants think that agents choose what’s best?

Introduction

According to the naive utility calculus theory (Jara-Ettinger et al., 2016), people use the assumption that agents behave in a utility-maximizing way to reason about observed actions, in order to infer goals, desires and beliefs, and to predict future behaviors. Evidence from research with infants supports the hypothesis that this sort of intuitive psychology is already available in the first year of life and allows them to draw rich inferences about others' various goals or even mental states. Here, we investigate whether infants use the naive utility calculus to ascribe to agents the choice of the better option among multiple available alternatives yielding different amounts of utility. While it is commonly taken for granted that infants apply a concept of choice in interpreting others' behaviors, we will argue that this has not been conclusively demonstrated, and that some fundamental questions about the processes involved in action understanding remain open.

The introduction to this section is structured as follows. First, we review the theoretical framework of teleological reasoning in infancy and the naive utility calculus. We will then discuss relevant research on the question of whether infants use a concept of choice to reason about others' actions. This literature will motivate the projects described in the subsequent empirical chapters, where we aimed to test whether infants would expect an agent to choose a goal option that yields higher rewards (Chapter 1.1), or that can be obtained at relatively lower cost (Chapter 1.2, Chapter 1.3).

Goal-directed action understanding, teleological reasoning, and the naive utility calculus

Humans rely on learning from and about others, and engage in a variety of social interactions, during which they need to rapidly interpret what their interaction partner is doing. For the purpose of extracting such information from the behavior of conspecifics, it is advantageous to interpret these behaviors as generated by a variety of unobservable, but causally efficacious mental states. The observable movements of other types of entities can be explained fully by reference to other mechanisms: To explain why an apple falls from a tree, it is unnecessary to ascribe to it an internal desire to reach the ground, but one can simply invoke the biological ripening process and the laws of gravity. In contrast, people often do things for reasons, because they want things, believe things to be true, are committed to things, and so on. Actions are caused by agents' intentions (Davidson, 1963), which rationalize them, or goal states, which can explain them. Therefore, to explain why an observed action has been performed, one needs to understand the intentions that motivated them, or the goals they aimed to bring about (Csibra & Gergely, 2007).

When taking an "intentional stance" (Dennett, 1987), humans happily generate mentalistic explanations even for the behaviors of beings who clearly do not actually possess mental states. This was famously illustrated by Heider and Simmel's (Heider & Simmel, 1944) seminal finding that people are willing to attribute complex, even sociomoral motivations to simple animated geometrical shapes, at a time in history when cartoons were not as widespread as they are now. After viewing video clips of black triangles and circles locomoting in particular ways, participants described what they had seen by referring to the shapes' underlying psychological states, as these could best explain the events in the stimuli.

However, action understanding, like for instance visual perception, is an inverse problem (Csibra & Gergely, 2007): As there is no direct evidence for the presence and nature of the psychological entities that produce observable movement patterns, they have to be inferred, much like the presence of objects from sensory data (Heider, 1958). Inverse problems don't have analytical solutions, but the solutions can be estimated by statistical methods and/or by adopting various prior assumptions.

One line of theories has proposed that people do this by running a similar process that would be used to plan an action in an inverse direction (Baker et al., 2009). To do so, they rely on the "principle of rationality", which specifies that intentional agents will choose those actions that bring about their desired goal states as efficiently as possible, given their beliefs about the world. In reverse, this entails ascribing those beliefs, desires, and goals to an agent that would rationalize the action the agent performed, given the state that she brought about. Because a large number of intentions may be compatible with any given action, inverse planning is proposed to be a Bayesian process that integrates the observed evidence with prior knowledge about the agent's mental states and situational factors (ibd.).

Research suggests that even at a very young age, infants can use the principle of rationality for action understanding, in a process that has been called "teleological reasoning" (Gergely & Csibra, 2003). They can reconstruct a goal by hypothesizing what action a rational agent would perform to bring about this goal. For instance, after viewing an agent (e.g., an animated ball, a hand) repeatedly approach a target by crossing a barrier, infants as young as 3 months of age expect the agent to modify her behavior in a novel context where the barrier is absent, and approach the target in an efficient, straight path, rather than continuing to perform the same motion trajectory (Gergely et al., 1995; S. Liu et al., 2019). On the other hand, if the agent initially acted inefficiently, this action could not rationalize the goal of approaching the target, so infants failed to generate such an expectation for future behaviors. This kind of teleological reasoning has proven inferentially rich and allows extrapolating goals, actions, and situational constraints an agent operates under (Csibra et al., 2003).

Importantly, this type of folk psychology does not require that people actually always behave rationally or efficiently (in fact, agents frequently make biased, "irrational choices"; see Kahneman & Tversky, 1979). It is sufficient that the model generally approximates how agents tend to behave so as to be useful in action understanding and prediction. Particularly in the domain of motor planning, people tend to select behaviors that are least effortful (Paulus & Sodian, 2015; Todorov, 2004; von Hofsten, 1980; Wolpert & Landy, 2012).

The Bayesian inverse planning process that is hypothesized to underlie infants' and young children's intuitive psychology has been described as a "naive utility calculus" by Jara-Ettinger and colleagues (2016) (for a related approach, see Lucas and colleagues' model of the "child as econometrician"; Lucas et al., 2014). They argue that, from early on, infants see other people as utility-maximizers, who choose action plans by which they can maximize the rewards obtained from goals while minimizing the action costs required to bring them about (i.e., acting efficiently).

The naive utility calculus (henceforth: NUC) formalizes intuitions about rational agency and makes precise quantifiable predictions. Research in this framework has in recent years accumulated substantial evidence that such a calculus indeed captures children's commonsense psychology across a range of domains (Aboody et al., 2021, 2022; Bridgers, Jara-Ettinger, et al., 2020; Jara-Ettinger, Gweon, et al., 2015; Jara-Ettinger, Tenenbaum, et al., 2015; Jara-Ettinger et al., 2020).

In preverbal infants, too, there is recent evidence that a NUC-like process allows them to draw rich inferences from observations of behavior. As mentioned before, infants expect agents to minimize action costs by minimizing the path length or jump height (Csibra et al., 1999, 2003; Gergely et al., 1995; Kamewari et al., 2005; S. Liu & Spelke, 2017; A. T. Phillips & Wellman, 2005; Sodian et al., 2004) and the number of steps in an action sequence (Scott & Baillargeon, 2013; Southgate et al., 2008). Moreover, they can deduce that an agent who is willing to invest more effort in reaching object A than object B has a relative preference for A over B (S. Liu et al., 2017), and similarly, that an agent who incurs a greater risk to reach one object compared to another will prefer the former (S. Liu et al., 2022). They can also compute and compare the costs of different action types (Pomiechowska & Csibra, 2020) and ascribe agent-specific cost functions from observing patterns of behavior (Pomiechowska & Csibra, 2022).

The NUC predicts that infants and children will expect knowledgeable agents to select, from among the available action plans, the option that will maximize expected utility (Jara-Ettinger et al., 2017). It seems to be a commonly held intuitive assumption that this is how infants ascribe goals and generate predictions of future actions. However, the question whether infants have a concept of choice which leads them to predict that an agent will choose the highest-utility alternative has rarely been addressed explicitly.

Infants' attribution of the concept of choice

A seminal paradigm for studying infants' goal-directed action understanding is the Woodward task (A. L. Woodward, 1998). In this task, participants are familiarized with an agent repeatedly approaching one of two available objects, whose locations are switched at test. Numerous experiments have found that infants expect the agent to continue approaching the same object in a different place, demonstrating that they encode the event's end state in terms of the acquired objects rather than in terms of its location (e.g., Eason et al., 2018; Luo & Baillargeon, 2005; Y. A. Shimizu & Johnson, 2004; Spaepen & Spelke, 2007; A. L. Woodward, 1999; A. L. Woodward & Sommerville, 2000). This suggests that they take the agent to aim to bring about a particular goal state, rather than merely perform a repetitive movement. In contrast, when there is only one object present during familiarization, infants do not hold an expectation for what an agent will approach when a novel second goal object appears at test (e.g., Y. Choi et al., 2018; Luo & Baillargeon, 2005). They only do so when the agent efficiently adjusts her behavior to environmental constraints in approaching the single object (Biro et al., 2011; Hernik & Southgate, 2012).

The standard explanation for this pattern of results is that infants ascribe a relative preference to the agent when she chose one of two available options during familiarization (Baillargeon et al., 2015; Y. Choi & Luo, 2023). If, on the other hand, there was only one object that could be reached for, they are naive with regards to how much the agent values that object relative to a novel object. Under such an account, infants in the two-objects familiarization condition would consider both objects as potential goals for the agent, use the agent's behavior during familiarization to infer which option yields relatively higher rewards for her, and thus assume that the agent will continue to choose the higher-value preferred object, even when it has moved to a different place. It is implied that infants in some way also represent the non-chosen option, if only to decide that the agent considers it relatively worse and will thus not choose it.

An alternative proposal has been put forward by Robson and Kuhlmeier (Robson & Kuhlmeier, 2019). They argue that rather than representing all goal options available to an agent, along with a preference ranking, infants may parsimoniously represent the goal of the agent with just as much detail as is needed in a given context (cf. Hobbs & Spelke, 2015). If there are two objects in the scene, infants may store specific features of the grasped item in order to be able to individuate it later, whereas if there is only a single object available at the time when the agent reaches for it, they may simple represent it as a featureless object (see Kibbe & Leslie, 2013 for the argument that such a representation is "cheaper"than one containing featural information). In support of this proposal, Robson and Kuhlmeier showed that infants, familiarized with an agent consistently approaching object A over B, ascribe goal A to the agent even when it is contrasted at test with novel object C (cf. Cesana-Arlotti et al., 2020). If infants interpreted the familiarization in terms of the agent displaying a preference for A over B, they should not hold an expectation at test, as they have no evidence for how much the agent values C relative to A. Relatedly, Feiman, Carey, and Cushman (2015) found that when infants are familiarized with an agent reaching for object A over a foil object which changes every trial, they expect the agent to continue reaching for A when it is paired with another novel object at test. This also speaks against the conjecture that infants in the Woodward task default on interpreting the agent's behavior as an expression of relative preference, unless they attributed a higher-order preference ("the agent likes A more than anything else"). It is however unlikely that merely approaching A over B, as in Robson and Kuhlmeier's experiment, would be sufficient to induce such a higher-order preference attribution. On the other hand, 7- and 14-month-olds in Feiman and colleagues' (2015) study did not attribute the goal of avoiding a particular object D to an agent, who always reached for a variable foil over D. This, too, is consistent with the idea that infants may simply not represent features of non-goal objects in the scene.

While these findings lend credibility to the claim that infants may not always and by default interpret an agent's reach to one of two objects as an expression of choice, other studies demonstrate that under certain circumstances, infants do construct preference rankings from observing agents' patterns of goal approach. Robson et al. (2014) found that 9-month-olds inferred an agent's preference hierarchies. They familiarized infants with an agent reaching for one of two varying object pairs: In some trials, objects A and B were present and the agent grasped A; others featured objects B and C and the agent grasped B. This could be seen as inconsistent behavior (as the agent sometimes pursues goal A and sometimes B), unless one appreciates that goal choice depends on context and that the agent has relative preferences. Going beyond this, Mou, Province and Luo (2014) showed that at 16 months, infants can use such information to draw transitive inferences (agent prefers A over C). Duh and Wang found that 14-month-olds ascribed to an agent a preference of relatively larger objects (Duh & Wang, 2019). Finally, a study by Liu et al. (2017) demonstrated that infants employ utility reasoning to infer how much an agent values a goal (relative to another) from how much effort he is willing to invest to reach it. The agent is shown to approach goal A at a low cost, but refuses at a medium cost, and approaches goal B at a medium cost, but refuses at a high cost. At test, infants expected him to approach B over A when both options are available (see Liu et al., 2022, for a similar design, but with risk—or counterfactual cost—as the variable from which infants infer preference). Note that this study also suggests that infants can appropriately interpret a non-action as refusal or choice not to act (cf. Behne et al., 2005).

If these studies show what they purport to show, infants have to go beyond merely ascribing a goal to an agent, and require a concept of choice. In terms of the cognitive operations needed for this, (1) infants must set up the alternative options for goals as well as means for reaching them. These have to be options from the perspective of the agent, i.e., the agent has to have epistemic and physical access to them (Y. Choi et al., 2022; Kampis et al., 2013; Luo & Baillargeon, 2007; Luo & Johnson, 2009) (2) infants have to calculate the expected utility for all of the options, and (3) compare these values to assess which option would yield the highest utility. Note that (1) is not a trivial task, as it is unclear how infants determine the relevant set of options: There is an infinite number of goals someone might pursue at a given time, and calculating the utility for all of them is an intractable task (Bear et al., 2020; J. Phillips et al., 2019; Smaldino & Richerson, 2012). (2) requires a simulation of the potential actions unfolding, so that infants can assess what the overall expected utility for each option would be. (3) necessitates that the respective expected utilities of the relevant options are kept in working memory, to allow the interpretation of ongoing and the prediction of future behavior.

It is an open question under what circumstances infants interpret an agent's behavior as a choice among alternatives, and what prompts them to do so. The issue under consideration relates to an ongoing debate on the developmental origins of modal reasoning (Leahy & Carey, 2020; Redshaw & Ganea, 2022). On the one hand, there is evidence that already preverbal infants can represent mutually exclusive possibilities (Cesana-Arlotti et al., 2018, 2020; Téglás et al., 2011). On the other, even preschoolers surprisingly fail at tasks that ought to be trivial for them if they were able to take multiple possible outcomes into consideration (Mody & Carey, 2016; Redshaw & Suddendorf, 2016). Clarifying the mechanism by which infants ascribe goals, and whether they do so by invoking choice, can make an important contribution to this larger project.

Choice of the higher-utility option

In the aforementioned studies ostensibly demonstrating infants' capacity to ascribe choice, the information about how much subjective value an agent places on the different options is derived from the agent's behavior. Thus, infants infer the utility function by considering which option is chosen in a given context or whether or not an agent is willing to incur a certain amount of cost to reach a goal. It is unclear whether infants can perform the inverse of this operation: If information about the relative utility of different options is provided, will infants assume that the agent will choose the best one?

This is one way to interpret the studies on infants' teleological reasoning. According to such a reading, infants would succeed in tasks like the one by Gergely et al. (1995) by considering a range of possible actions (or means) the agent might perform to bring about the goal, and expecting the agent to select the one that is least costly. The agent could in principle, for instance, choose whether to move towards the goal in a straight line, jump, detour horizontally, zig-zag, moonwalk, and so on; and, following the principle of rationality, should decide to perform the most efficient action. The shortest path is the optimal choice here because, holding other factors constant, it is the least costly for the agent. This assumption can be revised by other available information: For example, if the agent is a frog or the floor is lava, she may prefer moving in a different way; (cf. Pomiechowska & Csibra, 2022).

An alternative account would be, for instance, that infants generate an ideal reference trajectory for the action and detect any behavior that deviates from it as inefficient (Hudson et al., 2018; McDonough et al., 2019). This predicted trajectory can be simulated based on infants' previously established knowledge of the agent's goal and their understanding that he is a rational actor, as well as their assessment of the environment and the constraints it poses (Bach & Schenke, 2017; Csibra, 2008a). Thus, infants would not need to *a priori* consider and compare alternative possible options as such, so they could perform teleological reasoning even without this capacity. Note that the generation of the reference trajectory would be probabilistic, such that infants can tolerate small deviations from it without detecting an inconsistency. In fact, in some prior studies on teleological reasoning, the agent does not cut as close to the obstacle as would be possible (see e.g. S. Liu et al., 2017; S. Liu & Spelke, 2017)

With the experiments described in the following chapters, we investigated whether infants would expect an agent to choose an option that was in some way "better". We presented infants with an agent pursuing a goal (i.e., he efficiently approached a particular object), who at test faced a choice between two plausible alternative goals, one of which would yield a relatively higher utility to the agent. We then used a looking-time (Experiment 1.1 and 1.2) or eye-tracking measure (Experiment 1.3) to assess whether infants would find it more plausible that the agent approaches the "better" goal. Success in this task would demonstrate that infants possess the capacities previously described: They could identify and represent the relevant alternatives, compute their respective expected utilities, and adjudicate which is best. Merely simulating an ideal trajectory to the only available goal would not be sufficient.

Of course, what is "better" or how valuable something is to someone varies considerably between individuals. For this reason, we set up the alternatives in our studies to convey utility differences that we hypothesized infants would take to be agent-invariant, so that they need not learn the agent's idiosyncratic preferences from previous behavior. In Experiment 1.1, the "better" option consisted in a relatively higher number of goal objects (food items) to be acquired. If infants reason about utility in terms of energetic costs and benefits, then it is a reasonable conjecture that—with the cost of reaching either option being equal—more of a desired (edible) resource is preferable. In Experiments 1.2 and 1.3, the "better" option was one of two identical items that could be reached with less effort, where effort was a function of the height of a barrier (Experiment 1.2) or the path length around an obstacle (Experiment 1.3). Here, the less costly option should be preferable to any rational agent.

Chapter 1.1: Do infants expect an agent to choose “more”?

Experiment 1.1

With Experiment 1.1, we aimed to investigate whether infants have prior assumptions about reward value, specifically that *more* of a good thing is better, and leverage these assumptions to predict what an agent will choose in a context where goal objects are available in different quantities.

While infants’ expectations of efficient action have been well established, little research to date has focused on their reasoning about the benefits or rewards of action goals. Some studies have examined the interdependency of the two: The fact that rational agents will only perform an action when the cost of doing so is outweighed by the reward of reaching the goal can be used by an observer to infer how valuable at least an agent deems an outcome. This logic underlies the finding by (2017) that 10-month-olds could infer the relative reward value of different outcomes from how much cost an agent was willing (or refused) to incur (S. Liu et al., 2022). A study by Skerry and Spelke (2014) indicates that 8- and 10-month-olds take goal completion to be an outcome that an agent would deem positive, and subsequently respond accordingly. Smith-Flores and colleagues (Smith-Flores et al., 2024) found that infants even hold similar assumptions for vicarious emotions among friends, i.e., expecting an agent to show a positive response when her social partner successfully reaches her goal.

Some studies have investigated whether infants generally assume that goal rewards are agent-specific, such that a choice reveals something about an agent’s idiosyncratic preferences, or agent-independent, in which case the choice would convey that some objects are objectively preferable. This research has yielded mixed results (Buresh & Woodward, 2007; Egyed et al., 2013; Henderson & Woodward, 2012; Kampis et al., 2013). A seminal study by Repacholi and Gopnik (1997) found that in the second year of life, infants come to understand that other people can like different things than themselves: Eighteen-, but not 14-month-olds handed the experimenter a type of food the experimenter had previously expressed a preference for (broccoli), rather than the food they themselves preferred (though see Ruffman et al., 2018).

Overall, infants and toddlers seem to have some assumptions about the rewards that are associated with reaching a goal. However, in all these studies, infants use information about the agent’s behavior (e.g., investing effort, showing a particular response) to draw inferences about the value another agent expects to obtain. These cues serve as reliable indicators of an agent’s utility functions. It is an open question whether infants can extract information from the goal object itself to infer its value or predict an agent’s behavior.

This would be useful, as infants could generate predictions about what an agent will choose, without first having to observe her choices across different contexts. Further, such priors about benefits of different goals could allow the observer to assign precise reward magnitude to the options available in a novel context and derive a ranking of the agent's likely preference.

One indicator of benefit that is agent-invariant, straightforward, and could be cognitively accessible for even a young infant is the quantity of goal items. "More is better" may hold primarily in the domain of food: While having five cell phones may be no better than having one (and may actually make life more stressful), having more of a preferred edible item confers caloric benefits, which are crucial for a biological organism. This hypothesis is complementary with the idea that action costs, in their basic form, are energetic costs.

A reason to hypothesize that infants may expect another agent to prefer more comes from the fact that they themselves do. The cracker test, devised by Feigenson and colleagues (2002), had 10- and 12-month-old infants choose between two opaque containers, into which different quantities of crackers had been sequentially placed while the infant was watching. Infants reliably crawled towards the container which held the larger number of crackers (constrained only by their limitations to track larger numbers: As soon as one of the containers was baited with four or more crackers, infants failed the task). When infants had a choice and could assess the numerical quantities, they behaved as utility-maximizers by going for *more*.

Another line of evidence supporting the conjecture that infants may think that getting more is better comes from research investigating their intuitions on distributive fairness. Here, infants expected agents to distribute resources equally (Schmidt & Sommerville, 2011; Sloane et al., 2012), preferred to interact with fair distributors (Geraci & Surian, 2011; Lucca et al., 2018), and expected others to act positively towards fair over unfair distributors (DesChamps et al., 2016; Meristo & Surian, 2013, 2014). One way to interpret these findings is as indicating that infants take the receipt of a resource as a positive outcome, and hence that possessing more items of this resource is better than fewer.

In the present study, we tested whether infants share the following assumption: If the cost for reaching two possible goals is kept constant, and both goals contain tokens of a goal object type which an agent had previously approached, then the agent can maximize her utility by choosing the goal that contains a relatively higher quantity of the object. We familiarized 10-month-old infants with a Woodward task scenario in which an agent always approached the same one of two different available objects (a banana or a strawberry). At test, the agent faced a choice between one or three tokens of the previously selected object. If infants think that the agent will approach the larger quantity, they should look longer when the agent selects the lower-quantity goal. On the other hand, if infants do not have a default expectation that an agent will choose the highest available number of goal items, they should respond as in a classical Woodward task: The three objects may be represented by infants as a novel goal for the agent.

Methods

This experiment was preregistered at the OSF (https://osf.io/9h7yg).

Participants

Twenty-four 10-month-old infants (10 male, age range: 9 m 16 d - 10 m 12 d, mean age: 10.03 m) participated in Experiment 1.1. An additional 9 infants were tested but had to be excluded due to experimenter error (n = 4), failure to meet the predefined attention criteria (n = 2), fussiness (n = 2), or parental interference (n = 1). Participants were healthy, full-term infants who were recruited through a local database and received a small toy at the end of the testing session. The study received full ethical approval from the United Ethical Review Committee for Research in Psychology (EPKEB) in Hungary and was conducted according to the principles in the Declaration of Helsinki. Written informed consent was obtained from the infants' parents before the experiment.

Apparatus

Infants were seated in their caregiver's lap in a darkened, soundproof room, 80 cm away from a wide-screen 102 cm LCD monitor. The stimuli were 3D animated videos created with Blender animation software (https://www.blender.org) and presented from a Mac running with OS X Yosemite 10.10.4 with MATLAB (Mathworks) using the Psychophysics toolbox extension (Brainard, 1997; Kleiner et al., 2007). Infants were recorded during the session, and an experimenter watched the video live for online coding to determine the termination and onset of stimuli trials (infant-controlled procedure).

Procedure and stimuli

Caregivers were instructed to hold the infants by their hips without impeding their ability to attend or disengage from the screen. Caregivers' eyes were covered with opaque sunglasses. Before each trial, a short attention-getting clip (a randomly selected combination from 3 possible visual and 3 possible auditory cues) was shown until the infant attended to the screen. Trials ended either when the infant looked away for a minimum of 2 consecutive seconds after a video had stopped, or if 8 seconds (familiarization) resp. 60 seconds (test) had passed since a video ended.

Familiarization. Infants watched a total of 8 familiarization trials. Each trial consisted of a video that had a length of 7.5 seconds, and the display of a still image of the video's last frame. In all videos, an agent approached a goal object.

The videos started with a scene in which an agent (a pear-shaped blue figure with eyes) was located at the top of the screen, and two goal objects located at the bottom, on the left and right side respectively. The goal objects were a banana and a strawberry, slightly smaller than the agent himself. The scene contained obstacles on the left and right side which created a funnel-shaped clearing in the center. In the videos, the agent moved downwards in a straight line between the obstacles (3 s), then turned left or right and approached a goal object (2.5 s) and finally came to a standstill after making physical contact with the object while a ringing sound was played (2 s).

The agent always approached the same type of goal object, which was sometimes located on the left, sometimes on the right side of the screen (order: LRRLLLRR or RLLRRRLL).

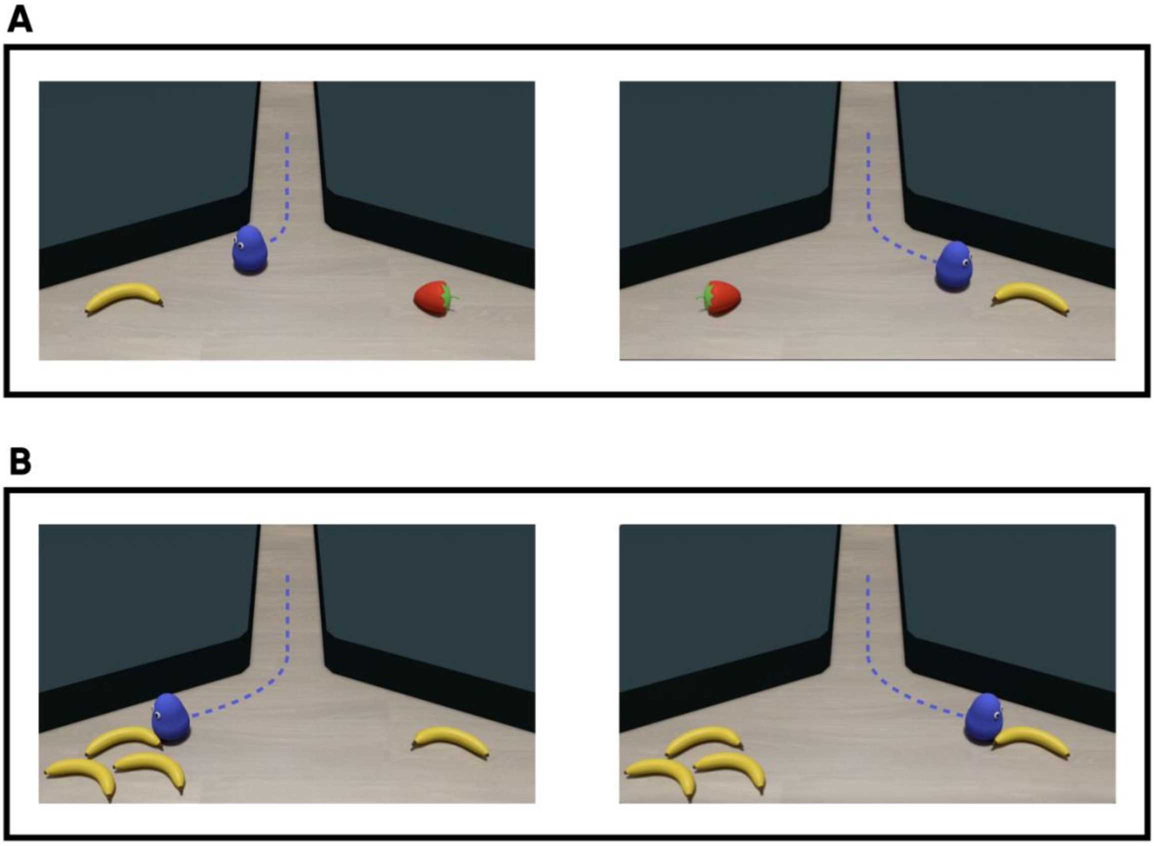

Figure 1.1. Stimuli used in Experiment 1.1. During Familiarization (A), the agent approached one of two available goal objects (e.g., the banana), which was sometimes located on the left, sometimes on the right side. At Test (B), the agent either approached three tokens of the goal object (Consistent test event, left) or a single token (Inconsistent test event, right).

Test. Two test trials were shown to participants. The videos were identical to the familiarization videos in terms of duration, behavior of the agent, and layout, apart from the goal objects. Now, there were only tokens of the previously approached goal object type present; however, on the one side, there was a single item, on the other, three items. The three items were close to each other but not touching (so that they could clearly be distinguished as separate objects). The single item and the topmost item from the group of three were located equidistant from the agent. The agent approached either the three items (Consistent test event) or the single item (Inconsistent test event).

Stimuli can be accessed at https://osf.io/6pf3b.

We counterbalanced four factors: the type of goal object approached (banana vs. strawberry); the order of the locations which were approached (LRRLLLRR vs. RLLRRRLL); the location of the three items at test (left vs. right); and the order of test events (approach-3 first vs. approach-1 first). The counterbalancing was not fully orthogonal, as this could not be accomplished with 4 factors in a sample of 24.

Coding and analyses

Infants' looking behavior was manually coded off-line to measure looking times using the same criteria as online coding and reviewed for the pre-registered exclusion criteria (fussiness; parental interference; experimenter error; lack of attention to crucial events). The looking times of 50% of the participants was reanalyzed by an independent second coder who was blind to the hypothesis and to the condition of the stimuli shown. The average absolute difference between coders was 0.28 s. Because of this high level of agreement, data from the first coder was used for analyses (in this and the following experiments).

The raw looking times were base-10 log-transformed for analyses (Csibra et al., 2016), but raw data are used for descriptive statistics and plots. As specified in the preregistration, we conducted both Bayesian and frequentist statistical analyses. For the Bayesian analysis, we used the method recommended by Csibra et al. (2016) for looking-time data. This method calculates a Bayes Factor which compares a H_1 assuming a moderate increase or decrease in looking times between conditions with a null model of no difference. For the frequentist statistical analyses, we conducted a paired sample two-tailed t-test on the data. Moreover, we conducted a 2x2 mixed ANOVA to check for order effects. Such a pattern is commonly found in looking-time studies with infants, who tend to look longer at the first test video they are shown, and slightly lose interest at the second. This tendency can interact with the predicted effect of test condition (consistent/inconsistent) (Baillargeon, 1987; Csibra et al., 1999; S. Liu et al., 2017; Mascaro & Csibra, 2012; Tatone et al., 2023). Statistical analyses and plotting were performed in R, version 3.4.1 (R Core Team, 2023).

Data are accessible at https://osf.io/6pf3b.

Results

Hypothesis-driven results

The Bayesian analysis suggested that infants looked longer when the agent approached the three goal objects ($M_{approach\text{-}3}$ = 19.31 s, $SD_{approach\text{-}3}$ = 15.14 s) than when the agent approached the single object ($M_{approach\text{-}1}$ = 14.98 s, $SD_{approach\text{-}1}$ = 14.89 s). This data yielded a BF of 73.45, which indicates strong effects (Csibra et al., 2016). In a two-tailed t-test, this looking-time difference was not significant ($t(23) = 1.99$, $p = 0.054$).

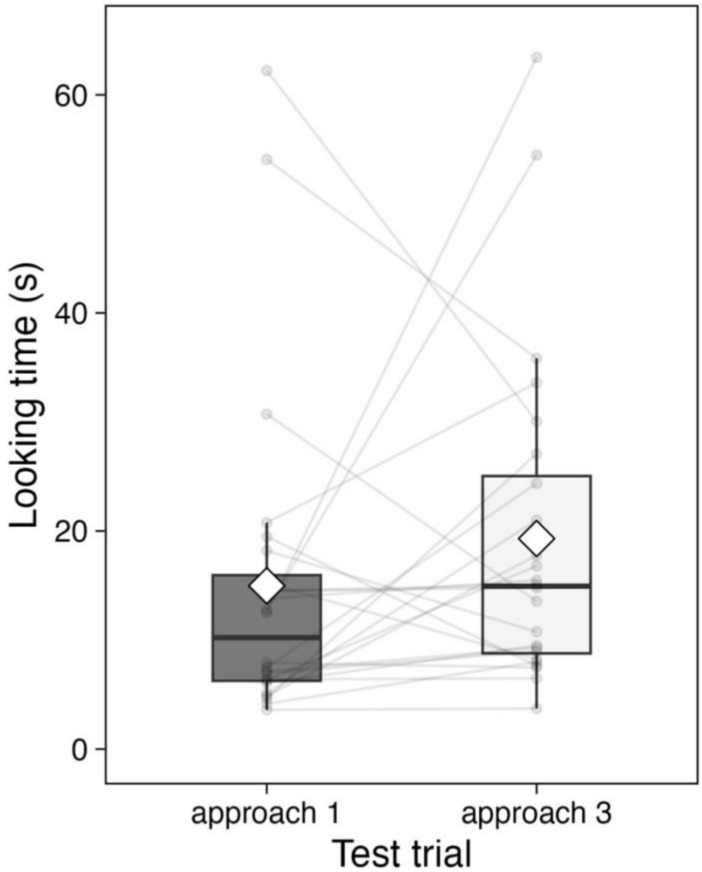

Figure 1.2. Boxplot of average looking times (in seconds) toward the test events in Experiment 1.1. Light grey lines connect the looking times of individual participants, white diamonds indicate means, horizontal lines indicate medians, boxes indicate middle quartiles, and whiskers indicate points within 1.5 times the interquartile range from the upper and lower edges of the middle quartiles.

Additional results

We analyzed whether there was an effect of the order of test events. A 2x2 mixed ANOVA did not show a significant Order x Condition interaction ($F(1,22) = 0.92$, $p = 0.348$). However, those infants who saw the approach-3 event first did look significantly longer at this ($M_{\text{approach-3}}$ = 16.03 s) compared to the approach-1 event ($M_{\text{approach-1}}$ = 9.98 s; $t(11)$ = 2.56, p = 0.027). On the other hand, there was no significant difference in looking times for those infants who saw the approach-1 event first ($M_{\text{approach-1}}$ = 19.97 s, $M_{\text{approach-3}}$ = 22.59 s; $t(11)$ = 0.63, p = 0.541).

Discussion

The results of Experiment 1.1 failed to provide evidence that infants expect an agent to maximize her utility by approaching a higher quantity of goal objects. On the contrary, infants looked longer when the agent selected more goal objects compared to when she continued to pursue the same goal as during familiarization, which consisted in approaching a single goal object. This indicates that they did not represent the larger set as a token of familiarized goal, but represented it as a novel goal.

Several explanations can account for these results. One possibility is that infants do not see quantity as a default indicator of how much an agent will value a goal: They may not assume that more of a good thing is necessarily better. Another possibility is that the novelty of the higher quantity of goal objects (which was never shown during familiarization) was disruptive and prevented infants from comparing the relative reward magnitude of the two options. If this is the case, infants may have problems using the NUC productively to reason about novel goals (at least in the domain of benefit maximization). Third, it is conceivable that infants may not even represent the scenario as a choice among alternatives. During familiarization, they might have merely attributed a goal to the agent (e.g., approach the banana) without taking into account the non-approached object (cf. Feiman et al., 2015), and at test looked longer at the outcome that was less consistent with the previously attributed goal.

To adjudicate whether infants had specific limitations when reasoning about reward magnitude, or whether they generally fail to consider and compare alternative options of varying utility, we conducted another experiment. Here, infants had to compute the relative utility of two identical goal options which could be reached at different cost.

Chapter 1.2: Do infants expect an agent to choose a goal that can be reached with less effort?

Experiment 1.2a

In previous studies on infants' assessment of efficiency in goal-directed actions, there was typically one goal towards which an agent directed her actions. As discussed before, one strategy for solving this task would be to calculate and compare the costs associated with different potential actions that could bring about the goal (e.g., move along a direct path vs. take a detour), and predict that the agent will select the one that minimizes her action cost.

Alternatively, they may simply expect agents to behave locally efficiently. Infants may anticipate that the agent will move towards his goal along a trajectory that is optimal in the given environment (Hudson et al., 2018). An observed action violates this prediction whenever it is not ideally adjusted to the relevant environmental constraints, for instance, if an agent detours with a lot more space between him and the obstacle than necessary (S. Liu & Spelke, 2017) or detours despite there not being any obstacle (Gergely et al., 1995).

In a scenario where multiple potential goals are available, such a strategy alone would not allow infants to identify the option that yields the highest utility. To do so, they would have to hold in mind and compare all alternatives. We hypothesized that if infants reason about actions directed towards a single goal by using a NUC, they should also do so when there is more than one plausible goal candidate obtainable. In their own actions, young infants succeed at monitoring the costs associated with different potential goals, and choose the least effortful one (Paulus & Hauf, 2011), or refrain from pursuing a goal they would otherwise enact once the required costs reach a certain threshold (Sommerville et al., 2018).

In Experiment 1.2, we tested whether infants would expect an agent to choose the one of two identical goal objects that could be reached with less effort. The reward from the two potential goals was constant and infants had to compare the respective cost. The experiment thus relies on the supposition that infants assume identical-looking objects to yield the same benefits.

The logic of this experiment is similar to that of a study by Scott and Baillargeon (2013). The researchers here familiarized 16-month-olds with a scene in which a human agent repeatedly selected one of two featurally identical objects (toy pigs), in a seemingly random manner. To obtain an object, the agent had to pull the handle of a platform on which it was located towards herself. At test, one of the objects was placed (by another agent) into a transparent container located on one of the platforms and covered with a transparent lid; the other object was placed in the container on the other platform but was left uncovered. The researchers found that infants looked longer when the agent selected the pig that was located in the container with the lid, which indicates that they expected her to minimize her effort by choosing the object which would require fewer steps to obtain.

We aimed to build on this finding by testing a sample of younger infants (10-, rather than 16-month-olds). In our experiment, the proxy of action cost was relative path length (rather than anticipated number of steps in the action sequence). We provided infants with direct evidence of the two relative action costs during familiarization, so that they would not have to generate the prospective costs of previously unseen behaviors.

Methods

This experiment was preregistered at the OSF (https://osf.io/pvy37).

Participants

Twenty-four 10-month-old infants (12 male, age range: 9 m 18 d - 10 m 15 d, mean age: 10 m) participated in Study 2. An additional 14 infants were tested, but were excluded due to failure to meet the attention criteria (n = 6), experimenter error (n = 2), or ceiling looking times at both test events (n = 6). Participant recruitment, compensation, consent, and ethical approval were the same as in Experiment 1.1.

Apparatus

The apparatus was the same as in Experiment 1.1, with the exception that stimuli were presented with PyHab 0.7.2 habituation software (Kominsky, 2019) in PsychoPy 3.0.6 (Peirce et al., 2019).

Procedure and stimuli

Caregivers' eyes were covered with opaque sunglasses, and they were instructed to hold the infants by their hips without constraining their ability to disengage from the screen. Before each trial, an attention-getting clip was shown. During familiarization, this was a short (2 s) clip containing a yellow moving square and a squeaky sound; before each of the two test trials, the attention-getter was a longer (15 s) clip of black and white moving geometrical pattern and a sliding xylophone sound. Each trial contained multiple instances of an event, such that the events were shown in a (quasi-) looped manner. This design, adapted from Liu et al. (2017), differs from that of Experiment 1.1, where infants saw only one event per trial and looking time was recorded during a still frame at the end of the videos. Trials ended either when the infant looked away for a minimum of two consecutive seconds, or if 46 seconds (familiarization) resp. 60 seconds (test) had passed since a trial's onset.

Familiarization. Infants watched a total of six familiarization trials. Each trial consisted of a maximum of five events (less if the infant ended a trial by looking away for two seconds before a trial ended), which each had a duration of 8.5 s. In each trial, there were two high jump events, two low jumps, and one straight (no jump) approach. After each event, a black screen was briefly displayed (0.5 s).

The scene shown in the stimuli always contained an agent (a pear-shaped blue figure with googly eyes facing forward) initially located in the middle of the screen. There was always either a low or a high wall, made of solid, dark grey material, to the left or right side of the agent. The walls were always in the same location and did not change sides within a subject (such that, for example, the low wall always appeared on the left side). Behind the walls, on the far left and right side of the screen, there were two flat, magenta disks that served as "landing pads" for the goal objects, which were yellow bananas of slightly smaller size than the agent.

Each video within a trial began with a bell sound, indicating the onset of an event (1.5 s). Then, a banana fell from the top of the screen onto one of the two magenta plates, which was accompanied by a whooshing sound (1 s). Following this, the agent turned to the side where the banana had fallen and moved towards it (4.5 s). If there was a wall in the way of the agent's path, he jumped over it. Upon making contact with the banana, the agent came to a stop, and a ringing sound was played (1.5 s). The timing was kept constant for all familiarization events (low jump, high jump, straight approach). The height of each jump was adjusted to the height of the wall, respectively.

Test. Infants watched two test trials. Each trial consisted of the same event, which was looped for a maximum of 60 s. As in familiarization, the videos within a trial were interspersed with a brief 0.5 s display of a black screen.

The layout was similar to the familiarization trials, except that now both walls (high and low) were present, in the same locations as before. The event also played out the same as in familiarization, with the exception that two bananas fell down at the same time (one onto each magenta plate), and that there was an additional 0.5 s pause before the agent started moving, to give infants a chance to attend to both potential goal objects. Test events were thus 9 s long. In the inconsistent test event, the agent approached the banana behind the higher wall, in the consistent test event, behind the lower wall.

Stimuli can be accessed at https://osf.io/7j58z.

We counterbalanced three factors: the location of the high and low walls (high left vs. high right); the side of the first approach during familiarization (LHLHLH vs. HLHLHL); and the order of test events (high jump first vs. low jump first).

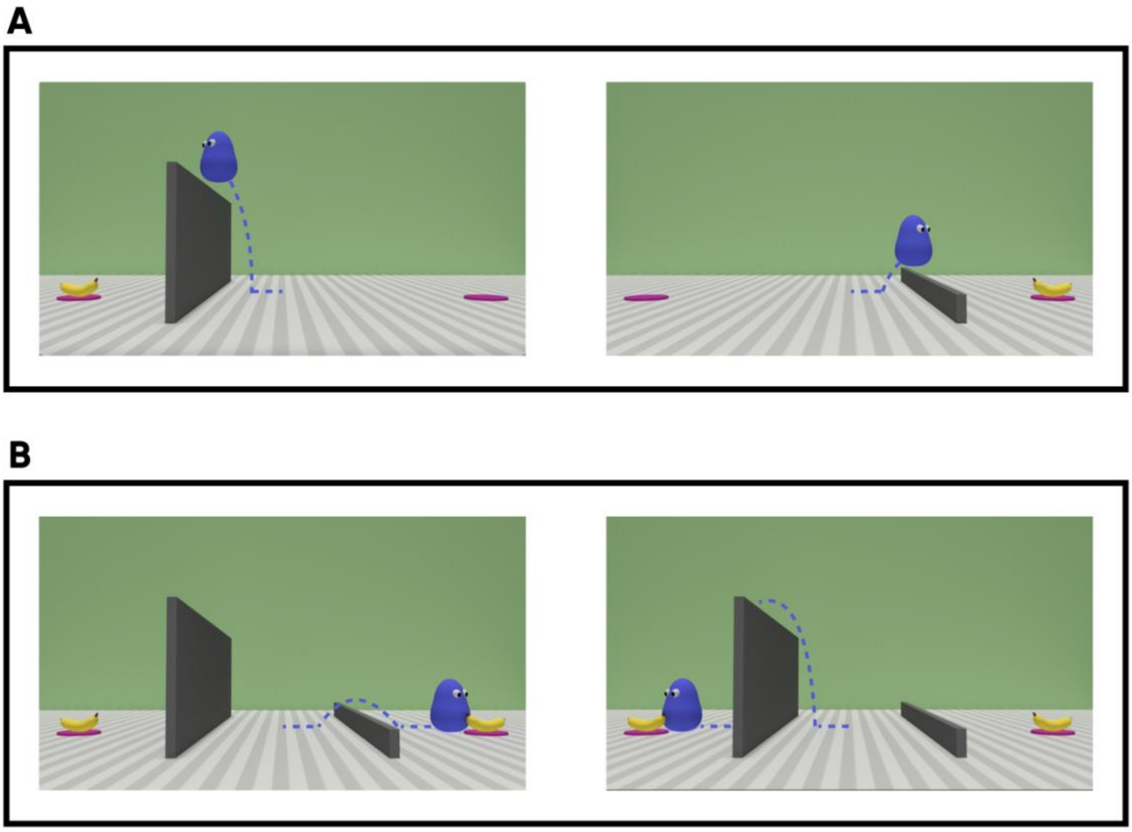

Figure 1.3. Stimuli used in Experiment 1.2. During Familiarization (A), the agent approached a goal object which was sometimes located on the left, sometimes on the right side. To reach the goal, the agent sometimes had to jump over a high (left) or low wall (right). At Test (B), both walls were in the scene and there was a goal object behind each. The agent either approached the object behind the low wall (Consistent test event, left) or the object behind the high wall (Inconsistent test event, right).

Coding and analyses

Infants' looking behaviors were coded and analyzed the same way as in Experiment 1.1. Again, a second coder, blind to the experimental condition, recoded 50% of participants' looking times. The average absolute difference between coders was 0.4 s. There was an unexpectedly high number of infants who did not disengage and look away from the screen during the experimental procedure. For this reason, we used an additional exclusion criterion that we had not preregistered: We excluded data from a participant if the participant did not end at least one of the two test trials with a 2 second look-away.

Data are available at https://osf.io/7j58z.

Results

Hypothesis-driven results

With the sample of 24 infants we tested, the Bayesian analysis provided some evidence supporting our hypothesis: We obtained a BF of 5.38. However, a t-test did not yield a significant result ($t(23) = 1.27$, $p = 0.216$), suggesting that infants did not look longer at the high jump action ($M_{\text{high jump}} = 29.93$ s, $SD_{\text{high jump}} = 15.9$ s) compared to the low jump action ($M_{\text{low jump}} = 27.13$ s, $SD_{\text{low jump}} = 20.13$ s).

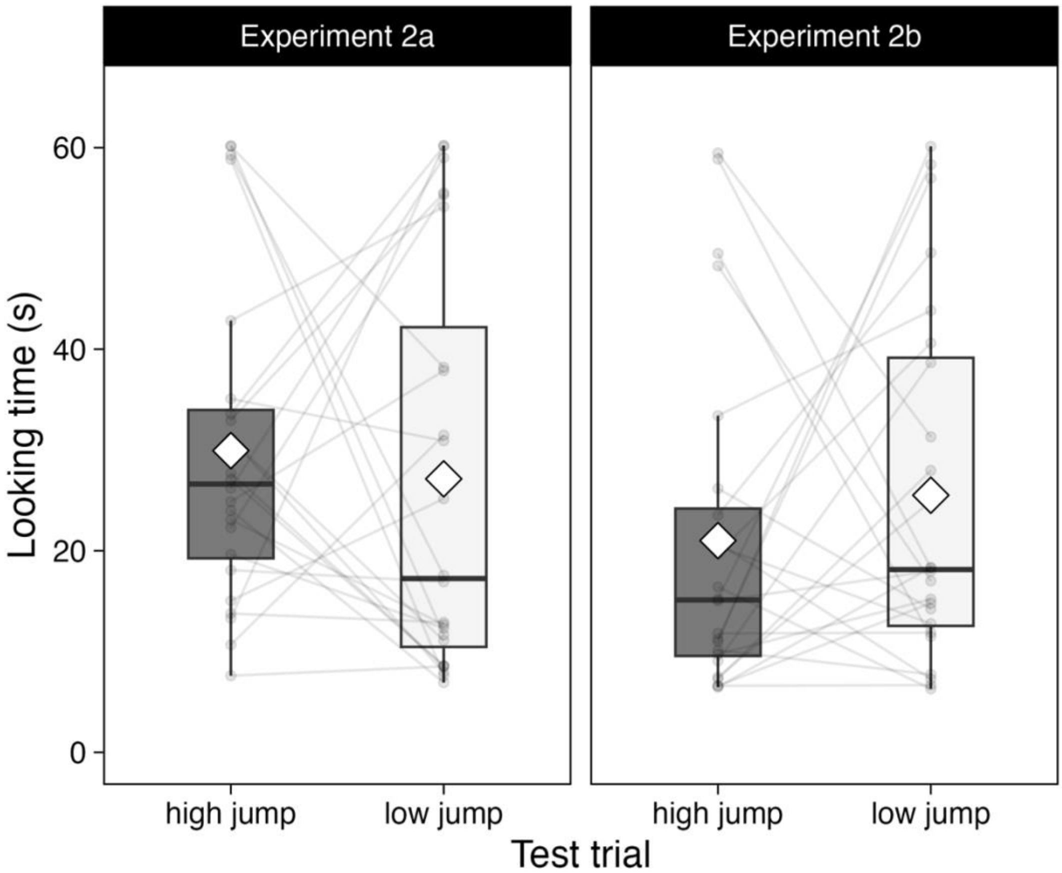

Figure 1.4. Boxplots of average looking times (in seconds) toward the test events in Experiment 1.2a (left) and 1.2b (right). Light grey lines connect the looking times of individual participants, white diamonds indicate means, horizontal lines indicate medians, boxes indicate middle quartiles, and whiskers indicate points within 1.5 times the interquartile range from the upper and lower edges of the middle quartiles.

Additional results

As in Experiment 1.1, we analyzed whether there is an effect of the order of test video presentation. A 2x2 mixed ANOVA did not show a significant Order x Condition interaction ($F(1,22) = 0.0004$, $p = 0.983$).

Discussion

The results of Experiment 1.2a were not conclusive with respect to our hypothesis. While the Bayesian analysis provided some evidence that infants indeed looked longer at the agent's high-jump action at test, and thus had expected him to choose the goal object that can be obtained at lower cost, this effect was weak (BF < 10), not supported by the frequentist analysis, and the pattern was only shown by 14 of 24 infants.

Because we used a procedure in which the test events were looped, it is possible that repeatedly playing the auditory cues that accompanied the events may have driven the infants' attention back to the screen, which prevented them from disengaging from the event even after they had lost interest. Analyzing only the data from the first looks, i.e., their looking times until they looked away from the screen for the first time, yielded a BF of 245.47 ($M_{\text{high jump}}$ = 23.55 s, $SD_{\text{high jump}}$ = 15 s; $M_{\text{low jump}}$ = 18.64 s, $SD_{\text{low jump}}$ = 16.76 s), indicating a strong effect. However, this analysis was post-hoc and cannot be considered confirmatory. Therefore, we decided to conduct a replication of Experiment 1.2a, removing the sound effects from the stimuli.

Experiment 1.2b

Methods

This experiment was preregistered at the OSF (https://osf.io/2h78y).

Participants

Twenty-four 10-month-old infants (age range: 9 m 18 d - 10 m 14 d, mean age: 10 m) participated in Experiment 1.2b. An additional 13 infants were tested but were excluded due to parental interference (n = 1), fussiness (n = 1), failure to meet the attention criteria (n = 5), technical failure (n = 5), or ceiling looking times at both test events (n = 1). Recruitment, consent, ethical approval, and compensation were the same as in the previous experiments.

Apparatus

The apparatus was the same as in Experiment 1.2a.

Procedure and stimuli

The procedure and stimuli were identical to the ones in Experiment 1.2a, except that we removed all sound cues from the familiarization and test trial stimuli.

Coding and analyses

Infants' looking behaviors were coded and analyzed the same way as in Experiment 1.2a. The average absolute difference between coders was 0.46 s. Data are accessible at https://osf.io/7j58z.

Results

Hypothesis-driven results

Infants did not look longer at the high jump test event: The Bayesian analysis resulted in a BF of 2.588, providing neither support for our hypothesis nor for the null hypothesis of no effect. The frequentist analysis also showed that looking times were not significantly different between conditions ($M_{\text{high jump}}$ = 21.01 s, $SD_{\text{high jump}}$ = 16.7 s; $M_{\text{low jump}}$ = 25.51 s, $SD_{\text{low jump}}$ = 17.59 s; $t(23) = -1.08$, $p = .29$).

Additional results

A 2x2 mixed ANOVA did not show a significant Order x Condition interaction ($F(1,22) = 2.03$, $p = .33$).

An analysis of the data from the first looks also indicates that infants did not look significantly longer at either test event (BF: 0.49, $M_{\text{high jump}}$ = 16.52 s, SD = 11.53 s; $M_{\text{low jump}}$ = 20.15 s, $SD_{\text{low jump}}$ = 16.48 s). Unlike in Experiment 1.2a, the pattern of first looks was thus not substantially different from that of the overall looking time.

Discussion

The results from Experiment 1.2b suggest that, contrary to our prediction, infants did not look longer when an agent chose to perform a costlier action over a less costly action to obtain the same reward. Analyzing the data from Experiments 1.2a and 1.2b together, a mixed ANOVA with Trial (Consistent vs. Inconsistent test event) as within-subject and Experiment (1.2a vs. 1.2b) as between-subject factors showed no significant main effects (Trial: $F(1,46) = 0.001$, $p = .916$; Experiment: $F(1,46) = 2.8$, $p = .101$) or interaction effect ($F(1,46) = 2.76$, $p = .103$), which supports this conclusion.

One possible reason for this null result is that infants did not assign equal benefits to the two identical-looking goal objects at test, which would be required for them to evaluate the relative utility of the outcomes. In fact, infants have a propensity to rationalize seemingly irrational actions (Csibra et al., 2003): For instance, in Liu et al. (S. Liu et al., 2017), infants resolved the apparent inconsistency of an agent sometimes performing a costly action (for goal A) and other times refusing to (for goal B) by inferring that goal A was more valuable to the agent than goal B. In our study, infants may have similarly reasoned that the object behind the higher wall provided a larger benefit, which made it plausible for the agent to approach this object, even though it required a more effortful action.

Under this account, both the “consistent” and the “inconsistent” test events may have satisfied infants’ rationality criteria. The actions of the agents in the two events were not equally efficient with respect to the goal description we had posited (“*reach **a** banana with as little cost as possible*”); however, since the agent only ever jumped as high over each barrier as was needed, each action was efficient with respect to the goal realized under another description (“*reach **this** banana with as little cost as possible*”).

If such an explanation is on the right track, it would provide support for the idea that infants at the age of 10 months, while they can consider whether a goal-directed agent acts locally efficiently by adjusting to relevant environmental constraints, may not explicitly represent and/or compare the utility of alternative goal options, and thus may not apply the concept of choice in interpreting and predicting an agent’s behavior.

In the next chapter, we describe an experiment where we addressed a similar question by testing older children, who previous research suggests may possess this capacity (Scott & Baillargeon, 2013).

Chapter 1.3: Do toddlers predict that an agent will choose a lower-cost goal?

Experiment 1.3

After having found no evidence for the hypothesis that infants expect agents to maximize their utility by selecting a lower-cost goal in Experiments 1.2a and 1.2b, we conducted another experiment where we tested this hypothesis with a sample of older participants. We had recruited 10-month-old infants in Experiments 1.1 and 1.2 as this was the age targeted by Liu et al. (2017), in whose experiment infants had to perform sophisticated inferences about costs and relative reward magnitudes. However, convergent evidence from different bodies of research raises the possibility that the operations required by the tasks we posed to infants may develop only at a later age. First, the study by Scott and Baillargeon (2013), which was conceptually similar to our Experiment 1.2, tested 16-month-olds. Second, in a similar study as Liu et al. (2017), probing infants' expectation that agents ought to minimize risk, i.e. hypothetical cost, participants younger than 1 year did not consistently show the same effects as those above this age threshold, hinting at a possible developmental change (S. Liu et al., 2022). Third, infants' reasoning about distributive fairness, which requires comparing unequal and equal payoffs obtained by different agents, has been shown to undergo changes between the first and second year of life (Geraci & Surian, 2011; Sommerville et al., 2013; Ziv & Sommerville, 2017). Fourth, Cesana-Arlotti, Varga and Téglás (2022) found that the pupil diameter of 14-month-olds, but not 10-month-olds, increased when they had to retain the multiple possible identities of a semi-occluded object in working memory. Taken together, these different findings tentatively point to the possibility that infants' capacity to keep in mind and compare alternative options may emerge only from around the first birthday (but see S. Liu et al., 2017).

Besides changing the target age range of our participants, we used a different method to assess their reasoning. The looking time measure we previously employed is potentially susceptible to the issue that infants could have rationalized the "inconsistent" event in unexpected ways, thus yielding null results. In Experiment 1.2, infants may have interpreted the familiarization events as showing that the two goal objects (left banana and right banana) possess different reward magnitudes corresponding to the amount of cost required to reach them (i.e., the banana behind the high wall is better, such that jumping over the wall is worth it). It would thus not be irrational to approach the high-cost, high-reward over the low-cost, low-reward option when both are simultaneously available. Infants may have easily generated such an explanation when the agent selected the former during the test events, and therefore not responded with longer looking.

In Experiment 1.3, we instead used eye-tracking to record toddlers' gaze during an unfolding action scenario, to assess whether they would preferentially look at the "better" goal candidate. A large body of prior research has found that infants direct their anticipatory gaze towards the goals of familiar actions, such as a hand reaching for an object (Cannon & Woodward, 2012; Falck-Ytter et al., 2006; Hunnius & Bekkering, 2010; Kochukhova & Gredebäck, 2010), with more mixed results when the action is performed by nonhuman agents, such as a mechanical claw, robot, or animated animal (Adam et al., 2016; Adam & Elsner, 2020; Biro, 2013; Kanakogi & Itakura, 2011; Manzi et al., 2023; Paulus et al., 2011).

There has been a debate on whether infants, after familiarized with a Woodward task, really anticipate goals or rather frequently approached locations (Cannon & Woodward, 2012; Ganglmayer et al., 2019; E. Y. Kim & Song, 2015). Paulus and colleagues (2011) also concluded that there is a primacy of frequency information in driving infants' anticipatory gaze, which trumps efficiency considerations. In their experiment, 9-month-old infants anticipated that an agent would take the same longer path she had taken during familiarization, rather than a shorter path that became newly accessible. Even adults, whose possession of teleological reasoning capacities is relatively uncontroversial, anticipated the longer, familiar path in the first test trial. Goal-directed anticipatory gaze may reflect a more effortful cognitive operation than location-based anticipatory gaze: The former was found to be slower than the latter (Krogh-Jespersen & Woodward, 2014). Taking these factors into consideration, we avoided the confounding effect of frequency information by familiarizing infants with both the longer and the shorter path. Moreover, we decided to record the proportion of infants' looks to the "low-cost" and the "high-cost" goal options during a pre-specified time window, rather than their first fixation to a target. We remain agnostic with regards to whether our measure reflects anticipation (for which the first fixation has been taken as a proxy), exploration, preference, or something else.

In our experiment, toddlers were familiarized with a scenario where an agent approached a goal object that was located behind a door. Sometimes the door and object were at a relatively small distance from the agent, sometimes at a greater distance on the far side of the screen. At test, both doors and objects were present (in the same locations as before). The agent started moving, but the video ended before it became clear which object he would approach. From this point, we measured where infants looked on the screen. We predicted that if they expected the agent to choose the object which would require less effort to reach, they should look more at the closer object, which would constitute a better goal candidate.

However, if we obtained this result, an alternative, simpler explanation could also account for it: Specifically, participants may have simply looked at the agent at the onset of the trial (who was located on the far-left side of the screen), then let their gaze wander across the screen, and fixated on the first item they encountered. To rule out this alternative, we also devised a control condition: Here, at test, the goal objects were removed, while the doors remained in the same locations as in the other test event. Given that without a goal, there is no utility to be maximized, we predicted that if infants rely on the NUC to interpret the agent's actions they should direct an equal amount of looks at the closer and the further door. On the other hand, if the simpler explanation is correct and visual attention and saliency drive infants' looking behavior in this task, they should behave similarly in the two conditions ("two-goals" and "no-goal") and look at the closest relevant element of the display (goal object or door) in each case.

Methods

This experiment was preregistered on the OSF (https://osf.io/cws8z).

Participants

We preregistered a Bayesian stopping rule to determine sample size, that is, data collection was to be concluded once one of the following criteria was satisfied: (i) either we collect 48 valid data sets or (ii) the Bayes Factor in the cross-condition comparison (see "Analyses") becomes equal to or greater than 10. The Bayes Factor calculation was first performed after collecting 16 valid samples, and after every 8 valid samples thereafter.

The final sample consisted of 48 14- to 16-month-old infants (27 male, mean age: 14.64 m). An additional 28 infants participated in the experiment but were excluded from the analysis for fussiness or inattentiveness, i.e. failing to provide a sufficient amount of valid on-screen data (n = 22), parental interference (n = 1), and due to being erroneously tested although the stopping rule criterion had already been met (n = 5). The study received full ethical approval from the University's Psychological Research Ethics Board (PREBO). Participant recruitment, consent, and compensation were the same as in the previous experiments.

Apparatus

Participants were seated in their caregiver's lap in a darkened, soundproof room, 60 cm away from the monitor. Their gaze was recorded using a Tobii Pro Spectrum Eye Tracker with an integrated 23.8-inch-diagonal monitor (resolution: 1920 × 1080; refresh rate: 60 Hz). A custom-made Python program[3] building on PsychoPy 2021.1.3 (Peirce et al., 2019) was used for calibration, presenting the stimuli, and collecting the gaze data. The stimuli were 3D animated videos created with Blender animation software.

[3] Written by Barbu Revencu.

Procedure and stimuli

Caregivers wore opaque glasses for the duration of the experiment and were instructed to hold their child by the hips without impeding their ability to attend or disengage from the screen. Before each trial, an attention-getting clip was shown.

Familiarization and test trials were presented in alternation in a blocked design, containing both two-goals and no-goal test trials. Infants were first shown four familiarization trials (short and long goal approach, in ABBA sequence), two of each type, then a test trial (two-goals or no-goal), then two familiarization trials (in AB sequence), then a test trial (the other type than was shown before), and so on (ABBA-X-AB-Y-BA-X-AB-Y…). Thus, they watched a maximum of 20 trials including 6 test trials, 3 per type.

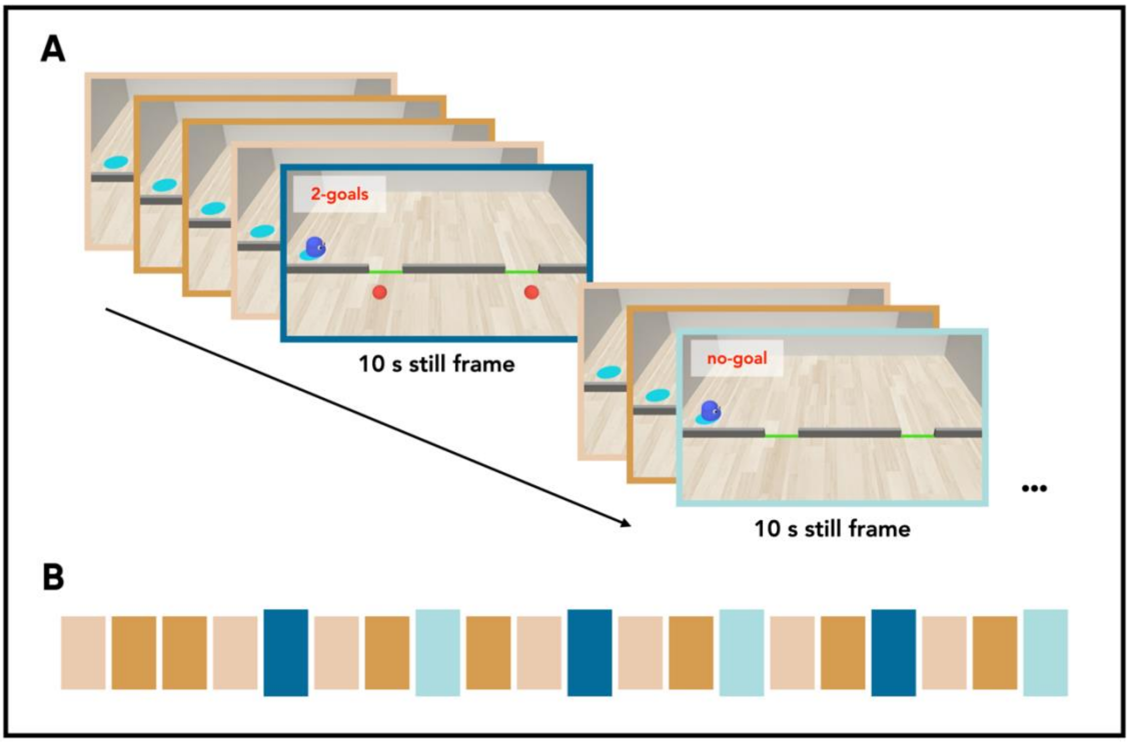

Figure 1.5[4]**.** Schematic illustration of the testing procedure in Experiment 1.3. (A) Familiarization and test trials were presented in blocks: Sets of familiarization trials were interspersed with test trials, in which infants' gaze was recorded during a 10 second still frame at the end of the video. (B) Infants received a maximum of 20 trials. Familiarization trials (long path and short path goal approach) are represented by tan rectangles; test trials (two-goals and no-goal) are represented by blue rectangles. Infants saw up to 6 test trials, 3 per type, presented in alternation (light blue vs. dark blue rectangles). The order of familiarization and test trials was counterbalanced.

[4] Figures for this experiment, as well as Experiment 2.3 and 3.1-4, were created using the Wes Anderson Palettes R package (https://github.com/karthik/wesanderson).

Familiarization. The familiarization videos showed a small blue agent who approached goal objects (red balls). Initially, the agent was facing forward and located on a small, circular, light blue plate on the left side of the screen, approximately in the center along the vertical axis. In front of the agent was a grey wall of a slightly lower height than the agent, spanning all the way across the screen horizontally. The wall contained a bright green segment which represented a door. The goal object was always located in front of this door.

The videos began by the red ball expanding and contracting twice (2 s). The agent then turned toward the ball and hopped in the air (2 s). Following this, the green door flashed yellow once and sank into the floor (3 s). Once the door had completely disappeared, leaving behind a green stripe on the ground (as a reminder where the door had been) and an opening in the wall, the agent started moving and, passing through the gap in the wall, approached the ball (5 s). Upon reaching the ball, the agent once again hopped in the air (1 s). Key events (ball and agent movement, door flashing and sinking) were accompanied by sound effects.

There were two types of familiarization videos: short-path approaches and long-path approaches. In the former, the door and ball were located at a relatively shorter distance to the agent, at 3 units on the left-right axis. In the latter, the door and ball were further away, at 7 units. Both video types had equal duration, which meant that in the long-path approaches, the agent moved at a slightly higher speed. At the end of the video, the final frame of the event was left on screen for 1 s before the onset of the next trial.

Test. Infants observed two types of test trial videos: two-goals and no-goal videos. Both contained a layout which was similar to that of the familiarization videos, except that there were now two doors in the wall, at the same locations as they had been in familiarization (at 3 and 7 units, respectively). Additionally, in the two-goals video, there was a ball behind each of the doors, whereas in the no-goal video, there was no ball.

The two-goals test video began similarly to the familiarization videos, in that the two red balls contracted and expanded. They did so sequentially (2 s each), and each ball's movement was met with a hop from the agent, who turned toward the ball which had just moved (2 s each) and finally turned to face the midpoint between the two balls. The green doors then blinked yellow, each once individually and subsequently both at the same time, and then simultaneously descended (5 s). At this point, the agent started moving, but only until he reached the edge of the blue plate on which he was located at the onset of the video (2 s).

The no-goal test video was identical, except for the fact that there were no balls behind the doors. The agent first turned to face the midpoint between the two doors, which then blinked and sank. As a result, this test video had a shorter duration than the two-objects video (9 s compared to 15 s).

At the end of a test video, the final frame of the event was left on screen for 10 s, which constituted the test measurement period.

Stimuli can be accessed at https://osf.io/fvwga/.

We counterbalanced the order of familiarization trials (short-path approach first or long-path approach first), order of test trials (two-goals first or no-goal first), and the order in which the two balls in the two-objects test trial moved (closer ball first or further ball first).

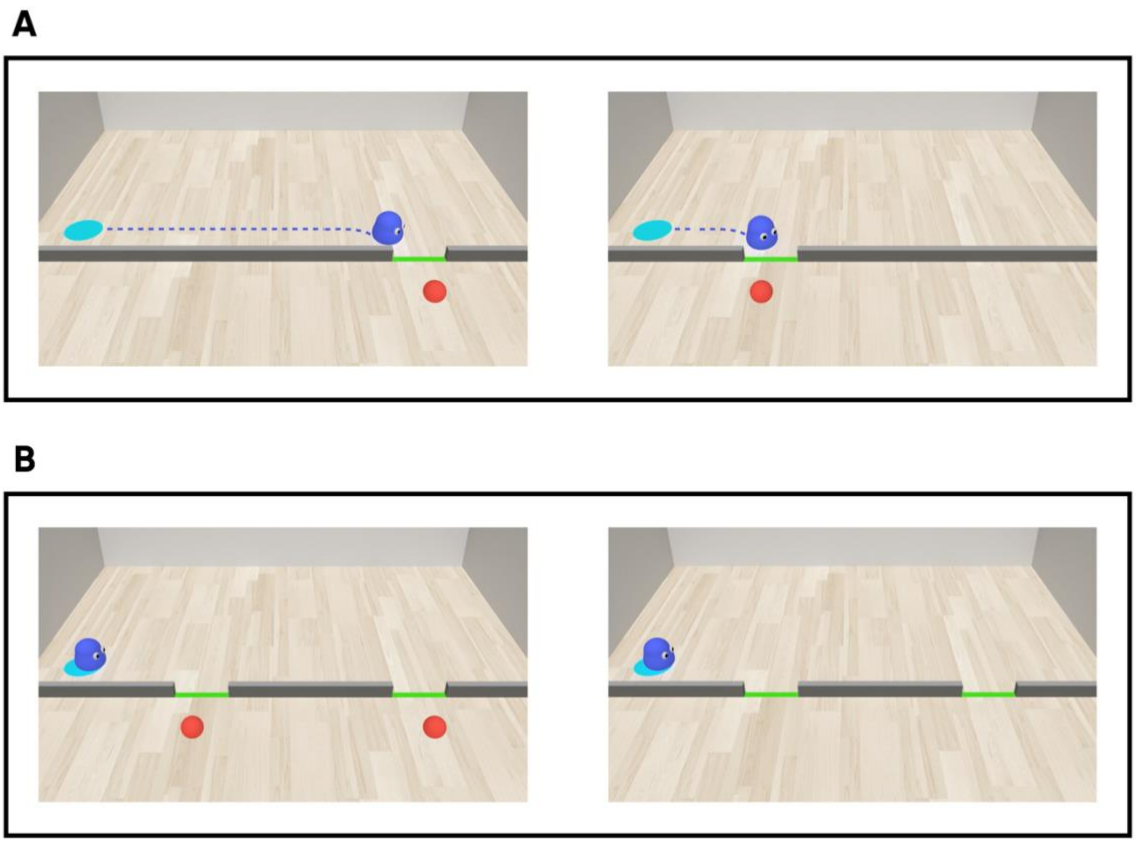

Figure 1.6. Stimuli used in Experiment 1.3. During Familiarization (A), the agent approached a goal object which was behind a door; the door was sometimes far away from the agent ("long path", left), sometimes closer ("short path", right). At Test (B), both doors were in the scene and there was a goal object behind each (two-goals trial, left), or there were no goal objects present (no-goal trial, right).

Data processing and analysis

Of primary interest for our hypothesis was the gaze data collected during the test measurement period (i.e., a 10-second freeze frame at the end of the trial, after the agent stopped moving but before he had approached a goal). The eye-tracker recorded sampled binocular gaze data every 16.67 ms. Gaze coordinates (x and y) of each sample were averaged across the eyes. Samples for which the x or y coordinate was missing, respectively, were removed. If there was data from only one eye available, these coordinates were used.

We defined areas of interest on the screen, specifically, the agent (A), the closer ball/door (C), and the further ball/door (F). The main dependent variable was the proportion of looking at the closer ball or door (i.e., the total looking to the closer ball/door divided by the sum of total looking to both balls/doors: propC = totalC / (totalC + totalF)).

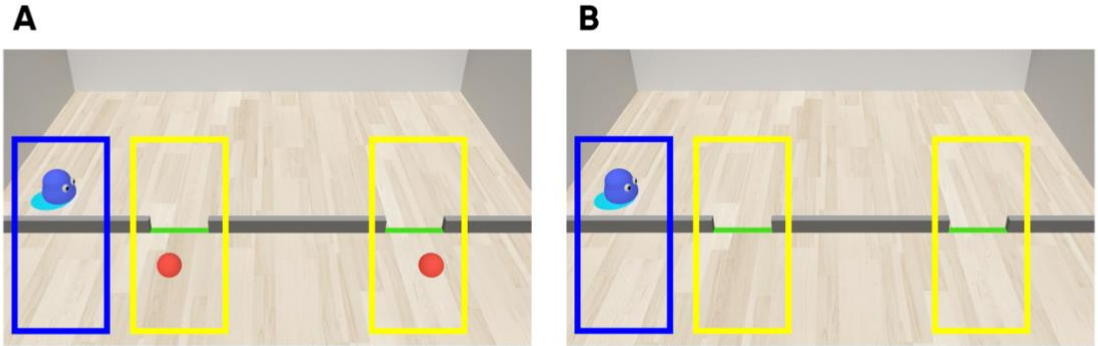

Figure 1.7. ROIs in Experiment 1.3, represented by yellow rectangles (for the two goal objects/doors, C and F) and blue rectangles (for the agent, A).

We performed two types of analysis on this variable. First, we directly compared the looking patterns in the two test trial types: To test whether infants on average looked longer at the closer option in the two-objects compared to the no-object trials, we used a one-sided Bayesian paired-samples t-test, predicting that the mean propC would be higher in the former than the latter. Second, we compared the mean propC in each condition to chance: To test whether infants would expect the agent to approach the closer rather than the further option in both conditions (or, as we hypothesized, only in the two-goals trials), we used one-sided Bayesian one-sample t-tests. Since there is little prior knowledge about the current effects, we used a default prior option, a Cauchy distribution (scale = 0.707). Statistical analyses were performed in R (version 4.3.2, R Core Team, 2023) with the BayesFactor package (Morey & Rouder, 2018).

To be included in the final sample, participants had to provide a valid data set, which minimally consisted of four valid familiarization trials and two valid test trials, one test trial per condition (two-goals and no-goal). A valid familiarization trial was defined as a trial during which the participant attended to the goal-approach action at least until the agent had crossed the line on the ground demarcating the opening in the wall (between

8 and 12 s of the animation[5]). A valid test trial was defined as one in which the participant (1) contributed at least 50% of on-screen data during the initial, animated phase of the video, (2) contributed at least four cumulative seconds of on-screen data during the test measurement period, and (3) gazed for a minimum of 300 ms (i.e., congruent with fixation) to at least one of the balls/doors during the test measurement phase.

Further exclusion criteria were caregiver interference (e.g., talking to the infant, pointing at the screen), experimenter error, external distractors (e.g., noise), and technical failure.

Results

Hypothesis-driven results

The mean proportion of looking to the closer option in the two-goals condition was $M_{\text{propC\_two-goals}} = .606$ ($SD_{\text{propC\_two-goals}} = .204$), and in the no-goal condition $M_{\text{propC\_no-goal}} = .535$ ($SD_{\text{propC\_no-goal}} = .265$). A direct comparison of the proportion of infants' looking to the closer option (propC) in the two-goals and the no-goal conditions showed no evidence that the looking behavior in the two events differed (BF: 0.82). However, when comparing propC to chance for each of the two trial types, we found that as predicted, infants looked longer at the closer option in the two-goals condition (BF: 73.61), but did not do so in the no-goal condition (BF: 0.38). Thus, one of the two preregistered analyses supported our hypothesis.

Additional results

During the test measurement period, infants attended to the screen for an average of 6.81 s (SD = 2.51 s) in the two-goals condition, and 6.68 s (SD = 2.25 s) in the no-goal condition.

We assessed whether the proportion of looking at the agent differed in the two test trial types, and found that infants looked more at the agent in no-goal compared to two-goals trials (BF: 1231.6; two-sided test), which is not surprising given that the former contained fewer elements to look at.

We further looked at the proportion of looking to the closer option during the movie phase of the test trials. Here, infants looked more at the closer option in both the two-goals (BF: 9152.8) and the no-goal condition (BF: 2480), and there was no evidence that this pattern differed between the conditions (BF: 0.3; all tests two-sided).

[5] We had preregistered this criterion without specifying whether infants had to attend for this entire time window, or whether brief look-aways were permitted. We later decided that the former would be too strict, forcing us to discard otherwise high-quality data, and applied a criterion according to which infants had to provide 90% of valid on-screen data during this event.

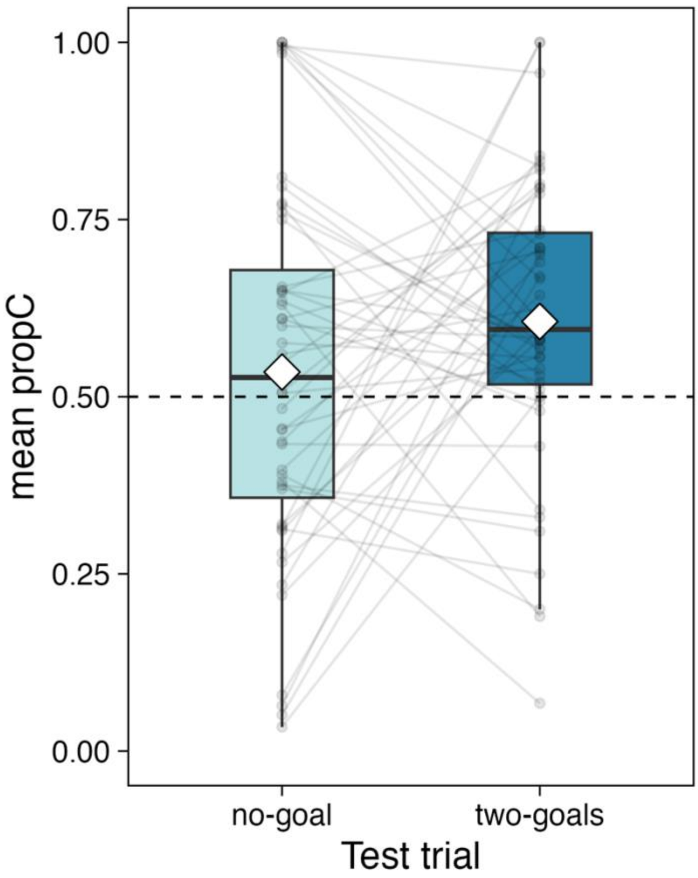

Figure 1.8. Box plot of mean proportions of looking to the closer option (propC) in no-goal and two-goals test trials. Light grey lines connect average looking proportions of individual participants, diamonds indicate means, horizontal lines indicate medians, boxes indicate middle quartiles, and whiskers indicate points within 1.5 times the interquartile range from the upper and lower edges of middle quartiles. The dashed horizontal line indicates chance (i.e., equal looking to the closer and further option).

Discussion

The aim of Experiment 1.3 was to test whether 14- to 16-month-old toddlers would predict that an agent would choose the one of two available goal options that could be reached at a lower cost. To this end, we used an eye-tracking design where participants were familiarized to an agent approaching a goal object which was located behind a door, either at a relatively shorter or longer distance from the agent. At test, both doors were present in the scene, and, in the two-goals condition, behind each there was a ball. We predicted that if infants use NUC and expect agents to maximize their utility by choosing a goal that requires less effort to reach, they should direct more looks towards the closer ball. In contrast, when there are no goal objects available in the scene, infants should look equally to the closer and further door, as neither affords a goal-directed action, and thus there is no utility to be maximized.

We found support for this prediction in one of the two analyses we had preregistered: While there was no evidence that the looking patterns in the two conditions were different from each other when comparing them directly, infants looked to the closer option above chance when the goal objects were present, but did not do so when they were absent.

This result tentatively indicates that infants considered the closer ball a better candidate for a prospective goal object in a context where multiple options became available, and thus ascribed choice of a better option to the agent. This conclusion differs from the one we drew from Experiment 1.1 and 1.2, where we did not find evidence that 10-month-olds assumed an agent would choose the outcome that yields a higher utility.

It is an open question which factors contributed to these divergent results. One possibility is that the participants' age difference mattered: We had speculated that there may be a developmental change occurring around the first birthday which contributes to infants' cross-domain capacity to represent and reason about multiple different alternatives.

Another possibility is that infants now succeeded because we used a different method to probe their thinking. Specifically, instead of measuring looking times to consistent and inconsistent outcomes, we recorded whether infants direct more looks to one of various potential goals which constitutes a better option. The former requires counterfactual reasoning, to compare the actually performed behavior with the non-chosen alternative, which requires both simulating an outcome that could have happened and representing the outcome that actually did. The latter, on the other hand, only requires future-directed hypothetical reasoning about different possible but unrealized outcomes, and thus no mental time travel to the past (Gerstenberg, 2022). Developmental research has shown that counterfactual reasoning may be more difficult for children than hypothetical reasoning (Robinson & Beck, 2000; Kominsky et al., 2021; Perner et al., 2004; Riggs et al., 1998; though even preschool-aged children seem to struggle with representing multiple mutually exclusive possible future states of the world; see Beck et al., 2006; Redshaw & Suddendorf, 2016).

While the present results provide initial evidence for utility-based reasoning about the hypothetical cost of different possible actions, it should be noted that the control condition only ruled out the possibility that infants' gaze was primarily guided by any perceptually salient item closer to the agent (such as the close door). There is, however, a second alternative that our control condition cannot account for: Infants may have measured the physical distance between the agent and the available goal objects and used this as a heuristic to select a preferred option, rather than simulating and comparing the actions that would be required to bring about each goal. This account predicts that infants should look more to the closer of multiple goal objects, but produce a similar proportion of looks to the two doors when the balls are absent, because the distance of non-goal items would not be relevant. Our results are compatible with both accounts, so future research will have to disambiguate between them, for instance, by conducting a follow-up experiment where the object that is closest to the agent as the crow flies is not the one that is least costly to reach.

Regardless, our study provides tentative support for the proposal that in the second year of life, children indeed apply NUC to predict which potential goal an agent will choose. The study is the first to use eye-tracking to demonstrate that infants expect agents to minimize their action cost (cf. Paulus et al., 2011), and points to promising directions for further research with this method.

General discussion

It is not known whether in making sense of agents' goal-directed behaviors by using NUC, infants ascribe a concept of choice to others. If so, they would explicitly represent the (relevant) alternatives available to an agent, calculate the utility of each option, and assume that the agent will behave rationally by selecting the one that allows him to maximize utility. Such a mechanism would serve to both explain past behavior and predict future actions as caused by utility-maximizing choice.

Here, we addressed whether infants possess such a concept by investigating whether they would expect an agent to choose the best available option. We presented participants with an agent who efficiently pursued a particular goal (to reach a target object). Then, in a novel context, multiple options became available which were consistent with the original goal, but differed in how much payoff they would yield (i.e., one vs. three target objects, objects that can be reached with lower vs. higher effort). In a series of looking-time experiments with 10-month-olds (Experiments 1.1 and 1.2), we found no evidence that infants found it more consistent for the agent to approach the higher-utility goal. In contrast, an eye-tracking study with 14- to 16-month-olds (Experiment 1.3) showed that infants looked more at the higher-utility goal, and this behavior did not seem to be driven by low-level visual saliency.

There are different possible explanations for these divergent results. On the one hand, it may be that the methods we used (violation-of-expectation experiment with a looking time measure versus eye-tracking with gaze recording during a freeze-frame before the goal-directed action occurred) prompted different cognitive processes in infants. On the other hand, we hypothesized that there may be a developmental change occurring around 1 year of age, such that infants acquire a (possibly domain-general) ability to reason about alternatives, which allows them to consider and compare the goal options an agent may pursue. Future research will have to disambiguate which of these explanations (if any) is on the right track.

The results we obtained speak only indirectly to the question whether infants in earlier research on teleological reasoning succeeded at these tasks by using a concept of choice. One reason for this is that these studies generally only featured one plausible goal in the scene, whereas in our experiments, there were multiple. Infants in the former context thus did not have to calculate and compare the expected utilities of mutually exclusive outcomes. Instead, if they compared alternatives at all, they only had to do so for different potential means that serve to bring about a fixed outcome (e.g., jumping versus moving in a straight line towards an object).

It is possible that these two tasks—comparing outcomes or means—recruit different cognitive mechanisms or differ in complexity. Investigating this hypothesis, Verschoor and

Biro (2012) found that when infants are given information on both the agent's "means selection" (i.e., efficiently adapting the actions to the environment) and "outcome selection" (i.e., approaching one of two different available objects), infants prioritize the former to generate a goal attribution. This study, however, does not explicitly address whether infants actually evaluate the efficiency of actions by representing and comparing different means, or instead use a simpler strategy. In Appendix A, we report an incomplete experiment with which we had intended to test whether 10-month-old infants would expect an agent to maximize utility both when (1) two goal options are available that can be reached with different amounts of effort (outcome comparison), and (2) there is only one goal available which can be reached on a long or a short path by detouring around opposite sides of an obstacle (means comparison). We stopped data collection after 14 participants, as results in the one-goal condition appeared inconclusive (BF: 2.34; infants had a tendency to look longer at the consistent event where the agent chose the *shorter* path).

The research we conducted overall shows that the exact mechanism by which infants interpret observed actions and generate predictions for future behavior remains unclear. Specifically, we have argued that it is an open question whether they possess a concept of choice and under what circumstances they deploy it. When and how do infants set up and compare alternatives? What options do they consider relevant, and how fine-grained are their utility comparisons (i.e., what makes one alternative meaningfully better than another)? What are babies' prior assumptions about goal rewards? From what age is the capacity to represent possible alternative goals available to infants, and what does their emergence depend on? Finally, which other cognitive operations rely on representing choice, or derive from a general ability to think about non-factual states of the world? Addressing these issues will be crucial to advance our understanding of the origins of both psychological and modal reasoning in infancy.

Section 2:

Infants’ and preschoolers’ understanding of third-party helping interactions

Introduction

Many behaviors are carried out not merely to generate a particular state of affairs for an agent herself, but with the aim of impacting someone else in a specific way, for instance by helping, sharing, cooperating, attacking, competing, and so on. Observers face the task of determining not only whether one agent's behavior affected another, but also whether such an effect was inadvertent or because the agent was pursuing a social, rather than a non-social goal (e.g., whether she opened the door for another person, or to let in some fresh air). These two alternatives yield different explanations for what motivated the agent, as well as corresponding predictions of how she will behave in the future. Young children, too, rely on drawing the right conclusions from the interactions occurring around them in order to make sense of the social world.

Based on findings from developmental studies, some researchers have argued that humans possess an innate moral core (Hamlin, 2013b): a set of sociocognitive abilities that allows infants to understand sociomorally relevant behaviors and respond to them by preferentially approaching or interacting with prosocial actors (Van de Vondervoort & Hamlin, 2016). One paradigmatic prosocial interaction that is employed in many of these studies is helping.

Appropriately interpreting helping interactions is not a trivial cognitive feat. Consider for instance a scenario often used in infancy research: One agent (the Helpee) tries but fails to open a box that contains a target object, and another agent (the Helper) comes and opens the box, allowing the first agent to grab the object inside (Figure 2.1). To make sense of this event as intended by the researcher, a naive observer not only has to infer the Helpee's unfulfilled goal (i.e., to reach the object inside the box), the necessary means to bring it about (i.e., to open the lid of the box), and relevant constraints and costs of actions (i.e., opening the lid is difficult or impossible for the Helpee), but also must work out how the Helper's behavior enables or facilitates the completion of the Helpee's goal. In doing so, the observer must discard alternative explanations of the Helper's behavior that do not directly relate to the Helpee's goal, for example, that the Helper had an object-directed goal (she may have simply preferred the box to be open for a different reason). As this example shows, naive onlookers often face the task of arbitrating among multiple available action interpretations. This is especially the case for second-order social goals such as helping, where an agent aims to generate rewards for another agent.

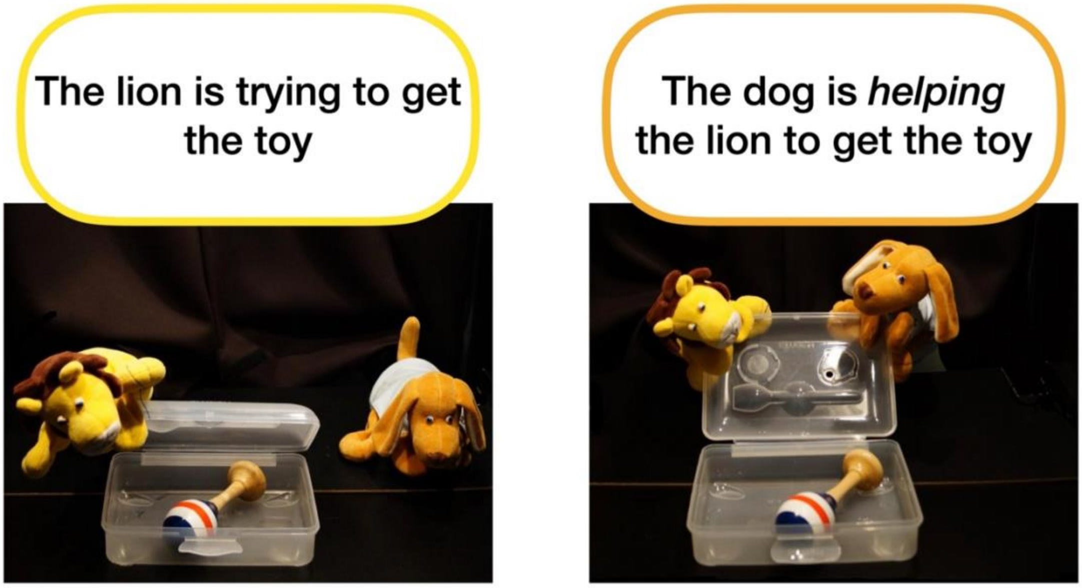

Figure 2.1. A helping scenario often used in infancy research. (*Left*) An agent (the lion puppet) is struggling to open a box, which is interpreted by the viewer as an instrumental action performed to reach the toy inside. (*Right*) Another agent (the dog) joins in the effort and gets the box open. Her intervention is interpreted as a helping action directed at the lion's instrumental action. Images from Study 1 in Salvadori et al. (2015).

Despite numerous findings on the purported sophistication of early reasoning about helping events, a comprehensive theoretical account of how infants and young children understand these events is currently lacking. Researchers often rely on their own (and the readers') intuitions when operationalizing helping events in their stimuli. Although this may be an inevitable or even justified approach, the growing developmental literature on this question does not detail the content and structure of helping representations with sufficient precision. The focus of this research has been to examine which inferences infants draw from observed helping interactions (e.g., inferences about the agents' intentions or dispositions) while neglecting questions concerning the minimal input criteria an event must satisfy to be represented as helping and the cognitive operations that generate its representation.

Recent proposals have argued that people, including infants, instead of possessing a number of domain-specific capacities for interpreting various social behaviors, rely on a naive utility calculus to reason about social actions, similarly as for non-social actions (Powell, 2022). To do so, they would have to integrate the utility functions of both agent and recipient, and categorize actions based on who benefits and who pays a cost. In such a framework, helping would be characterized as an action directed at increasing the utility of another goal-directed agent by intervening on that agent's action constraints.

In the following, we first review the existing literature on infants' understanding of helping. Then, we describe in more detail the proposal that a concept of helping consists in that of a hierarchical, utility-based second-order goal (henceforth: H-NUC), and critically discuss whether this is the concept that infants adopt when representing observed instances of helping. Finally, we outline alternative candidates for an early helping concept.

The subsequent chapters then present empirical studies probing infants' and children's understanding of helping. Chapter 2.1 contains a replication of a seminal study on the topic: Hamlin's (2007) finding that when infants were presented with displays of characters who help or hinder another agent, they subsequently tended to reach for the Helper. We conducted a replication of this paradigm as it is the one most widely used to probe infants' intuitions about third-party helping actions. Chapter 2.2 and 2.3 describe studies that were meant to test the hypothesis that infants and young children possess H-NUC. In Chapter 2.2, we report a series of experiments with infants, where we investigated whether they ascribe the goal of helping to an agent who lowers another agent's action cost. Chapter 2.3 introduces an experiment conducted with preschoolers in which we asked whether they would themselves help another agent in a way that maximizes that agent's utility, and whether they would identify an agent who acted in this way in a third-party context as the one who helped. Note that in the remainder of this introduction, we focus on reviewing research with infants, as the main focus of this work was to analyze a potential early-emerging helping concept; pertinent studies with older children are discussed in Chapter 2.3.

Insights into infants' helping concept have an important bearing on theories of early action understanding. Spelling out the mechanism by which young children comprehend certain events as helping can contribute to evaluating the proposal that NUC may guide the representation of both nonsocial instrumental actions and social interactions (Powell, 2022), and to specifying which processes drive learning in these two domains. Moreover, insights into how infants and children represent third-party helping interactions can inform the research program investigating their own helping behavior. From the second year of life, infants provide help to others, for example, by handing over an out-of-reach object (Warneken & Tomasello, 2007). It is not clear, however, how infants in such contexts conceive of their actions or precisely what goal they pursue (Hobbs & Spelke, 2015). There is an ongoing debate as to whether helping in young children reflects a genuine altruistic concern for others and an attempt to address the Helpee's specific unmet needs and desires (Hepach & Warneken, 2018; Warneken, 2018), a mere preference for goal completion (Michael & Székely, 2019), or a general motivation to interact with people (Allen et al., 2018; Cortes Barragan & Dweck, 2014; Dahl & Paulus, 2019). Knowing whether representations of third-party helping interactions are undergirded by a utility-based concept may provide a relevant piece of evidence to this debate by suggesting that infants factor the utility of the Helpee into their own helping behavior (for related evidence that infants take into account their own costs when deciding whether to help, see Sommerville et al., 2018).

While helping can encompass many means of assistance, such as providing information, support, or material benefits (for example, we talk of helping lost tourists by giving them directions or helping flood victims by donating money), these means fall outside the scope of instrumental helping that we discuss here. This is the case also when helping is achieved by handing a recipient an out-of-reach object—a scenario that has been employed in several studies on early sociomoral evaluation (e.g., Hamlin et al., 2011; Hamlin & Wynn, 2011; Köster et al., 2016). Considering that infants default to interpreting the active transfer of an object as a prosocial action, irrespective of whether the object was sought by the recipient (and hence constituted her goal or not) (Geraci & Surian, 2011; Tatone et al., 2019), in such scenarios infants may be leveraging assumptions about giving actions in general rather than about instrumental helping per se. This possible source of ambiguity motivates our decision to focus on instances of helping in which the Helper intervenes on the constraints that the Helpee faces in reaching her goal.

Contrary to previous work on helping, which framed the question of how infants engage in (Dahl, 2020) or evaluate (Van de Vondervoort & Hamlin, 2016; Woo et al., 2022) such a prosocial behavior in the context of moral development, the present analysis deliberately sidesteps this issue. While spelling out the representational constituents of helping that infants leverage may be relevant to adjudicate whether, for instance, there is ontogenetic continuity in the construals that people recruit to apprehend such behaviors, this analysis is orthogonal to the question of how sociomoral intuitions develop. Recognizing an action's goal and evaluating it as good or bad are, in fact, functionally distinct operations and as such are likely subserved by different mechanisms (Mikhail, 2011). Insofar as our primary aim is not to identify the types of sociomoral inferences that infants draw from observing helping, but rather the conceptual structures they may use to understand this behavior, here we sidestep the question of moral development.

Infants' responses to the observation of helping

Beginning with a study by Premack & Premack (1997), a wealth of experiments has probed preverbal infants' understanding of third-party instrumental helping interactions. These experiments generally implement the act of helping as follows. One agent, a Helpee, tries and fails to perform an instrumental action, which is directed at either gaining access to an out-of-reach target or arriving at a particular location. Another agent, a Helper, intervenes either on the environment or directly on the Helpee such that as a result of the intervention, the Helpee reaches her goal. The majority of studies used either the hill scenario (Kuhlmeier et al., 2003), where one agent tries to reach the top of a hill and is pushed there by another, or the box scenario (Hamlin & Wynn, 2011), where one agent tries to access an object inside a box and another agent opens the box so that the first agent can retrieve the object. Other studies presented infants with the Helper lifting or pushing the Helpee toward a goal object or location (Holvoet et al., 2019; Premack & Premack, 1997), acting on other kinds of physical constraints impeding the Helpee's action toward her goal (Hamlin et al., 2013; Woo et al., 2017), or transferring an out-of-reach goal object directly to the Helpee (Hamlin, 2014; Hamlin et al., 2011; Hamlin & Wynn, 2011; Jin & Baillargeon, 2017; Köster et al., 2016, 2019; Singh, 2020; Steckler et al., 2018).

These studies measured either (*a*) the infants' own preference for the Helper, operationalized as the propensity to manually choose this character over another (typically, a Hinderer), or (*b*) the infants' expectations about subsequent interactions among the participating agents, probed via common gaze measures (looking times) and neurophysiological measures (electroencephalogram).

Manual choice has traditionally been interpreted as a measure of social preference under the assumption that infants would preferentially explore agents they seek to affiliate with. Thus, infants' selective preference for Helpers over Hinderers (Hamlin, 2015; Hamlin et al., 2007) as well as neutral agents (Chae & Song, 2018; Hamlin et al., 2007) has been taken as evidence that infants consider helping a prosocial action. Furthermore, this early social evaluation appears to be surprisingly nuanced (Hamlin, 2013a, 2014; Hamlin et al., 2011, 2013; Steckler et al., 2017; Woo et al., 2017; Woo & Spelke, 2020a). Note that replication attempts using this measure have yielded mixed results (Chae & Song, 2018; Colaizzi, 2016; Cowell & Decety, 2015; Fortin, 2019; Maxwell & Rafetseder, 2015; Nighbor et al., 2017; Salvadori et al., 2015; Y. Shimizu et al., 2018; Vaporova & Zmyj, 2020). Chapter 2.1 describes a replication of the original study by Hamlin and colleagues (2007), in which we found that 15-month-olds did not show a preference for Helpers. We discuss the context and potential reasons for this replication failure in that chapter.

Beyond preferring Helpers over Hinderers, infants expect the recipients of helping actions to exhibit a similar preference (Fawcett & Liszkowski, 2012; Kuhlmeier et al., 2003), even if the Helper merely attempts to help without succeeding (Y. Lee et al., 2015). Infants have also been shown to infer social groups from selective helping patterns. For instance, infants consider agents who helped each other to belong to the same group and to be united against an out-group (Rhodes et al., 2015). Conversely, social structure affects infants' expectations about the occurrence of helping; specifically, infants expect helping to occur when directed toward an in-group, rather than an out-group, member (Jin & Baillargeon, 2017; Pun et al., 2021). In the absence of group-relevant information, infants and toddlers default on expecting agents to help those in need rather than to withhold help (Hepach et al., 2012, 2016; W. Lee et al., 2020), as well as to preferentially help agents in greater need (Köster et al., 2016, 2019; Schuhmacher et al., 2019).

As this brief literature review attests, preverbal infants seem readily able to apprehend third-party helping actions and to generate expectations and preferences on that basis. These findings suggest a sophisticated understanding of helping as a prosocial behavior. Infants seem to understand in which situations helping is more likely to occur, who may be an appropriate recipient, and what type of intervention would satisfy the Helpee's goal; further, they seem to draw rich social inferences from the observation of helping. However, while this work sheds light on the broad range of factors that influence infants' preferences and expectations, the fundamental question of how infants represent social interactions that adults commonly interpret as helping has been largely overlooked.

The mature helping concept: Increasing the utility of the Helpee's action

What kind of helping concept are these findings evidence of? To address this question, we first sketch what a mature folk concept of helping might consist in. Adults' understanding of helping is plausibly more sophisticated and nuanced than that of infants. Adults have more knowledge about the types of goals that agents may pursue, the different rewards that action outcomes yield, the different types of costs that their completion might entail, and the various means by which these goals can be brought about. Adults also come to learn that helping may be carried out preemptively, which entails predicting what goal the Helpee may require assistance with (e.g., handing a tool to someone who does not yet know she will soon need it), or paternalistically, which requires appreciating that actions that do not align with the Helpee's short-term goals may nevertheless promote the Helpee's welfare (e.g., a mother putting a scarf and hat on their resistant child who is impatient to play outside in the snow).

That said, even for the more basic instances of instrumental helping such as the ones discussed earlier, it is not immediately clear how adults solve the task of ascribing the goal of helping. One promising approach is to ground the understanding of helping within the general framework of the NUC (Jara-Ettinger et al., 2016; Powell, 2022), which was discussed in Section 1.

Social actions can be captured in a utility-based framework by modeling the interdependency of the agents' utility functions (Baker et al., 2008; Kleiman-Weiner et al., 2016; Shum et al., 2019; R. E. Wang et al., 2020). While an agent performing a nonsocial object-directed action can maximize her own utility directly by bringing about the intended outcome (e.g., by acquiring a valuable good), a Helper does so indirectly by maximizing the Helpee's expected utility (e.g., by allowing the Helpee to retrieve her own valuable goods). In other words, the Helpee's expected utility function is embedded in that of the Helper. Under this account, an observer can attribute the goal of helping to an agent by recognizing that this agent is acting efficiently with respect to the goal of increasing the utility of another agent's own action. Thus, the interpretation of an action as helping is a second-order goal attribution to the Helper, whose proximate goal is dependent on the goal attributed to the Helpee.

The assumption that helping generates rewards for the Helper follows directly from the normative standard of rational action theory, according to which the hallmark of goal-directed behavior is the increase of the agent's utility. While it may seem paradoxical to hold such an assumption for prosocial behaviors, as these result in the voluntary imposition of net cost for the agent, it is worth noting that rewards need not be either immediate or direct. For instance, the Helper's action may positively affect the well-being of valuable kin, engender subsequent reciprocation from the assisted partner, or produce reputational gains in the eyes of third parties (well-known examples of indirect rewards). Which types of rewards (if any) infants may infer from the occurrence of helping behavior is orthogonal to the validity of this assumption and is not the subject of our inquiry.

The Helper can increase the utility of the Helpee's action in two ways: either by reducing the action cost that the Helpee has to pay to bring about her goal or by increasing the reward that she obtains because of the Helper's intervention. Using the rationality assumption, the observer can infer that the increase in the Helpee's utility, however small or large, offsets the Helper's costs of intervening on the Helpee's behalf. People may hold prior assumptions about what is a reasonable trade-off between Helper's and Helpee's utility (Jara-Ettinger et al., 2020), but they may also treat skewed cost-reward distributions as indicative of the Helper's evaluation of her social target[6]. For example, when witnessing a Helper voluntarily incurring large costs to bring about a negligible utility increase for the Helpee, one may infer that the Helper greatly values the Helpee. Importantly, however, not all agents' utility functions are the same (Pomiechowska & Csibra, 2020, 2022): What is costly for one agent may not be so for another.

The intuition that people possess a H-NUC concept was formalized through Bayesian modeling and tested by Ullman and colleagues (2009). They found that their inverse planning model matched participants' judgments about an animated agent's likely goal (nonsocial object approach versus help/hinder) better than a cue-based model, which disambiguates goals by simple cues such as physical proximity (cf. Netanyahu et al., 2021; Shu et al., 2020). This study suggests that actions do not need to exhibit particular visuospatial cues for adults to identify them as instances of helping; instead, what matters is that the Helper's behavior is apprehended as promoting the utility of the Helpee in the

[6] See work on Welfare-Tradeoff Ratios (e.g., Delton & Robertson, 2016; Tooby & Cosmides, 2008; Howard et al., 2018).

most efficient way possible, given the constraints that both agents operate under. Corroborating this suggestion, He et al. (2020) showed that adults were most likely to interpret an ambiguous transitive action as helping (*a*) when it was inefficient with regard to the nonsocial goal of approaching a target object and (*b*) when the Helpee's costs of goal fulfillment were reduced as a result.

These studies suggest that people may infer the goal of helping by recruiting the same mechanisms that guide the interpretation of nonsocial instrumental actions. They also suggest that, while possibly facilitating the identification of helping, behavioral cues that may commonly accompany these actions in the real world (e.g., behavioral synchrony, contingency, communication) are neither necessary nor sufficient for adult observers to identify an interaction as helping. A second, perhaps more important, reason to doubt the usefulness of perceptual cue-based approaches is that, unlike other social goals such as giving (Tatone et al., 2015; Yin et al., 2022) and chasing (Gao et al., 2009), which can be exhaustively defined with respect to particular event features (giving meaning that one agent transfers the possession of an object to another agent and chasing meaning that one agent moves in the direction of another in a heat-seeking manner), helping can occur in a multitude of ways and can be directed toward any conceivable first-order goal. Furthermore, as different goals can generate similar-looking behaviors, reliance on perceptual cues alone is of limited use in allowing observers to disambiguate helping from other types of social scenarios, for example, agents pursuing individual goals side-by-side or jointly working toward a shared goal. As these considerations suggest, cue bootstrapping is unlikely to play a constitutive role in guiding the representation of helping for two reasons: On one side, because paradigmatic helping cases are not restricted to particular means or outcomes, and on the other, because the social cues that tend to accompany helping events also occur across a suite of social interactions modeled on different payoff distributions.

This is not to imply that people do not exploit information conveyed by interaction cues in naturalistic settings. These cues play an important role in reducing ambiguity, as they can reflect processes that play a functional role in social interactions; for instance, communication or eye contact can be used to signal common knowledge or commitment (Siposova et al., 2018; Wyman et al., 2013), and spatiotemporal coordination can help make behaviors more predictable to one's social partner (Vesper & Sevdalis, 2020). Moreover, especially in ambiguous situations, where an action results in an outcome that may be helpful for another but that may also constitute an end in itself for the agent (e.g., removing the cap on a bottle of soda held by another person), people may leverage contextual cues (e.g., the person holding the bottle shows gratitude) or prior knowledge (e.g., bottle caps are not valuable goods) to constrain interpretation.

Do infants understand helping as a second-order goal with an embedded utility function?

The research on adults' interpretation of helping events (and other collaborative interactions; see e.g. Kleiman-Weiner et al., 2016; Shum et al., 2019) discussed above suggests that the folk concept of helping is captured by H-NUC. Such a concept is flexible enough to allow observers to identify even unfamiliar helping actions. Considering the rich literature suggesting that infants recruit the naive utility calculus to infer instrumental goals (Gergely & Csibra, 2003; S. Liu et al., 2017), it seems natural to ask whether infants interpret helping by leveraging an adult-like concept.

Here, we articulate three proposals for what an early helping concept might consist in (though this is not intended as an exhaustive list): (*a*) a mature concept of helping, where infants interpret helping the same way as adults do (H-NUC); (*b*) helping as enabling, where infants interpret helping as an action that makes it possible for the Helpee to fulfill her goal; and (*c*) helping as joint action, where infants interpret helping as an interaction where two or more agents work together toward a shared goal (for an illustration, see Figure 2.2; for a summary of features, see Table 2.1). In the sections below, we discuss these accounts in more detail, and spell out their respective requirements in terms of the cognitive operations to be carried out.

Helping concept	Cognitive requirements			
	Utility calculus	**Second-order goal ascription**	**Counterfactual outcome**	**Representing coordination**
Mature	Yes	Yes	No	No
Enabling	No	Yes	Yes	No
Joint action	Yes	No	No	Yes

Table 2.1. Summary of the different cognitive operations entailed by the proposed helping concepts. Utility calculus refers to evaluating an agent's action in terms of a graded magnitude increase (or decrease) to an individual (or aggregate) utility function (compared with enabling, where the utility change that the Helper brings about is nongraded and qualitative). Second-order goal ascription refers to setting up hierarchical representations of distinct agent-specific goals (compared with joint action, where the agents' actions can be construed by a common reference to a shared goal). Counterfactual outcome refers to comparing two different outcomes (the realized state of affairs in which the Helper's intervention leads to the Helpee's goal achievement and an unrealized one involving the Helpee's failure to achieve her goal on her own, which is required to contrastively assess the Helper's contribution). Although counterfactual reasoning is involved in all three helping accounts, only in the enabling case do the terms of the counterfactual comparison involve two outcomes rather than the same outcome achieved at different (individual or aggregate) costs. Lastly, representing coordination refers to interpreting agents' actions as directed toward a single shared goal.

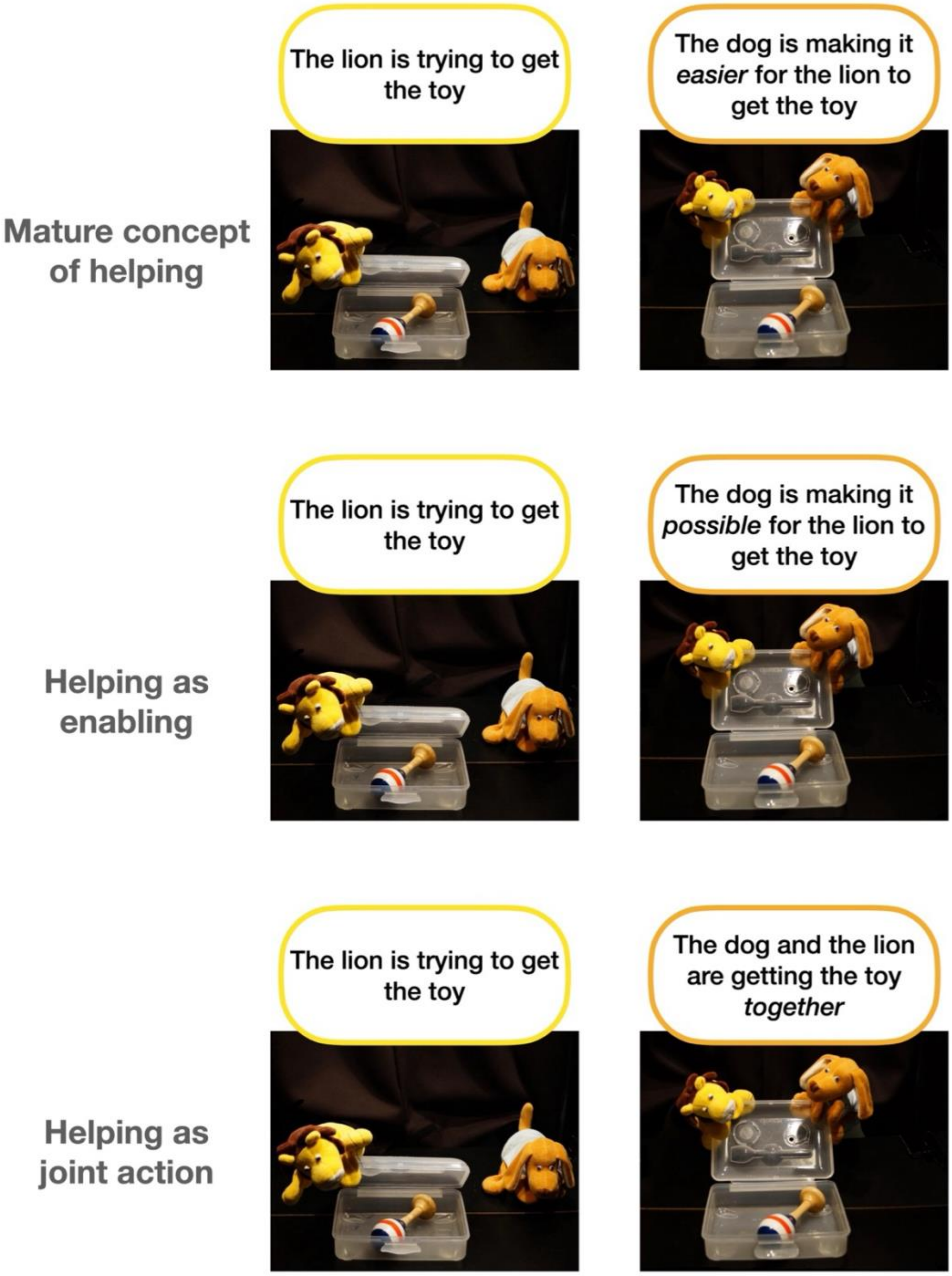

Figure 2.2. Schematic illustration of the three accounts of helping using the scenario depicted in Figure 2.1. The captions specify the goal description that the infant is expected to attribute to the Helper and Helpee in each of the three accounts. In a mature concept of helping (*top*), the Helper's goal is to reduce the effort of the Helpee in her goal pursuit. In helping as enabling (*middle*), the Helper's goal is to allow the Helpee to reach a previously unattainable outcome. In helping as joint action (*bottom*), the Helper has a shared goal with the Helpee of bringing about a particular outcome together. Images from Study 1 in Salvadori et al. (2015).

Before proceeding further, however, it should be noted that alternative proposals attempting to explain infants' responses to helping events without leveraging a proper helping concept also exist. Most notably, Spelke and Powell (Powell & Spelke, 2018a; Spelke, 2016, 2022) have argued that infants first conceive of helping interactions as instances of imitation, insofar as Helpers often reproduce some of the same actions that the Helpee directed toward her first-order goal. Under such an account, infants' preference for Helpers can be reframed as a preference for imitators, who display affiliation by aligning their behavior to that of the imitated agent (Powell & Spelke, 2018b; for similar arguments, see also Benton & Lapan, 2022; Premack & Premack, 1997). Crucially, however, such an interpretation of infants' preference is at odds with studies reporting a bias toward Helpers even when (*a*) Helpers perform actions that are highly dissimilar from the Helpee's (Hamlin et al., 2013; Woo et al., 2017) and (*b*) Helpers and non-Helpers differ in their knowledge about the Helpee's goal, but not in their actions (Woo & Spelke, 2023). More importantly, this account does not explain how H-NUC could emerge by leveraging such an imitation-based concept. In fact, Powell has since changed her views, arguing that a utility-based concept of helping may be already present in infancy (Powell, 2022). Considering the fundamental dissimilarity between a representation of helping as imitation and its mature counterpart (with which it does not share crucial cognitive prerequisites; see Table 2.1), as well as the lack of an identifiable developmental pathway for conceptual change, we do not discuss this proposal further.

Infants possess the mature concept of helping

It is possible that infants recruit the concept we argue that adults possess, H-NUC. If they do, a number of predictions follow: Infants should be able to distinguish between actions that are more or less helpful on the basis of their relative cost mitigation; and they should expect a Helper to intervene when doing so would reduce the Helpee's cost, but refrain from acting when her action would have no such effect. In the empirical chapters of this section, we tested these predictions.

While it is possible that infants have H-NUC from early on, this has not been conclusively demonstrated (though see Woo et al., 2022, for a study with toddlers, discussed further in Chapter 2.3). Appropriately setting up such a representation presupposes several cognitive operations of non-negligible complexity. One of these operations is counterfactual reasoning. Infants would have to infer that the Helper's action increased the utility of the Helpee by comparing the actual costs the Helpee incurs after receiving help with the (higher) prospective costs she would have incurred had she been left unassisted. As discussed in Section 1, although evaluating the efficiency of instrumental actions may rest on counterfactual reasoning, simpler inferential mechanisms could also support this type of analysis. Such a strategy, however, could not scaffold the type of contrastive analysis (i.e., utility of being helped versus utility of acting unassisted) required to assess whether the Helper's action qualifies as helping—i.e., whether it increased the utility of the Helpee.

Another operation demanded by second-order goal attribution is utility embedding. To represent the goal of the Helper as directed at reducing the costs of the Helpee's goal fulfillment, infants need to nest the agents' individual utility functions into one another. Doing so presupposes selecting as the Helper's goal not the immediate state of affairs that this agent brings about (e.g., opening a box) but the distal effects that this outcome has on the Helpee's action options (e.g., allowing her to retrieve the object inside). Appropriately solving this selection problem is not trivial, especially considering that explanatory parsimony may encourage them to privilege structurally simpler action interpretations. For instance, upon being exposed to an agent taking an object from a patient, infants interpret the agent's action as directed toward acquiring a resource, without considering the action's consequences on the object's original possessor (Tatone et al., 2015; see also Yin et al., 2022). While adults likely solve this selection problem by applying their knowledge about prototypical goal states (i.e., outcomes that generate direct or derived rewards), this strategy may not be available to infants due to their limited goal repertoire (S. Liu et al., 2019) and their still developing understanding of means–ends relations (Woo & Spelke, 2020b). This challenge is further compounded for second-order social goals, such as helping, insofar as these actions cannot be construed by a common reference to a single agent's utility (Hobbs & Spelke, 2015). As such, these actions open the possibility that infants may construe (or misconstrue) the Helper's action as directed to a change of state that is immediately rewarding for the agent herself, irrespective of its effects on the Helpee's goal fulfillment.

Considering the suite of operations that the deployment of H-NUC entails, it is plausible to assume that cognitively leaner concepts may be adopted by infants to make sense of helping interactions. If this is the case, it remains to be explained what the representational content of these concepts is, how they allow infants to generate behavioral responses congruent with researchers' predictions, and, more importantly, how children transition from these early construals to H-NUC. In the following sections, we sketch two alternative concepts of helping, explain how they differ from the mature helping concept, and identify the most relevant developmental steps necessary for their transition into the mature state.

Infants conceive helping as enabling

The helping interactions used in the studies reviewed above often feature the Helpee trying and failing to bring about a goal (e.g., climbing a hill, opening a box, or reaching for an object) before any helping occurs. These events are meant to demonstrate that the Helpee has a goal that she is unable to attain on her own, implying that the Helper's intervention is necessary to goal completion. In contrast, in the real world people often help with tasks that could have, in principle, been carried out alone (at higher costs) such that the intervention makes it easier for the Helpee to reach her goal while not being strictly necessary. Indeed, the stimuli used to study adults' helping concept involved scenarios where the Helper intervened to facilitate the fulfillment of outcomes that the Helpee could have realized by herself (He et al., 2020; Ullman et al., 2009).

It may not be a coincidence that the scenarios that infants are exposed to generally feature failed attempts. Infants' success in the studies reviewed above raises the possibility that infants recruited a concept of helping as enabling: an action whose goal is to make goal fulfilment possible for the Helpee. Such a concept would still require infants to embed individual goals into one another, insofar as the Helper's goal achievement is to enable the completion of the outcome that the Helpee is trying to bring about (thus attributing a second-order goal to the Helper in the same sense as the mature concept does). However, this concept does not require a proper utility analysis (Table 2.1). While conceiving of helping as facilitating an agent's goal achievement requires comparing graded utility magnitudes (i.e., how much it costs for the Helpee to bring about an outcome with or without the Helper's assistance), conceiving of helping as enabling goal achievement only requires comparing two discrete states: one in which the Helpee's goal is not realized (failure) and one in which it is (success). Utility analysis is not completely eschewed by adopting such a concept; infants still need to evaluate the efficiency of the Helper's action with respect to whether it minimized her own potential costs incurred to bring about her goal (to help). Doing so, however, does not additionally require computing the utility increase that the Helper's action generates for the Helpee. Here, the Helper's goal is to enable the Helpee's success, nothing more. Thus, under this concept of helping falls any intervention that has the effect of bringing about a qualitative change in the Helpee's action options (i.e., making the outcome realizable).

Because of its simplicity, the application of this concept is fairly restrictive. Yet, even if helping as enabling is predicated on a simple transition between two discrete states (failure to success), its application would still require counterfactual analysis insofar as the Helper's contributions can only be contrastively assessed against the Helpee's failed action. This holds true even if (as in many studies) the Helpee's failing is explicitly shown before the Helper intervenes, because interpreting the outcome of the Helpee's action as a failure demands consideration of the unattained goal of the action (i.e., a counterfactual outcome). Although some research suggests that infants can infer goals from failed attempts (Behne et al., 2005; Brandone & Wellman, 2009; Hamlin et al., 2008, 2009; Meltzoff, 1995), it is not known under what circumstances infants interpret incomplete actions as failed attempts to bring about a goal state rather than as complete actions (e.g., an agent moving up and down a hill may be interpreted as playing rather than as struggling to climb to the top). The actions employed in these studies most often consist of behaviors that are likely familiar to the infants, the goal of which they know from experience (e.g., opening a box to retrieve an object). However, considering the limited repertoire of goal states that young infants can leverage to understand actions, it is unlikely that they would appropriately recognize failed actions whose goal state they have never seen fulfilled.

Nevertheless, if infants can infer the goal of a failed action, the observation of the Helpee's failed attempts could serve as a potential cue for construing the Helper's subsequent behavior in relation to the Helpee's goal. This construal is based on the understanding that the actions of the Helpee with and without the Helper's intervention were directed to the same goal. Comprehending the Helpee's failure and subsequent success after receiving help thus entails an appreciation that she would not have reached her goal without the Helper's assistance. In order to transition toward a mature concept of helping, infants endowed with the enabling concept would have to eventually abandon the assumption that helping can occur only in situations where this counterfactual holds. Instead, they would have to come to appreciate graded differences between utility states that could have been obtained by the Helpee with or without the Helper's assistance.

Infants conceive helping as joint action

It is also possible that infants' concept of helping is subsumed under the concept of joint action. Under such a construal, infants would teleologically link the Helpee's and Helper's actions without embedding their individual goals into one another. Joint actions can be defined as social interactions "whereby two or more individuals coordinate their actions in space and time to bring about a change in the environment" (Sebanz et al., 2006). In contrast to helping actions that are proximately altruistic, joint actions are studied under the premise that the agents engaged in them pursue a shared goal from which both benefit in some way. In addition, unlike in helping, the individual goals of the agents participating in a joint action are not derived from each other but from a shared goal. In this sense, they are not second-order goals but rather subgoals of the shared goal in the same sense as subgoals serve further goals in individual action hierarchies (see e.g. Csibra, 2008a).

A number of studies on infants' understanding of third-party joint actions suggest that they can attribute a shared goal to the collaborating agents involved such that its completion generates positive utility for all participants. This research suggests that infants take joint actions to be directed at bringing about a shared goal, from which both collaborators stand to benefit (Y. Wang & Henderson, 2018) and where each collaborator may perform the action steps required to bring about this goal (Fawcett & Gredebäck, 2013, 2015; Henderson et al., 2013; Henderson & Woodward, 2011). Moreover, infants seem to adjust their evaluative standards for rational behavior in a joint action context. Agents directing their efforts toward a shared goal should behave in a coefficient way that minimizes aggregate, not individual, action costs (Török et al., 2019). Preliminary research suggests that, when presented with two agents coordinating, infants do not hold expectations of individual efficiency, but of joint efficiency (Begus et al., 2020; Mascaro & Csibra, 2022). Importantly, much like in helping, the efficiency of joint action can warrant an asymmetric distribution of effort such that one agent reduces aggregate costs by reducing the partner's effort at the expense of her own and thus acts in a locally helpful way toward the other within the joint-action context (Török et al., 2021).

Some of the research on infants' joint action understanding employed events very similar to the box scenario used by Hamlin and colleagues to study helping (Henderson et al., 2013; Henderson & Woodward, 2011; Y. Wang & Henderson, 2018). Unlike in helping, however, representing the two agents as collaborating toward a shared goal requires only that the two actions be ordered in a causally adequate means–end structure (e.g., the container must first be opened and then the object retrieved), without having to nest one agent's goal into the other.

Helping actions can resemble joint actions when one agent (the Helper) adopts the Helpee's goal and treats it as if it were shared. However, unlike in joint actions, in helping scenarios it is only the Helpee who directly benefits from bringing about her goal, while the Helper solely stands to gain from the fact that the Helpee's own utility has increased. Observers with a mature helping concept can appreciate this difference and appropriately infer who benefits from a given state of affairs, even if jointly realized. To do so, they may use contextual knowledge or behavioral cues (e.g., one of the two agents expressed an inclination for, or attempted to bring about, a certain goal). In contrast, infants may be unable to differentiate between interactions geared toward mutualistic versus altruistic outcomes. Even in the presence of skewed cost investments, infants may nevertheless interpret the Helper and Helpee as working together toward a shared goal and correspondingly evaluate the efficiency of their contributions in terms of their aggregate utility. To interpret helping as a joint action, they would not need to entertain a second-order social goal involving nested utility functions but rather a first-order goal shared by the cooperating agents (Table 2.1).

Just as failed attempts may serve as cues to guide the interpretation of an agent's action as helping (of the enabling kind), several cues commonly used in infant studies may help infants identify helping scenarios as instances of joint actions. Examples of such cues include communicative interactions (e.g., gesturing or vocalizations), agents orienting toward each other (e.g., eye contact), proximity or physical contact, acting on the same object, and spatiotemporal contingency of the agents' behaviors (for evidence of such cues being used in the early representation of third-party interactions, see Augusti et al., 2010; Beier & Spelke, 2012; Fawcett & Gredebäck, 2013; Powell & Spelke, 2013; Tauzin & Gergely, 2018; Thiele et al., 2021). These cues help set up an interpretive prior that the observed scenario may involve coordination between the two agents, and thus link the two agents' behaviors. Conversely, this account also implies that when assistance occurs without conspicuous cues of interaction or collaboration between the parties, naive learners leveraging such a concept may fail to appropriately recognize the episode as an instance of helping.

To transition from a concept of helping as joint action to a mature concept, infants would have to understand that the presumed joint goal originates from the Helpee's goal, which the Helper adopts to increase the Helpee's utility. This developing understanding should be accompanied by a transition from interpreting the two agents as contributing toward a shared goal to embedding the goal of one agent (the Helpee) into that of another (the Helper). A full-fledged concept of helping would finally allow infants to identify altruistic goals even for interactions devoid of perceptual proxies of collaboration or in which helping occurs distally.

The present research

A version of the H-NUC account seems to underlie many infant researchers' implicitly recruited helping concept. However, it is currently an open question whether this is actually how infants understand the goal of helping. As we pointed out, representing helping as H-NUC entails several cognitive operations that may be challenging for young infants. These include counterfactual reasoning (to assess the Helpee's utility with and without the Helper's intervention; cf. Section 1), embedding the individual goal of one agent (the Helpee) into that of another (the Helper), and appropriately relating the Helper's action to the distal effects on the Helpee's behavior and selecting these, rather than proximal state changes, as the Helper's goal.

In the following chapters (2.2, 2.3), we describe studies that aimed to probe whether infants and children possess H-NUC and can thus interpret helping interactions merely by conducting a cost-benefit analysis. First, however, we report an attempt to replicate the finding by Hamlin and colleagues (2007) which first demonstrated that infants not only make sense of observed helping interactions, but prefer agents who help over those who hinder.

Chapter 2.1: Toddlers' preference of Helpers over Hinderers

Experiment 2.1

A growing literature suggests that, from a very young age, infants spontaneously engage in third-party social evaluation, drawing inferences about agents' sociomoral dispositions on the basis of their interactions. This proliferating research project was launched by the seminal 2007 study of Hamlin et al. (2007), which showed that 6- and 10-month-olds presented with two characters interacting in a helpful or harmful manner towards a common patient subsequently preferred the former when prompted to choose among the two.

Follow-up studies by Hamlin and colleagues showed this to be a nuanced and sophisticated phenomenon. Already in their first year of life, infants appear sensitive to epistemic states and overt intentions: They prefer intentional over accidental Helpers, but accidental over intentional Hinderers (Woo et al., 2017), and unsuccessful Helpers over unsuccessful Hinderers (Hamlin, 2013a). Additionally, infants show a preference for Helpers only when these know the particular goal the Helpee is trying to accomplish (Hamlin, 2015; Hamlin et al., 2013). Moreover, infants do not choose characters on the basis of the mere valence of the actions they performed, but interpret them in context, preferring a character who 'punishes'—i.e. acts antisocially towards—a previous Hinderer over a character who helps her (Hamlin, 2014; Hamlin et al., 2011).

Beyond instrumental helping, a preference for prosocial characters has been found in a number of other sociomoral domains. In the domain of physical aggression, for instance, infants preferred victims over perpetrators (Kanakogi et al., 2013; Uzefovsky et al., 2020), and third-party characters intervening in a conflict to shield victims from ongoing aggression over passive bystanders (Kanakogi et al., 2017). Similarly, in the domain of resource allocation, infants have been repeatedly shown to prefer fair distributors over unfair ones (Burns & Sommerville, 2014; Geraci & Surian, 2011; Lucca et al., 2018). Modified versions of the manual choice paradigm have also been recently used to investigate whether similar evaluative tendencies exist in non-human animals, such as bonobos (Krupenye & Hare, 2018), capuchin monkeys (Anderson et al., 2017), dogs (Chijiiwa et al., 2015; McAuliffe et al., 2019), and cats (Chijiiwa et al., 2021).

Despite the recent growth of the literature on early sociomoral evaluation, attempts to replicate the findings by Hamlin et al. have yielded mixed results. For example, using the original 'hill scenario', Cowell & Decety (2015) found no significant preference for Helpers in 12- to 24-month-olds (see also Colaizzi, 2016). Similarly, Scarf et al. (Scarf et al., 2012) suggested that low-level perceptual features, rather than inferred sociomoral dispositions, could adequately explain infants' preference for prosocial characters (though see Hamlin, 2015 for a rebuttal of this claim). Using the 'box scenario', Salvadori et al. (2015) found no preference for Helpers across two experimental attempts. A similar lack of preference was documented by Nighbor et al. (2017) with 5- to 16-month-olds, by Vaporova and Zmyj (2020) with 9-, 14-month-olds and 4-year-olds, and by Maxwell and Rafetseder (2015) with preschoolers. Conversely, using the 'ball scenario', Scola et al. (2015) reported a significant preference for prosocial characters in 12- to 24- and 24- to 36-month-olds, as did Chae and Song (2018) with 6- and 10-month-olds, whereas Shimizu et al. (2018) documented a similar, albeit weaker, preference in 15- to 18-month-olds, but not in younger age groups. It is worth noting, however, that previous replication attempts have followed the methods of the original studies to varying degrees of fidelity. Differences in stimuli materials and procedural details might have conceivably affected infants' responses.

In a meta-analysis, Margoni and Surian (2018) reviewed 26 published and unpublished studies using manual choice measures to investigate early sociomoral evaluation. While their analysis revealed an overall significant tendency to prefer prosocial characters across studies, the authors cautioned about the possibility of publication bias and the under-reporting of negative findings (file drawer problem). Importantly, Margoni and Surian also attested the presence of a laboratory effect: Research conducted by Hamlin and collaborators tends to generate larger effect sizes compared to studies done by independent laboratories. On these grounds, the authors called for more and sufficiently powered replications.

We conducted a conceptual replication of the original study by Hamlin et al. (2007). Our study differs from the original in three potentially important ways. Firstly, we tested 15-month-old infants, an age group slightly older than the infants tested in similar studies. While Margoni & Surian's meta-analysis (2018) found no significant effect of age on infants' preference for prosocial characters, the participants' mean age in the studies reviewed was approximately 13 months (390 days). Secondly, we did not present the stimuli in the form of a live puppet show, but as video animations on a screen, which were generously provided to us by Woo and Hamlin. Although Margoni & Surian (2018) found no effect of presentation mode (live versus video), a majority of the studies in their sample were based on live puppet shows. Thirdly, instead of using a habituation procedure, we employed a familiarization procedure, presenting the stimuli for a fixed amount of time across infants. This was aimed at mitigating the problem of fussiness and high drop-out rates, common with older infants when using habituation designs.

Crucially, these modifications were implemented under recommendation of Woo and Hamlin, who used the same video stimuli and familiarization procedure for their own in-laboratory replication of the original Hamlin et al. (2007) study. Here, Woo and Hamlin found a significant preference for the Helper character in a sample of 32 infants (23 of 32; reported in Margoni & Surian's meta-analysis (2018)).

Methods

The article reporting this study received results-blind in-principle acceptance (IPA) at Royal Society Open Science. Following IPA, the accepted Stage 1 version of the manuscript, not including results and discussion, was preregistered on the OSF (https://osf.io/krms8). The preregistration was produced after data collection and analysis.

Piloting phase

Before testing our experimental sample, we conducted a pilot with 24 infants aged 14–16 months. During the piloting phase, we sent video recordings of the participants to Hamlin (written permission for data sharing was obtained from the parents), who kindly provided helpful feedback on the procedure, and we subsequently implemented her suggestions. Testing of the experimental sample began only after Hamlin had confirmed that our procedure was sufficiently close to the original.

Participants

Thirty-two 14- to 16-month-old infants participated in the study (mean age: 15.18 months, range: 431–492 days). The sample size was determined prior to data collection and was twice the sample of 10-month-olds and more than twice the sample of 6-month-olds tested in Hamlin et al. (2007). An additional 19 infants were tested but not included in the final sample due to failing to produce a choice at test (n = 7), inattentiveness during familiarization (n = 5), fussiness (n = 4), experimenter error (n = 2) and technical failure (n = 1). Participants were full-term infants with no reported health or developmental issues. Infants were recruited from the database of the Cognitive Development Center.

Caregivers were informed about the nature and possible consequences of the study, and gave informed consent for their child to participate. We obtained ethical approval for this research from the United Ethical Review Committee for Research in Psychology (EPKEB) in Hungary.

Materials and apparatus

During the familiarization phase, infants were seated in their caregiver's lap in a dimly lit room, approximately 60 cm away from a TV screen of 100 cm diagonal size. The stimuli were generated by Woo and Hamlin using Blender animation software, and were presented on a screen using PsyScope X (Cohen et al., 1993) controlled by a Mac Mini computer.

The objects for the manual choice procedure were printed-out versions of the blue square and yellow triangle characters from the stimuli videos (square: 13 × 13 cm, triangle: 15.5 × 13.5 cm). The printout graphics were converted from RGB to CMYK color space and adjusted, so that the color of the printed characters matched those on screen as closely as possible. Printouts were glued onto figures made of stacked cardboard, to mimic the three-dimensional appearance of the characters in the video. The figures were then wrapped with a transparent plastic cover, to protect them from wear. The figures were attached with removable adhesive putty onto a white board (50 × 36 cm) at a distance of 19 cm from each other, 3 cm from the sides of the board and 3 cm from the bottom of the board.

During the familiarization, Experimenter 1, who ran the study and coded the infants' looking behavior online, was seated in the same room as the child, hiding behind a black curtain. Experimenter 2, who performed the manual choice task, also hid behind the curtain during the familiarization phase. To ensure that Experimenter 2 was blind to condition, she had no visual access to the screen displaying the stimuli.

Procedure and stimuli

Before the familiarization phase, Experimenter 2 briefed the carer on how to position herself for the manual choice task. The carer was instructed to turn her chair away from the screen, place her feet behind a tape marking on the floor and have the child sit on her knees while supporting the child by the ribcage. After this training on the choice phase, the carer was asked to turn back towards the screen for the familiarization phase and to keep her eyes closed for the whole duration of the study.

Familiarization phase. Infants watched a total of six familiarization trials featuring three helping and three hindering events, alternated. Each trial was preceded by a brief attention-getter (a flashing checkerboard accompanied by the sound of a xylophone slide) which played until the child gazed back at the screen. The two familiarization events were matched in timing and overall duration (17 s).

Both events took place on a light-blue sky background and a dark green hill, extending from the bottom left to the top right corner of the screen. The hill plateaued halfway and at the top.

Each event started with a character (a small red circle with eyes pointing to the top of the hill; hereinafter, Protagonist) located at the bottom of the hill. After a bell sound, the Protagonist moved to the intermediate plateau and bounced up and down twice with her eyes directed towards the viewer (2 s). He then attempted to climb the top plateau twice, each time reaching up to two-thirds of the steep incline and sliding back down to the intermediate plateau (8 s). At this point, the Helper or Hinderer appeared, again to the sound of a bell (Helper: from the bottom left of the screen; Hinderer: from the top right of the screen). As the Protagonist attempted to climb the steep incline to the top plateau for a third time, the Helper/Hinderer (whose eyes were directed to the top or bottom of the hill, respectively) moved towards the Protagonist and, with two repeated shoves (accompanied by a knocking sound), pushed the Protagonist up to the top plateau or down to the bottom one (4 s). Upon reaching either the bottom or the top of the hill, the Protagonist came to a standstill, while the other character exited the scene from the location where he initially appeared (3 s).

Each trial ended with a still frame, kept on screen until the infants had looked away for a minimum of two consecutive seconds or until 30 s had elapsed.

Test phase. Immediately after the end of the familiarization phase, the screen turned black and a soft guitar tune started playing (also provided by Woo and Hamlin). Experimenter 2, following a cue from Experimenter 1, entered the testing room from behind the curtain, turned on the light and instructed the carer to assume the previously practiced position for the manual choice task and to close her eyes again afterwards. Experimenter 2 kneeled down in front of the child and addressed her while making eye contact: 'Szia [name of child]! Kivel szeretnél játszani?', which translates to 'Hi [name of child]! Who would you like to play with?'. Then, she lowered her gaze to the chin of the child and flipped over the board with the two characters. The board was moved towards the infant and turned slightly downward at approximately a 30° angle, so that the figures were within the infant's reach but required participants to stretch out their arms in order to touch them. After the board had been flipped over, the experimenter made sure not to pull the board away while the infant was reaching out for a character, as this might convey to the infant that her intended choice was 'wrong' (JK Hamlin 2017, personal communication).

If the infant did not produce any visually guided reaching after approximately 30 s, Experimenter 2 verbally encouraged the infant by saying, for instance, 'Figyelj!' (Pay attention!), 'Nézd meg őket!' (Look at them!), or 'Bátran!' (Be brave!), and repeating the original question. If no choice was produced after 2 min, the experiment was terminated.

The following factors were counterbalanced in the study: the identity of Helper and Hinderer during familiarization (blue square vs. yellow triangle), the order of event presentation (helping first vs. second) and the position of the characters on the board (Helper on the right vs. left side). The condition that each infant was assigned to was randomly selected before testing.

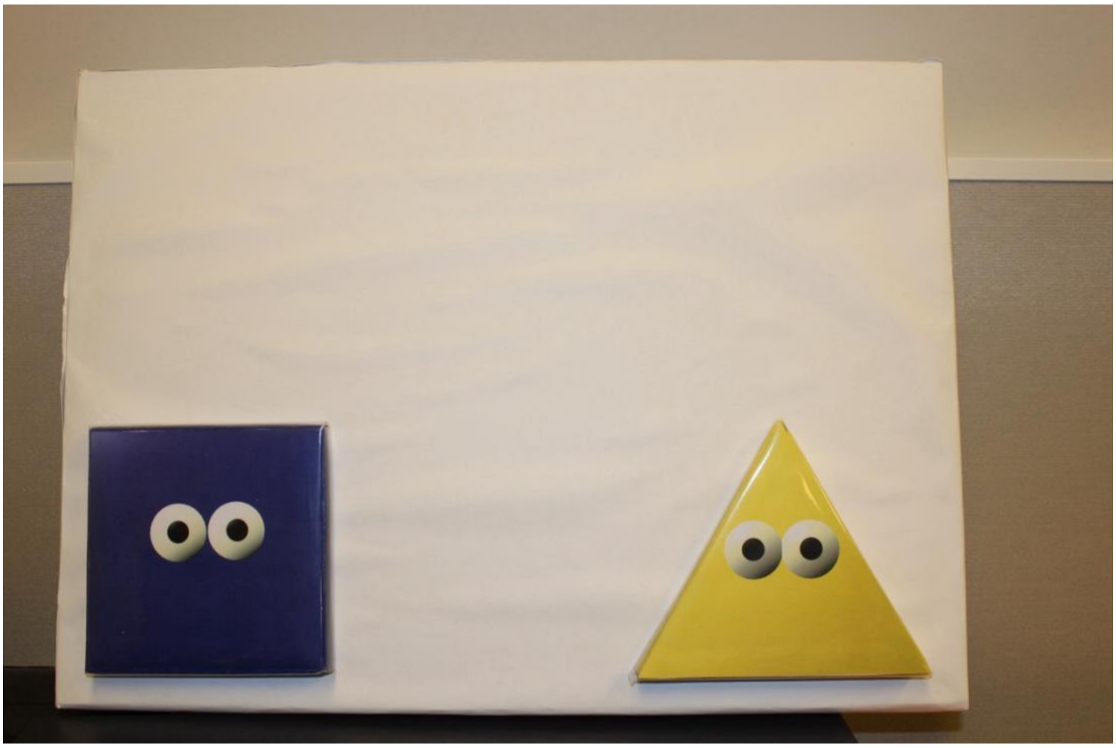

Figure 2.3. Figures used in the manual choice task.

Coding and analyses

The dependent variable was the infants' choice of the Helper or Hinderer character, assessed by their reaching to one of the figures on the board. In order to be counted as a choice, the reaching had to be visually guided: i.e. infants had to look at a character before and while touching it. If infants reached for a figure while looking elsewhere, they were given the opportunity to produce another reach within the 2 min time window. If infants touched both figures, but looked only at one prior to establishing contact, this was coded as a choice for the figure they looked at.

Experimenter 2 judged online whether the infant had reached unambiguously for one of the figures and thus whether to terminate a trial. Choices were coded offline from the videos by Experimenter 1, and recoded by an independent second coder blind to the experimental condition, reaching 93.75% of agreement. Two infants judged by the second coder to have made no clear choice were removed from the final sample and replaced.

In order to be included in the final data analysis, infants had to watch at least 50% of the duration of each helping/hindering event (from the onset of physical contact between the protagonist and the Helper/Hinderer to the end of the pushing action) in all six trials. This stringent criterion of attentiveness was meant to ensure that each infant attended to the crucial social interactions differentiating Helper and Hinderer for a sufficient number of times. Including the manual choice data from the infants who did not meet this criterion did not affect the results.

In order to assess whether infants showed preference for the Helper character, we performed a one-tailed binomial test on the number of infants who chose the helper and the total number of infants included in the analysis against the probability of 0.5 as chance level (as was done by Hamlin et al., 2007), and calculated Bayes factors to assess the strength of evidence for H0 or H1. Statistical analyses were performed with R, the lme4 package (Bates et al., 2015) and the BayesFactor package (Morey & Rouder, 2018). Data are available at https://osf.io/kh5r4.

Results

Hypothesis-driven analyses

Sixteen out of 32 children directed their first visually guided reach to the Helper (one-tailed binomial test $p = 0.57$; 95% CI: 0.344–1.0). Thus, infants did not display a preference for either the Helper or the Hinderer character. When including in the analysis the 5 infants who were excluded due to inattentiveness during the familiarization phase, 20 out of 37 reached for the Helper (one-tailed binomial test $p = 0.371$; 95% CI: 0.394–1.0).

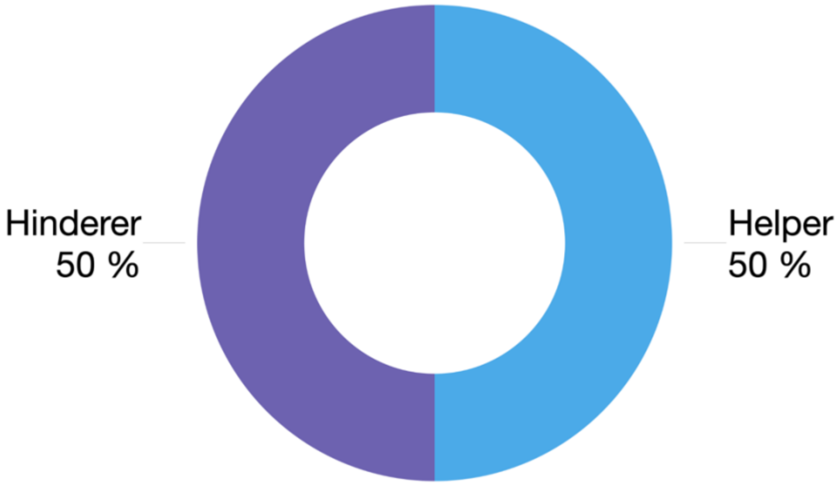

Figure 2.4. Proportion of infants' choices for the Helper and Hinderer in Experiment 2.1.

Additional results

In a Bayesian analysis with a null model of $p = .5$ and an alternative model with a uniform prior (implemented in the BayesFactor package by an 'ultrawide' scale parameter of 1), the data from our study yielded a Bayes factor of 4.62 in favor of H0, indicating moderate support for the null hypothesis of no effect.

Infants' choice was not significantly influenced by their gender (9 of 20 male infants chose the Helper, while 7 of 12 females did), characters' features (20 of 32 infants reached for the yellow triangle), characters' location on the board (17 infants reached for the figure on the right) and order of familiarization events (12 infants reached for the agent they last saw).

During the manual choice, a subset of infants did not unambiguously direct their gaze at both characters before producing a choice. This, however, did not affect the results: 12 of 24 of those infants who looked at both characters chose the Helper, whereas 4 of 8 of those who only looked at one character reached for the Helper.

Since we presented infants with a fixed number of trials in a familiarization design, the present failure may also be due to insufficient exposure to the two characters' actions. Indeed, infants' average looking times from the first three trials (12.81 s) to the last three trials (9.99 s) decreased by 22%, thus failing to meet the habituation criterion previously adopted by Hamlin (i.e. decrease in looking by 50% from the first three to the last three trials). To assess the effects that the overall weak level of habituation had on infants' choices, we examined whether stronger habituation predicted a higher likelihood of reaching for the helper, but found no support for this hypothesis ($\beta = -0.002$, SE $\beta = 0.007$, $p = .816$, logistic regression model).

We also analyzed whether the amount of looking to the two types of familiarization events may have influenced the infants' behavior at test. In line with previous studies, we found no difference in the mean looking times to the still frames following the two events (helping: 11.41 s; hindering: 11.39 s). We fit a mixed-effects linear regression model predicting log looking time from familiarization event type with a subject-random intercept. Model comparison revealed no significant looking time difference between the event types ($\chi 2 = 0.03$, $p = 0.864$). Moreover, infants did not tend to choose the agent they attended to longer on average during familiarization (16 of 32 reached for the character they had looked at longer).

Discussion

In the present study, we attempted to replicate Hamlin et al.'s (2007) finding that infants preferentially reach for helpful over hindering characters. In that study, 92.9% of infants exhibited such preferences (14 of 16 10-month-olds and 12 of 12 6-month-olds). By contrast, only 50% of infants did so in our study (16 of 32). Therefore, we could not reproduce the original findings. There are several potential explanations for such a failure. Our study differed from the original in three potentially relevant ways: Firstly, we tested infants from an older age group (15-month-olds, rather than 6- and 10-month-olds); secondly, we used three-dimensional animated videos rather than a live puppet show to expose infants to the helping and hindering events; and thirdly, we used a familiarization rather than a habituation design.

Any of these deviations from the original study may have potentially contributed to our results. For instance, it is conceivable that six familiarization trials were insufficient for infants to learn about the agents' respective dispositions. Supporting this possibility, the average decrease in looking times during familiarization was insufficient to reach the habituation criterion adopted by Hamlin in previous studies (decrease of 50% from the first three to the last three trials). It should be noted, however, that prior studies (Geraci & Surian, 2011; Hamlin et al., 2013; Kanakogi et al., 2017) and the in-laboratory replication onto which our study was modeled successfully elicited a preference for prosocial characters by means of familiarization.

Alternatively, infants may have had trouble mapping the cardboard replicas of helper and hinderer to the three-dimensional animated characters they were familiarized with. While this remains a genuine possibility, several studies reported preferential reaching for replicas of prosocial characters presented on screen (Geraci & Surian, 2011; Kanakogi et al., 2017; Powell & Spelke, 2018b; Scola et al., 2015), and the meta-analysis found no effect of presentation type on infants' preferences (Margoni & Surian, 2018).

Our study used animations and familiarization following recommendations by Woo and Hamlin, who found these stimuli and design to be suitable for eliciting social evaluation in infants older than 12 months of age. It should be noted, however, that the percentage of infants reaching for the helper in their in-laboratory replication was lower than in the original study (Hamlin et al., 2007), and failed to show the effect in two additional samples of younger infants (8-month-olds: 21/32; 10-month-olds: 15/32; as reported in the supplementary materials of Margoni & Surian, 2018). These differences raise the possibility that familiarizing infants to animations may not be as effective in eliciting social evaluation as habituating them to live-action puppet shows.

It is also possible that other unforeseen methodological differences, some of which may be hard or impossible to control for, contributed to our failed replication. Such differences may concern, for instance, the physical set-up of the testing environment, the cultural background of the population tested or, more likely, the behavior of the experimenters involved in the study. On this note, it is, however, worth noting that, unlike other replication attempts, ours benefited from the close and careful scrutiny of the experimenters' behavior by Hamlin herself. Her feedback during the piloting phase allowed us to fine-tune the procedure of character presentation in ways that other studies could not.

Finally, current evidence suggests that the underlying effect size of infants' preference for helpful characters may be smaller than originally assumed. The meta-analysis by Margoni & Surian (2018) estimated that on average 64% of infants in the studies reviewed reached for the prosocial character. Importantly, however, the strength of infants' preference was affected by the sociomoral domain tested: 77% of infants preferred the prosocial character after observing giving versus taking events, 69% after observing fair versus unfair distributions and only 63% after observing helping versus hindering events. Although instrumental helping represented the domain with the lowest percentage of infants' choice of the prosocial agent, this was nevertheless considerably higher than the percentage (50%) obtained in our study.

Margoni & Shepperd (2020) have argued that individual replication studies ought not to be considered as confirming or disconfirming an effect, but rather should be pooled together to produce a better estimate of the true underlying effect size of the phenomenon at hand. If original studies are underpowered, as is often the case in infant research, replications with a relatively wide range of results may technically be taken as confirming the original finding if they fall within a 'prediction interval' of potential outcomes. This said, our proportion of 50% helper choices falls outside the value range (0.59–1.0) defined by the 95% prediction interval proposed by Margoni & Shepperd for a replication of Hamlin et al.'s (2007) study with n = 32, and thus cannot be considered confirmatory.

Recently, there has been a large-scale, multi-lab replication attempt of the original Hamlin et al. (2007) study (Manybabies 4: Lucca et al., 2024; results presented at BCCCD24). Five- to ten-month-old infants from 35 labs across the world participated in a habituation design, using videotaped stimuli of the live "hill" puppet show, and styrofoam replicas of the Helper and Hinderer figures in the manual choice task. As in the original paper, infants were either assigned to the "social" condition, in which the protagonist was an agent attempting to climb up a hill, or a "non-social" condition, in which an inanimate ball was pushed up or down the hill. Out of 352 infants in the social condition, 175 chose the Helper (49.7%), while in the non-social condition, 174 of 332 (52.4%) chose the push-up agent. These latest results also indicate that the effect may not be as robust as initially assumed.

Replicability issues taken aside, the manual choice measure has some further caveats which have thus far been insufficiently considered in the literature. First, it inevitably conflates infants' understanding (of helping) and their preference (for Helpers), limiting its use in assessing the former. For instance, the finding that infants prefer a Helper who is knowledgeable about the Helpee's goal but not one who lacks such knowledge (Hamlin et al., 2013) could be taken as evidence that infants do not consider the latter a genuine helping action or that they are not motivated to associate with an unwitting aide. Second, it presupposes that observing the Helper interacting with third parties is enough for infants to attribute to this agent a broad prosocial disposition—sufficiently broad, in fact, to encompass interactions with unrelated others, such as the infants themselves (Wynn, 2009). The reason why infants are expected to produce such an inductive leap are, to our knowledge, yet to be clarified. It is not obvious why infants would have such a dispositional bias in interpreting others' social behaviors. Several studies suggest that infants do not generalize (affiliative or antagonistic) social behaviors to novel targets [e.g., dominance (Mascaro & Csibra, 2012), giving (Tatone et al., 2015), helping (Pepe & Powell, 2023), and comforting (Kudrnova et al., 2023)]. Instead, they form stable representations of the particular interactions within which such behaviors originally occurred, which allow them to monitor underlying social relations. Rather than being prepared to inductively infer individuals' dispositions, this literature suggests that infants may instead attend to social interactions primarily to discover the relational make-up of their social surroundings. Compounding the issue, it is not clear how infants relate the characters presented on stage or on the screen to the replicas they are prompted to choose from during the manual choice test; i.e., whether they assume identity between the two sets, treat them as different tokens of the same abstract concept, or interpret them as equivalent symbols standing for a fictional agent (Revencu & Csibra, 2020).

In conclusion, the present replication contributes to broader methodological debates on the replicability of findings in developmental science, and sheds further light on the robustness of the phenomenon of early sociomoral evaluation and the conditions under which it can be reliably elicited.

Chapter 2.2: Infants' understanding of helping as a second-order goal

Although some prior studies purport to show that human infants possess an early-emerging concept of helping behavior, there is a dearth of research investigating what this concept precisely entails. Previously, we outlined what we hypothesize a mature understanding of helping to consist in (H-NUC, see the introduction to this section). In the experiments that are subject of this chapter, we tested whether 12-month-old infants would attribute such a goal to an agent after having been familiarized with a scenario where the agent minimizes (Experiment 2.2.1) or decreases (Experiment 2.2.2) the action cost a Helpee has to incur to reach her goal.

We decided to focus on infants around one year of age, as at that point they have been found to possess prerequisite capacities for H-NUC. First, as reviewed in Section 1, they have a robust understanding that agents should act efficiently and thus minimize their action costs. Second, as discussed in the introduction to this section, they were found to preferentially reach for Helpers over non-Helpers, and this preference seems grounded in sophisticated and nuanced interpretation of Helpers' motives.

Note that there are two slightly different versions of H-NUC that an observer might use: The Helper may aim to *maximize* the Helpee's utility, or to *increase* it compared to a situation where the Helpee does not receive any help. The latter is satisfied whenever the Helpee, as a result of being helped, is somewhat better off than if she were acting alone. Think about someone carrying some of an old lady's heavy shopping bags: She now has to bear less weight. On the other hand, someone who maximizes another's utility chooses, from the available options for intervention, the one that yields the highest possible utility for the Helpee. For example, if my partner offers to do a load of laundry for me, it may be most useful to me at that moment if she washes jeans rather than bedsheets, even though both would lower the amount of housework I have to do. In Experiment 2.2.1, we tested whether infants would expect a Helper to minimize the Helpee's action cost (i.e., utility *maximization*), and in Experiment 2.2.2, to reduce the Helpee's action cost (i.e., utility *increase*).

Experiment 2.2.1

In Experiment 2.2.1, infants watched an agent (the Helpee) approach a goal object by detouring around a barrier. During familiarization, the Helper (when present) reduced the Helpee's action cost by opening a door and thus allowing him to take a shorter path. At test, two doors were blocking direct access to the goal, one closer and one further away from the goal, but equidistant from the Helper. The Helper either opened the closer door which freed a relatively shorter path, as during familiarization ("Consistent" test trial), or opened the further-away door (“Inconsistent” test trial). If infants interpreted the familiarization events as a helping interaction and take the Helper’s goal to be maximizing the Helpee’s utility, they should find it inconsistent with this goal when the Helper, at the same cost to himself, selects the option that is inferior for the Helpee.

Methods

Participants

Twenty-four 12-month-old infants (13 male, age range: 11.2 – 13 m., mean age: 12.1 m) participated in Experiment 2.2.1. An additional 17 infants were tested, but had to be excluded due to fussiness (n = 5), lack of attention to test events (n = 8), outside noise from a construction site during data collection (n = 3), and parental interference (n = 1). Recruitment, ethical approval, consent, and compensation were the same as in Experiment 2.1.

Apparatus

Infants were seated in their caregiver’s lap held by the hips in a darkened, soundproof room, 80 cm away from a 40-inch monitor. The stimuli were 2-D animated videos created using Adobe Animate CC software and presented with MATLAB (The MathWorks) using the Psychophysics toolbox extension (Brainard, 1997). Videos of the infants were recorded during the session. Infants’ looking behavior was coded on-line to determine when to start a new trial, and later manually coded off-line to measure looking time.

Procedure and stimuli

Caregivers were instructed not to interact with the infants during the experiment. Their eyes were covered with opaque sunglasses. Before each trial, a short attention-getting clip was shown until the infant looked at the screen. Trials ended either when the infant looked away for a minimum of 2 seconds consecutively after the video had stopped, or if 8 seconds (familiarization) resp. 60 seconds (test) had elapsed since the end of the video.

Familiarization. Infants watched eight familiarization videos. We presented two types of videos: “Solo” and “Helping”; the presentation order was SSHHSSHH. In half of the videos, the goal object was located on the top part of the screen, in the other half on the bottom; the order was counterbalanced.

In the videos, a goal object (strawberry) was always located on the far right side of the screen (either in the top or bottom corner). In the middle of the screen, there was a vertical barrier containing two openings. One of the openings – the one located closer to the goal – was blocked by a light blue door.

In the Solo familiarization videos (13 s), a character (the Helpee, a yellow circle with googly eyes directed rightward) appeared from the left side and moved toward the right (3 s), then paused in front of the barrier, in the middle between the two openings (1 s). Then, the Helpee moved through the unblocked opening towards the goal (5 s). After making contact with the strawberry, the Helpee bounced up and down (4 s).

The Helping familiarization videos (16 s) were similar, except that here, another character – the Helper (a green square with googly eyes directed leftward) – was located on the right side of the barrier, equidistant to the two openings and close to the strawberry. Again, the helpee entered from the left and paused in front of the barrier (3 s). The Helper then moved toward the door, opened it, and returned to his initial position (5 s). The Helpee then approached the strawberry as in the Solo familiarization clips (8 s).

Key elements of the videos (the Helpee moving, beginning the goal approach through the opening in the barrier, reaching the goal; the door opening) were accompanied by sound effects.

Test. Infants received two test trials: a consistent and an inconsistent trial (15 s each). The scenario in the test videos was similar to the one in the helping familiarization trials, except that now both doors were closed. The reward was located in the opposite location from the last familiarization trial (so, if it was on the bottom of the screen in the last familiarization video, it was at the top at test, and vice versa). In both videos, the Helpee again approached from the left (3 s) and paused in front of the barrier (1 s). Then, he briefly moved towards each of the doors (3 s). This was meant to highlight to infants that both possible paths were now blocked.

In the Consistent test event, the Helper opened the door that was closer to the goal (6 s). In the Inconsistent event, she opened the one further away from the reward. After this, in both events, the sound that, during familiarization, preceded the onset of the Helpee's goal approach movement was played again, but now the video ended before the Helpee started moving.

Stimuli can be accessed at https://osf.io/adyqn/.

We counterbalanced the location where the goal object was located in the first trial (top vs. bottom), the order of the Helpee's partial approach movements in the test trials (to the upper opening first vs. to the lower opening first) and the order of test trials (Consistent first vs. Inconsistent first).

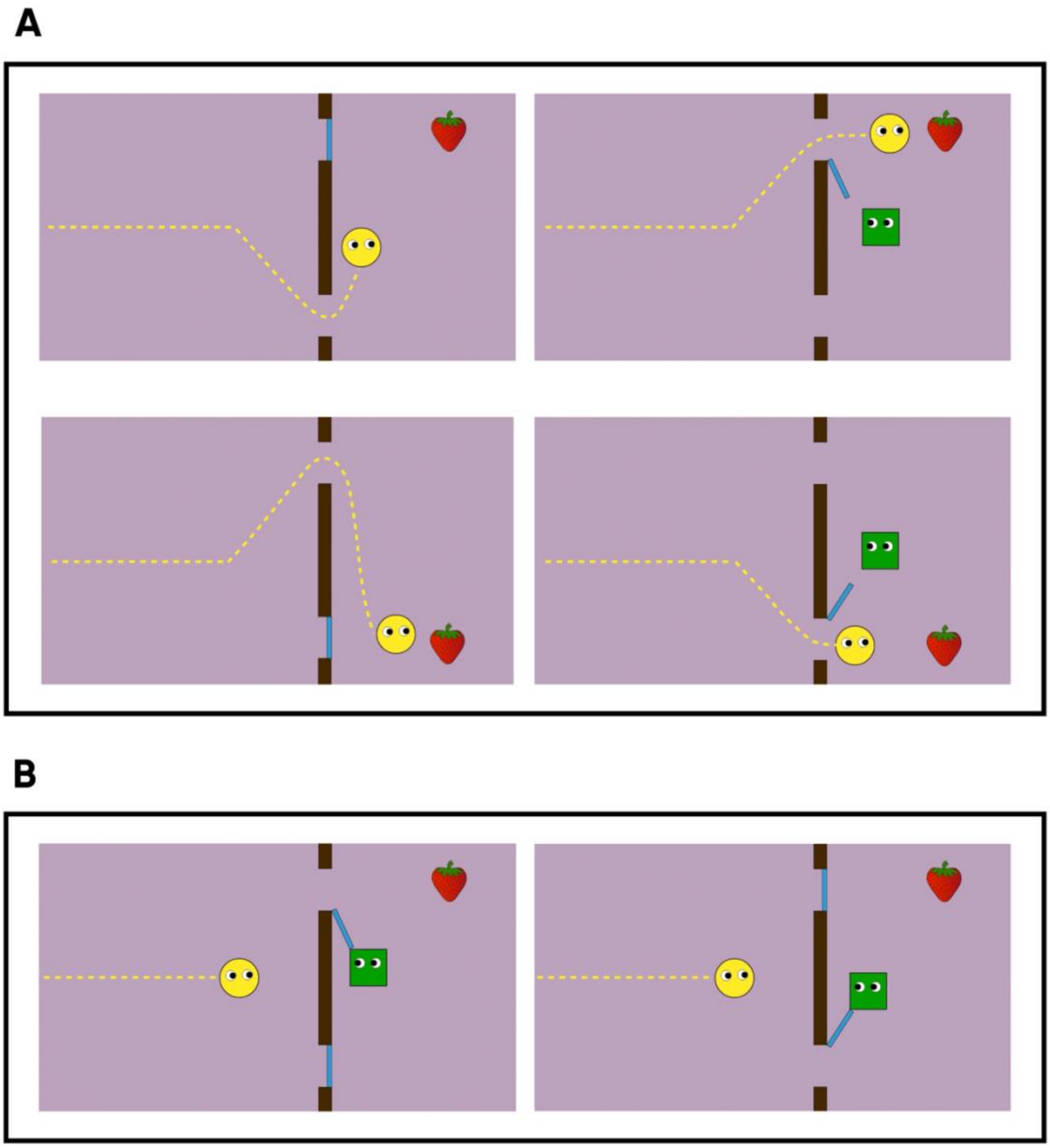

Figure 2.5. Stimuli used in Experiment 2.2.1. During Familiarization (A), the Helpee (yellow) approached the goal object. When alone, the Helpee had to take a longer path (left column); when the Helper (green) was present, he could take a shortcut. At Test (B), the Helper opened either the door that allowed the Helpee to take the short path (Consistent test event, left) or the long path (Inconsistent test event, right).

Coding and analyses

Infants' looking in the test trials was measured from the point of time when the two test events began to diverge, i.e. when the Helper started approaching one of the doors to open it. The looking behavior was manually coded off-line to measure looking times using the same criteria as online coding and reviewed for the pre-defined exclusion criteria (fussiness; parental interference; experimenter error; lack of attention during the test trials: i.e. failing to look for at least 50% of each trial in total, as well as at least 50% to each crucial door-opening action). The looking times of 50% of the participants was reanalyzed by an independent second coder who was blind to the hypothesis and to the condition of the stimuli shown. The recoded data were strongly correlated with the original data ($r = 0.99$, $p < .001$). Because of this high level of agreement, data from the first coder was used for analyses (in this and all following experiments in this chapter).

The raw looking times were base-10 log-transformed for analyses (Csibra et al., 2016), but for descriptive statistics and plots we use the raw data. We conducted both Bayesian and frequentist statistical analyses. For the Bayesian analysis, as in Section 1, we used the method recommended by Csibra et al. (2016) for looking-time data. For the frequentist statistical analyses, we conducted a paired sample two-tailed t-test on the data, and a 2x2 mixed ANOVA with Order as a between-subject and Trial as a within-subject factor to check for order effects. Statistical analyses and plotting were performed in R, version 4.3.1 (R Core Team, 2023).

Data are available at https://osf.io/adyqn/.

Results

There was no significant difference in looking times to the two test videos ($M_{\text{consistent}}$ = 17.78 s, $SD_{\text{consistent}}$ = 11.97 s; $M_{\text{inconsistent}}$ = 20.75 s, $SD_{\text{inconsistent}}$ = 13.38 s; $t(23) = 1.24$, $p = 0.229$), see Figure 2.6. An ANOVA showed no significant main effects, but a significant effect of the Order by Trial interaction ($F(1,22) = 7.61$, $p = 0.012$). Subsequent t-tests showed that infants who saw the Consistent test trial first did not look significantly longer at either test event ($t(11) = 0.92$, $p = 0.379$), whereas those who saw the Inconsistent trial first looked significantly longer at the Inconsistent event ($t(11) = 3.09$, $p = 0.01$). This pattern suggests that there was an effect of Trial type, which interacted with infants' tendency to look longer at the first test trial they saw.

In the Bayesian analysis, we obtained a BF of 0.54, which constitutes anecdotal evidence for the null hypothesis of no effect.

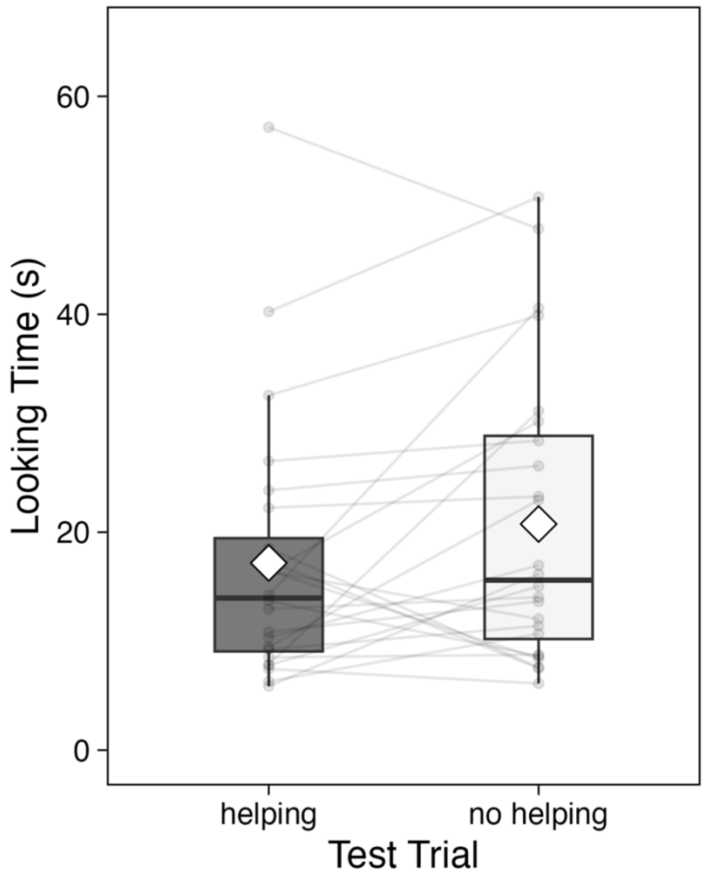

Figure 2.6. Boxplot of average looking times (in seconds) toward the test events in Experiment 2.2.1. Light grey lines connect the looking times of individual participants, white diamonds indicate means, horizontal lines indicate medians, boxes indicate middle quartiles, and whiskers indicate points within 1.5 times the interquartile range from the upper and lower edges of the middle quartiles.

Discussion

In Experiment 2.2.1, we aimed to test whether 12-month-old infants would expect an agent who previously reduced a Helpee's action cost to select an action which allowed her to take the shortest path possible to her goal. The results we found were inconclusive. On the one hand, there was evidence for an order effect interacting with the effect of the test trial type, such that infants tended to look longer at the Inconsistent event, but only if this was presented in the first test trial. However, while this type of order effect is not uncommon in violation-of-expectation looking time designs (Baillargeon, 1987; Csibra et al., 1999; S. Liu et al., 2017; Mascaro & Csibra, 2012; Tatone et al., 2023), the looking difference between the test events in our experiment was not strong enough to yield a significant main effect. The Bayes factor similarly did not indicate a difference in looking times to the test trials, instead providing anecdotal evidence for the null hypothesis.

One way to interpret this pattern of results is that while some of the participants may have interpreted the stimuli as we intended and ascribed the goal of utility maximization to an agent who had previously reduced a Helpee's action cost, the task we posed to infants may have been too demanding. A concept of helping which specifies that the Helper's goal be to maximize the Helpee's utility requires a sophisticated comparison of multiple hypothetical or counterfactual scenarios: Specifically, the observer has to compare the effects of all the possible action options the Helper may pursue—some of which may increase, some of which may decrease the Helpee's utility, and some of which may be irrelevant.

In contrast, infants may instead possess a concept of helping where the Helper aims to increase the Helpee's utility, without requiring that the Helper assist in the optimal way. In this case, infants would merely have to compare the scenario where the Helpee does not receive aid with the one in which the Helper intervenes, and determine whether the Helper's action led to a better outcome for the Helpee.

This could explain the results we found in Experiment 2.2.1: In our stimuli, opening even the further door left the Helpee in a better state compared to a situation where no aid is given, in which case the Helpee could not access the goal at all. In other words, infants may have merely checked whether the Helpee was somewhat better off as a result of the Helper's action, and because this was the case in both test events, found both consistent with the goal of helping.

To address this possibility directly, we conducted Experiment 2.2.2.

Experiment 2.2.2

In Experiment 2.2.2, we tested whether 12-month-old infants would expect a Helper to increase the Helpee's utility by performing an action that would allow the Helpee to take a relatively shorter path, instead of a superficially similar-looking action which did not have any bearing on the Helpee's goal pursuit.

Infants were again familiarized with a scenario where a Helpee had to take a longer path to reach a reward when he was by himself, and could take a shortcut when a Helper intervened. At test, the Helper again removed an obstacle, whose location differed across test trials: In the Consistent test event, the obstacle was obstructing the Helpee's most direct path to the goal object, therefore, the Helper's action reduced the Helpee's cost. In the Inconsistent test event, the obstacle was placed at a location in the scene where it was not blocking the Helpee's path. When the Helper moved this obstacle aside, the action looked similar in its first-order action features to the one performed during familiarization, but was not compatible with the goal of helping.

Methods

Participants

Twenty-four 12-month old infants (10 female, age range: 11.5 – 12.5 m., mean age: 12 m) participated in Experiment 2.2.2. An additional 12 infants were tested but had to be excluded from the sample due to fussiness (n = 2), lack of attention to test events (n = 4), outside noise from a construction site during data collection (n = 2), experimenter error (n = 1), and having been retested a second time due to a scheduling error (n = 3). Recruitment, ethical approval, consent, and compensation were the same as in the previous experiments.

Apparatus

The apparatus was the same as in Experiment 2.2.1.

Procedure and stimuli

The procedure was the same as in Experiment 2.2.1.

Familiarization. Infants watched eight familiarization videos. Again, two types of videos were shown: “Solo” and “Helping”; the presentation order was SH-SH-SH-SH. Within each of the Solo-Helping trial pairs, the physical layout of the scene was the same, but it varied across pairs. This allowed infants to observe the cost-reducing effect of the helping action directly, as the Helpee could take a relatively shorter path in the same environment after being helped, but also provided variability in the Helpee’s goal-approach motion paths across trials. The layout in the third and fourth pair of trials was the same as in the first and second, but rotated horizontally and vertically. The duration of trials varied, such that the first and third pair of videos were longer (due to a longer approach motion being necessary to reach the goal in this layout). There was always a goal object (strawberry) located somewhere in the scene, along with a wall containing two openings, of which one was obstructed by a black square block.

In the Solo familiarization videos, a character (the Helpee, a yellow circle with googly eyes as in Study 1) appeared on the screen (2 s) and paused (1 s). Then, the Helpee moved along the shortest possible trajectory through the non-obstructed opening (9/7 s) and finally reached the strawberry (1 s).

In the Helping familiarization videos, another character—the Helper (a green square with googly eyes)—was located in the middle between the openings in the barrier. Again, the Helpee entered and paused (2 s). The Helper then moved toward the opening that was obstructed by a block, pushed the block aside, and returned to his starting position (7/5 s), upon which the Helpee approached the strawberry as in familiarization (3/4 s).

Pre-test. The last trial of the familiarization phase served as a pre-test event. Here, the physical layout of the scene was the same as in the subsequent test events (save for the number of obstacles present), but the Helpee was alone, as in the Solo familiarization events. There was a strawberry located at the center of the screen, and two horizontal barriers at equal distance from the strawberry, one above, one below it. Both barriers had openings that were blocked by an obstacle, respectively. The Helpee approached from the bottom of the screen and paused (3 s), then approached the strawberry by detouring around the lower barrier (4 s).

Test. After the familiarization and pre-test events, infants watched two test videos: a Consistent and an Inconsistent trial (10 s each). Each test event featured both the Helpee and the Helper. The layout was the same as in the pre-test event, except now only one opening was blocked: In the Consistent trial, a block covered the gap in the lower barrier (thus obstructing the Helpee's most direct path to the goal), while the upper barrier's gap was unobstructed; conversely, in the Inconsistent trial, the opening in the upper barrier was covered by a block, while the gap in the lower barrier was free (such that the Helpee could approach the goal on a direct path). In both videos, the Helpee again approached from the bottom (2 s) and paused in front of the barrier. At this point, the Helper moved away the block from the opening (4 s), and the Helpee approached the strawberry on a straight, upward path (4 s). The behavior of the Helper was thus similar across the two test events, except that she moved upward or downward to remove the obstacle; the behavior (including the movement trajectory) of the Helpee was identical in the test events.

We counterbalanced the order of the familiarization videos (goal at the top vs. at the bottom of the screen in the first familiarization trial), and the order of test trials (Consistent first vs. Inconsistent first).

Coding and analyses

The coding procedure, exclusion criteria, and data analyses were the same as in Experiment 2.2.1.

Data recoded by an independent second coder (50% of participants) were strongly correlated with the original data ($r = 0.98$, $p < .001$).

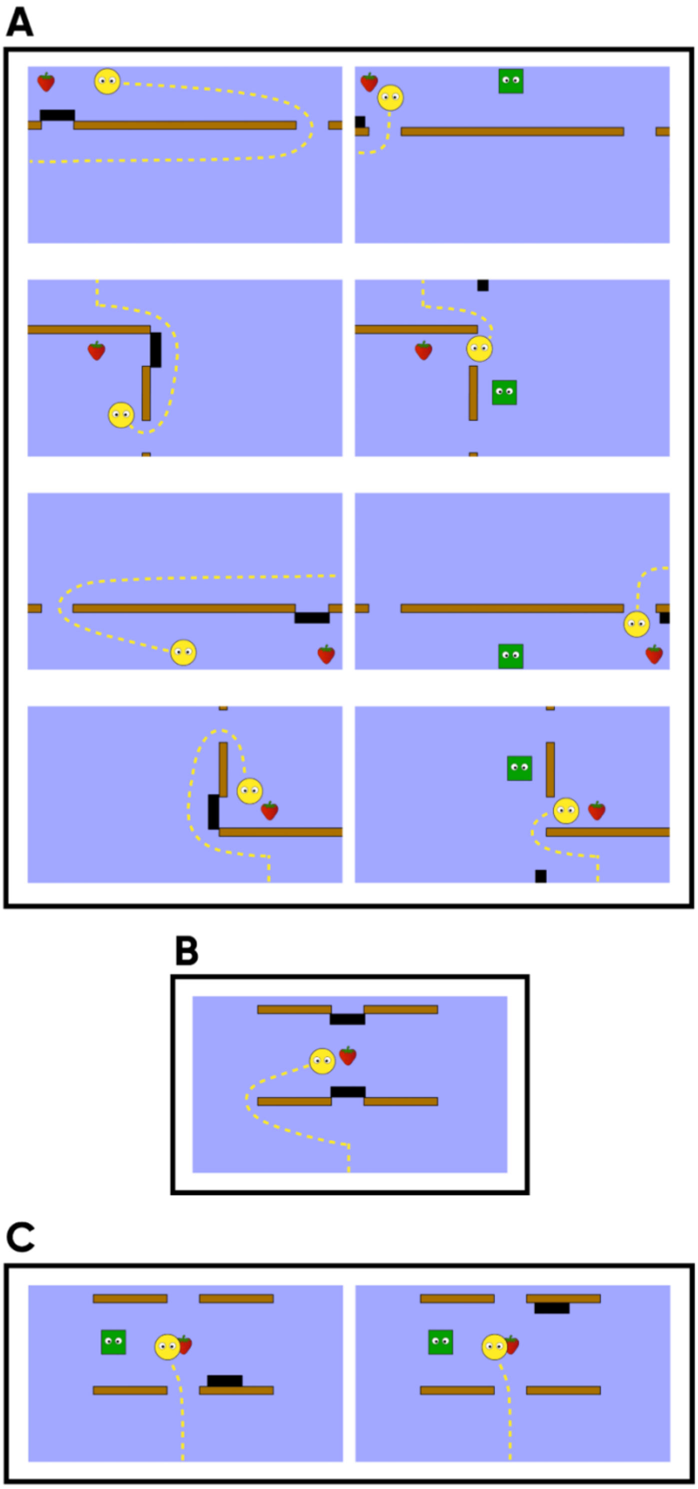

Figure 2.7. Stimuli used in Experiment 2.2.2. During Familiarization (A), the Helpee (yellow) approached the goal object. When alone, the Helpee had to take a longer path (left column); when the Helper (green) was present, he could take a shortcut. In the pre-test event (B), which had the same spatial layout as the test events, the Helpee approached the goal by detouring around the obstacle. At Test (C), the Helper either removed an obstacle that was in the way of the Helpee so that the Helpee could take the most direct path (Consistent test event, left), or removed an obstacle that was located elsewhere, even though the shortes path for the Helpee was already free (Inconsistent test event, right).

Results

There was no significant difference between the looking times to the test trials ($M_{consistent}$ = 16.35 s, $SD_{consistent}$ = 10.28 s; $M_{inconsistent}$ = 15.92 s, $SD_{inconsistent}$ = 8.84 s; $t(23) = 0.38$, $p = .71$), see Figure 2.8. There were no significant main effects in the ANOVA, but the Order by Trial interaction was significant ($F(1,22) = 7.53$, $p = .012$). Subsequent t-tests showed that when splitting the sample by Order, looking time patterns did not differ significantly for either those infants who saw the Consistent trial first nor for those who saw the Inconsistent trial first (Consistent-first: $t(11) = 1.89$, $p = .085$; Inconsistent-first: $t(11) = -2.006$, $p = .07$). However, a paired t-test on looking times grouped by trial position (first vs. second trial presented to infants) showed that overall, infants looked significantly longer at the first event they saw (M_{first} = 18.45 s, SD_{first} = 10.87; M_{second} = 13.82, SD_{second} = 7.38; $t(23) = 2.8$, $p = .01$).

In the Bayesian analysis, we obtained a BF of 0.11, which constitutes substantial evidence for the null hypothesis.

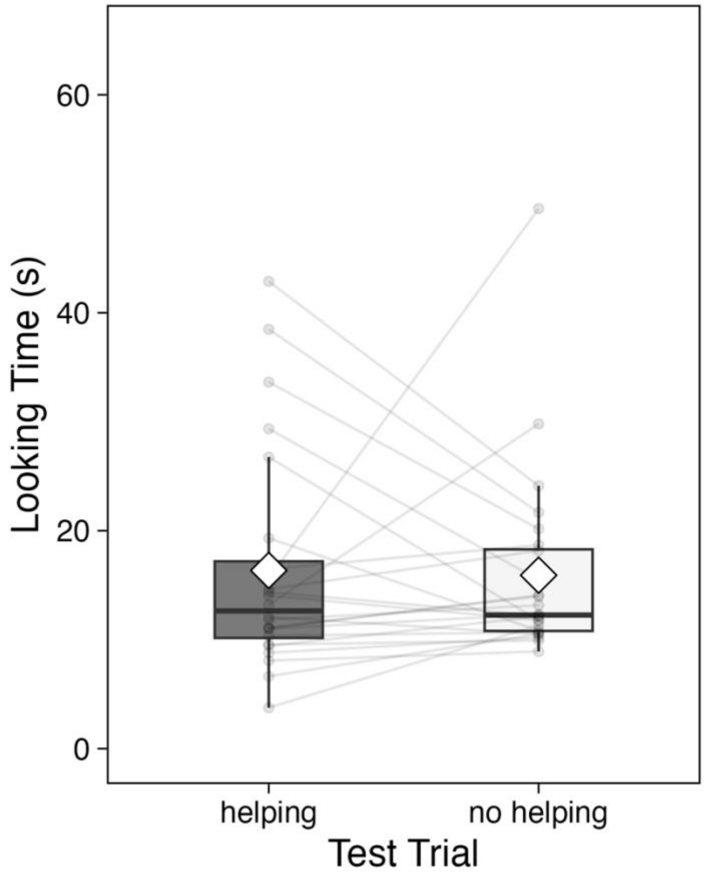

Figure 2.8. Boxplot of average looking times (in seconds) toward the test events in Experiment 2.2.2. Light grey lines connect the looking times of individual participants, white diamonds indicate means, horizontal lines indicate medians, boxes indicate middle quartiles, and whiskers indicate points within 1.5 times the interquartile range from the upper and lower edges of the middle quartiles.

Discussion

The results of Experiment 2.2.2 did not support our hypothesis that infants take the goal of a Helper to be reducing the Helpee's action cost. On the contrary, the Bayesian analysis provided evidence for the null hypothesis. Unlike in Experiment 2.2.1, although we also found a significant interaction of Order and Trial type, this was due to an order effect: Infants simply looked longer at whichever test trial they saw first.

We had expected that the task posed to participants in Experiment 2.2.2 would be easier than the one in Experiment 2.2.1, especially if infants possess a concept of helping as second-order utility increase rather than utility maximization. If infants ascribed the goal of helping as we conceived it, they should have looked longer in the Inconsistent test trial, where the Helper performed an action that looked similar to the helping action but did not have an effect on the Helpee.

We found that 12-month-olds in our experiment did not show this pattern in their responses. (In Appendix C, we report an incomplete replication of Experiment 2.2.2 with 18-month-old toddlers. We abandoned data collection after testing 13 participants, since at that point the analysis of the looking times yielded a BF of 0.14, also indicating support for the null hypothesis of no effect.) The data are instead consistent with the possibility that infants simply ascribed first-order non-social goals to both agents: The Helpee approaches a strawberry, while the Helper pushes a block. In this case, any (efficient) block-pushing action at test would be consistent with the goal previously set up for the Helper.

Another possibility is that the stimuli were overall too challenging for infants, such that they failed to establish goal representations for any of the agents. We showed infants a helping interaction in four different spatial layouts (see Figure 2.7), which we hoped would help infants represent the figures as efficient and goal-directed agents (Csibra, 2008b; Csibra et al., 1999). It may be, however, that infants' working memory capacities were overly burdened by processing the changing environments and figures' movements within them. To rule out the possibility that participants were simply confused and did not reason about agents' goals and efficiency at all, we ran Experiment 2.2.3 as a control condition.

Experiment 2.2.3

In Experiment 2.2.3, we tested whether infants successfully attributed a non-social goal goal to the Helpee. We presented participants of the same age group with the familiarization stimuli from Experiment 2.2.2, and at test varied whether the Helpee acted individually efficiently in approaching her goal, either moving towards it directly, or detouring. The Helper behaved identically across test trials, in both cases removing an obstacle which blocked the Helpee's direct path. We predicted that if infants managed to ascribe a goal to the Helpee, they should look longer when she performed an unnecessarily costly action. Thus, we aimed to conceptually replicate previous findings showing that infants expect agents to behave efficiently.

Methods

Participants

Twenty-four 12-month-old infants (13 male, age range: 11.5 – 12.5 m, mean age: 11.9 m) participated in Experiment 2.2.3. An additional 16 infants were tested but had to be excluded due to fussiness (n = 4), failure to meet the pre-defined attentiveness criteria (n = 5) experimenter error (n = 4), parental interference (n = 2) and technical problems (n = 1). Recruitment, ethical approval, consent, and compensation were the same as in the previous experiments.

Apparatus

The apparatus was the same as in Experiments 2.2.1 and 2.2.2.

Procedure and stimuli

The procedure was the same as in Experiments 2.2.1 and 2.2.2.

The stimuli were the same as in Experiment 2.2.2, except for the test trials. In both test trial videos (11 s each), the block was in front of the opening in the lower barrier, thus obstructing the Helpee's direct path to the reward, and the Helper pushed this block aside (as in the Consistent test trial in Experiment 2.2.2). In the Consistent trial video, the Helpee approached her goal in the most direct path, moving straight upward (4 s). In the Inconsistent trial video, the Helpee approached the goal by detouring around the side of the barrier, moving along the same curvilinear path as in the pre-test trial where the direct path had been blocked (4 s). To equate the duration of the two videos, the Helpee moved faster in the Inconsistent trial, as here her path was longer.

Coding and analyses

The coding procedure, exclusion criteria, and data analyses were the same as in Experiments 2.2.1 and 2.2.2. The recoded data were again strongly correlated with the original data ($r = 0.99$, $p < .001$).

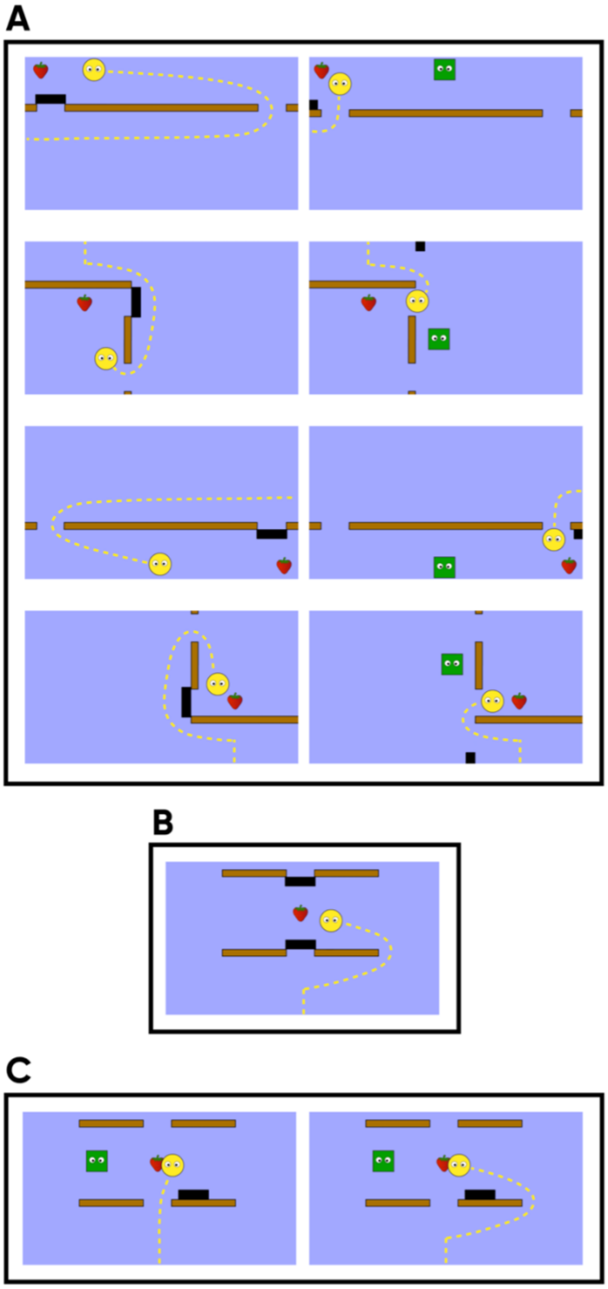

Figure 2.9. Stimuli used in Experiment 2.2.3. The Familiarization (A) and pre-test event (B) were the same as in Experiment 2.2.2. At Test (C), the Helper helped in both test trials, while the Helpee approached his goal either efficiently, on the most direct path (Consistent test event, left), or inefficiently, by detouring around the side of the wall (Inconsistent test event, right).

Results

Infants looked longer to the Inconsistent compared to the Consistent test event ($M_{inconsistent}$ = 17.32 s, $SD_{inconsistent}$ = 16.54 s, $M_{consistent}$ = 10.02 s, $SD_{consistent}$ = 7.59 s, $t(23) = 2.6$, p = .016). A 2x2 mixed ANOVA with Order as between-subject and Trial as within-subject factor showed only a significant main effect of Trial ($F(1,22) = 6.5$, p = .018), there was no significant effect of Order ($F(1,22) = 1.81, p = .193$) and no significant Order by Trial interaction ($F(1,22) = 0.07, p = .801$).

In the analysis using Bayesian statistics, we obtained a BF of 323.59, which constitutes strong evidence in favor of the hypothesis.

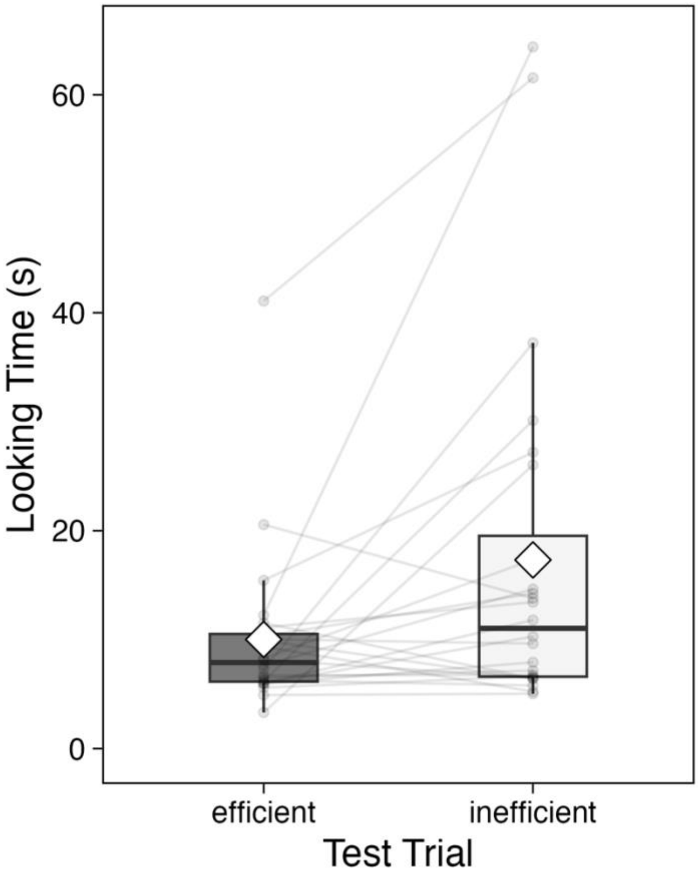

Figure 2.10. Boxplot of average looking times (in seconds) toward the test events in Experiment 2.2.3. Light grey lines connect the looking times of individual participants, white diamonds indicate means, horizontal lines indicate medians, boxes indicate middle quartiles, and whiskers indicate points within 1.5 times the interquartile range from the upper and lower edges of the middle quartiles.

Discussion

As predicted, 12-month-old infants looked longer when an agent, the Helpee, moved towards the goal object in an inefficient manner after having previously done so efficiently, showing that infants succeeded in attributing the goal of approaching the strawberry to the Helpee.

This result replicates previous findings that preverbal infants expect agents to adhere to the principle of efficiency. It also extends these previous findings by demonstrating that one-year-olds can also set up a goal representation in a context where multiple agents perform distinct actions. The finding speaks against the possibility that infants in Experiment 2.2.2 did not ascribe a helping goal to the Helper because the stimuli were generally too complex for them to track.

Experiments 2.2.1-3: Discussion

The aim of the research presented in this chapter was to investigate whether infants understand helping as an action whose goal is to lower another agent's action costs (H-NUC). We hypothesized that if infants use a naive utility calculus to make sense of others' goal-directed actions, they may also apply its principles to interpret helping behaviors and ascribe a second-order goal to a Helper.

The results from Experiments 2.2.1-3, taken together, do not support this hypothesis. Infants' looking behavior in Experiment 2.2.1 tentatively suggested that they may have found it less consistent with a Helper's previous behavior when he did not minimize the Helpee's action cost, but that this effect was masked by an additional order effect. However, Experiment 2.2.2 provided evidence for the null hypothesis: Infants did not distinguish between an event where the Helper performed a utility-increasing action and a similar-looking non-helpful action. Finally, Experiment 2.2.3 demonstrated that infants familiarized to the same stimuli as those in Experiment 2.2.2 successfully ascribed a non-social, instrumental goal to the Helpee in this scenario, ruling out the possibility that the stimuli or experimental procedure failed to elicit any kind of goal attribution.

There are different explanations for these results. One option is that infants possess H-NUC, but the stimuli or experimental design that we used were not well-suited to prompt participants to apply it. For instance, the cost differences the Helpee would incur as a result of being helped may have been too small to be salient for infants.

Another possibility is that infants at the age we tested struggled with the means-ends-reasoning required by our task, i.e., understanding that moving an obstacle freed a relatively shorter path for another agent. They may therefore have attributed only first-order goals to both agents (Helpee: reach goal object, Helper: move a door/block), towards which the agents behaved efficiently (save for the Helpee in the inconsistent trial of Experiment 2.2.3). Ascribing a hierarchical means-ends structure, where subgoals are merely performed in the service of facilitating an ultimate goal, may be especially difficult in a social context, as infants have to override a potential prior assumption that agents tend to perform actions to acquire personal benefits. However, even for non-social action contexts, it has not been directly tested whether infants understand that means or subgoals can serve the sole purpose of making the overall action sequence less costly. In Appendix B, we report an incomplete experiment with which we sought to address this question; data collection was terminated after 10 subjects (BF: 0.44).

A further possibility is that at 12 months of age, infants cannot understand helping at all. However, this conclusion would be at odds with the large body of literature suggesting that even much younger infants prefer helpful agents (but see Lucca et al., 2024), and that this preference depends on the intentions infants ascribe to them (Woo et al., 2023).

Finally, another option is that initially, infants conceive of helping in a different way, as we discussed at length in the introduction to this section. If, for instance, young infants possess a concept of helping as enabling or as joint action, they could not have succeeded in our experiments. With a concept of helping as enabling, an observer would only consider an action helping if it allowed the Helpee to reach a previously inaccessible goal. In our stimuli, the Helper did not perform an enabling action in any of the familiarization events; in the test events of Experiment 2.2.1, both of the Helper's actions were enabling; and at test in Experiment 2.2.2, neither was. Therefore, a possessor of an enabling concept would not have set up a representation of helping during familiarization, and would not have distinguished between test events. With a concept of helping as joint action, on the other hand, an observer might rely on the presence of particular social interaction cues, which help establish an interpretive prior that the agents in the scene are participating in a collaborative endeavor. Our stimuli were largely void of such cues: The agents did not engage in communication or eye contact, were not in close physical proximity to one another while pursuing their respective goals, and did not act on the same objects. A possessor of a joint action concept may thus not have related the behaviors of Helper and Helpee, and accordingly not have established a shared goal for them.

In conclusion, our results don't support the hypothesis that infants have, or recruit, H-NUC when observing a helping event. It is therefore still an open question how infants understand helping actions, and how this understanding emerges in ontogeny. Follow-up research will have to uncover when, and how, young children come to acquire H-NUC, what its predecessor—if infants initially rely on a simpler concept of helping—looks like, and how the former develops from the latter.

Chapter 2.3: What do preschoolers mean by "helping"?

Experiment 2.3

The experiments from Chapter 2.2 investigated whether preverbal infants can engage H-NUC, and found no support for this hypothesis. The study described in the following builds on this result and explores the question with three-year-old preschoolers. At this age, children understand and use the term "help"[7], which implies that they possess some concept of helping. We exploited this fact to adapt the experimental stimuli from Chapter 2.2 to be used in a design with verbal prompts.

Just like in the infant experiments, we depicted a helping interaction in such a way that children could only succeed in the task if they possess H-NUC, rather than, for example, a concept of helping as enabling or joint action. Instead of recording children's looking behavior as a proxy for how much an observed outcome was consistent with their goal representation, we devised two different tasks to probe their intuitions. In one, children were asked to themselves help a goal-directed agent (the Helpee): To do so, they could choose from two different options for intervening on the Helpee's constraints, one of which enabled a relatively less costly action for the Helpee. In the other task, children observed two novel agents perform superficially similar-looking actions, only one of which reduced the Helpee's action cost, and were then asked to identify which one helped.

The aim of the present study was thus to assess whether preschoolers recruit H-NUC when prompted with the term "help". If they do, this would establish that by the age of three, children acquire the mature understanding of this behavior and can perform the required complex cognitive operations. If, on the other hand, they fail at this task, this could suggest that there may be distinct limitations to children's NUC (particularly when reasoning about complex actions with a second-order social goal), or that the representation of helping they have in mind differs in specific ways from that of adults.

The naive utility calculus, which we argued in Section 1 to be a good model of how infants make sense of others' goal-directed actions, also approximates social reasoning in older children (Jara-Ettinger et al., 2016). For instance, from observing an agent who displays the seemingly inconsistent behavior of choosing a cookie over a cracker when both are equidistant, but reaching for a cracker instead of a cookie when the latter is further away, 5-year-olds understood that the agent prefers cookies (and only picked the cracker in the latter scenario because the effort of reaching the cookie was not worth it). Further, children can infer, for example, agents' cost functions, skills, and epistemic states (Jara-

[7] See e.g. http://wordbank.stanford.edu/data?name=uni_lemmas (Meaning: HELP).

Ettinger, Tenenbaum, et al., 2015) on the basis of the assumption that agents make utility-maximizing choices.

Conflicting evidence was found by Gönül and Paulus (2021) in a study on 3-6-year-old children's rational action prediction. Children were shown images of an agent who wanted to move towards a goal and could do so on one of two sides of a bifurcated path, one of which was substantially longer than the other, and were asked to indicate which path the agent would choose. While 4-6-year-olds predicted that the agent would take the shorter path (i.e., the less costly option), 3-year-olds' responses were at chance. Further, 3-year-olds showed similar levels of performance when the agent could choose not among different paths leading to the same goal object, but among paths of different lengths each leading to two separate identical-looking goal objects. This finding may point to limitations in younger preschoolers' explicit reasoning about action efficiency when they are asked generate behavior predictions.

Research on children's understanding of third-party helping interactions, like that with infants, has focused mainly on what follows for children from observed instances of helping. That is, a helpful act is often used as a paradigmatic example of prosocial behavior, and researchers' main interest is in children's moral judgments. For example, it has been studied whether and under what circumstances children prefer or positively evaluate a helpful character (Franchin et al., 2019; Kenward & Dahl, 2011; Kishimoto et al., 2018; Li & Tomasello, 2018; Vaish et al., 2010; Van De Vondervoort & Hamlin, 2017), what role an agent's intentions play for such evaluations (Van de Vondervoort & Hamlin, 2018), or under what circumstances children deem helping to be obligatory or desirable (Dahl et al., 2020; Eisenberg-Berg & Neal, 1981; Hepach et al., 2012, 2016; Killen & Turiel, 1998; S. Kim et al., 2014; Sierksma et al., 2014). Overall, the studies corroborate and extend findings with infants that a sophisticated concept of helping is in place and used for social reasoning early on.

Many of the studies probing children's reasoning about helping interactions use vignette stimuli that make explicit reference to one agent helping another (Dahl et al., 2020; de Cooke, 1992; Eisenberg-Berg & Neal, 1981; Killen & Turiel, 1998; Nucci et al., 2017; Sierksma et al., 2014, 2014; Weller & Hansen Lagattuta, 2013), or experimenter prompts that include the term "help" (Paulus & Moore, 2011). Because these studies employ explicit linguistic primes to label the interactions, they fall short of telling us whether naive observers would categorize such behaviors as helping.

A few studies to date seem to suggest that preschool children and even toddlers can recruit utility reasoning when interpreting third-party helping situations. In a study by Jara-Ettinger and colleagues (Jara-Ettinger, Tenenbaum, et al., 2015), two-year-olds used the anticipated cost agents would incur in assisting someone with a task (activating a toy) to evaluate the agents' prosociality: They judged a competent agent who refused to help as less nice than an incompetent one, presumably because they accepted the relatively higher effort the latter would have to invest as a reasonable excuse (cf. Sierksma et al., 2014). Jara-Ettinger et al. (2020) found that 4-5-year-old children expected Helpees to take Helper costs into consideration: Here, children concluded that when a Helpee ambiguously requested aid from two potential Helpers, she was likely addressing the one for whom it would be less effortful to help. However, when that agent was unable to help, children thought that the Helpee was addressing the other agent (cf. Paulus & Moore, 2011)

While these studies suggest that children consider *Helpers'* action costs and expect them to be minimized, they do not directly give insight into whether children take the goal of helping to be a utility increase for the *Helpee*. Tentative evidence for this is provided by some recent experiments. Bridgers and colleagues (2020) asked preschoolers to judge the relative helpfulness of two Helpers, where helpfulness was based on different counterfactual outcomes. Here, 3-5-year-olds judged that a Helpee would thank the Helper who prevented a relatively worse event from happening over another Helper who prevented a harmless event. Therefore, this "better" Helper—although performing a similar first-order action as the other Helper—averted a larger counterfactual loss for the Helpee and was thus considered more praiseworthy by the participants.

Relatedly, in a study by Bennett-Pierre, Asaba and Gweon (2018), 3-5-year-olds could direct a puppet to help one of two other puppets, one of whom had to build a complex, the other one a simple block tower. Children selected the former, thus appreciating that the potential Helpee with the more difficult task would stand to benefit more from receiving assistance. In this design, the motivation to help was taken as a given, and children, helping by third-party puppet proxy, had to identify the recipient who faced higher action costs.

Finally, Woo and colleagues (2024) found that 16-month-old toddlers preferred a Helper who assisted one of two potential recipients facing a more costly task. In this study, one potential Helpee could be inferred to be struggling more than another in pushing a boulder up a hill, either because the hill was steeper or the agent was presented as weaker. This suggests that toddlers take relative need (i.e., the costs agents would have to invest to reach their goals) into consideration when engaging in social evaluation.

These studies are consistent with the possibility that children take the goal of helping to be enabling or bringing about a joint outcome. A need for help might not mean facing a relatively high cost to reach one's goal, but a high probability that without assistance, one would fail to do so (cf. Köster et al., 2016). Thus, for instance, children in the studies by Bennett-Pierre et al. (2018) and Woo et al. (2024) may have prioritized a Helpee who they believed was less likely to reach her goal alone. The experiments leave open the question whether children take the goal of helping to be intervening in such a way that the Helpee's utility is maximized.

The aim of the present study was to address this question and investigate whether preschool-aged children possess H-NUC. We implemented our question in two contexts: (1) where children themselves were asked to help someone, and (2) where they had to interpret the goal of another (Helper) agent from a third-party perspective. Children encountered an animated agent whose goal it was to approach and collect tokens of a target object. In (1), at test the agent was prevented from reaching her goal by obstacles blocking the way. Children could help by moving one of two obstacles aside, one of which would allow the agent to reach her goal on a relatively shorter path, or to reach the closer one of two goal objects. In (2), children saw the same agent in an environment with two other agents: One of them removed an obstacle that blocked the Helpee's shorter path to her goal (i.e., helped), the other moved either an unrelated obstacle (an action that was irrelevant for the Helpee) or moved an obstacle such that the agent's action cost actually increased (i.e., hindered). Children were then asked "Which one helped?" In both tasks, we considered children's response "correct" if it indicated an action that increased, or maximized, the utility of the Helpee.

In these tasks, participants could only answer correctly if they take helping to mean H-NUC. If, on the contrary, they have a concept of helping as enabling, children should be at chance in both tasks: In (1), both potential helping interventions enable the Helpee to reach her goal; and in (2), neither of the agents allows the Helpee to reach a previously unattainable goal. With a concept of helping as joint action, children may not categorize the scenario as helping, since it lacks diagnostic social interaction cues.

It should be noted that in (1), both actions constitute a utility increase for the Helpee (by way of enabling), but only one results in utility maximization for the Helpee (i.e., the shortest path made available). Thus, with a version of H-NUC requiring only utility increase, but not maximization, for the Helpee, both options in (1) may be considered helping. In (2), the helping action constitutes both a utility increase and utility maximization for the Helpee, while the action performed by the foil agent is neither (i.e., it either doesn't affect the Helpee or results in utility decrease). Therefore, in (2), any kind of utility-based helping concept should pick out the former.

Methods

The experiment was preregistered at the OSF (https://osf.io/v4yeg).

Participants

Our sample consisted of 64 children (age range: 3;0 to 4;0; mean age: 42.8 months). An additional 16 children participated in the experiment but were excluded for showing a side bias by indicating the same side (left or right) across all four trials (n = 10), failing to provide a valid response in at least one trial per block (n = 5), or technical failure (n = 1). Participants were recruited via Hungarian-language advertisements on social media. Informed consent was obtained from both children and caregivers before the experiment. The study received full ethical approval from the United Ethical Review Committee for Research in Psychology (EPKEB) in Hungary.

Apparatus

The experiment was conducted remotely via video chat (Zoom), with stimuli presented on the participants' web browser via Slides.com. The Zoom call was recorded for later off-line coding of responses. Video stimuli were created using the Blender 3D animation software.

Procedure & stimuli

The experiment consisted of 3 phases: a warm-up task, and two experimental blocks. All stimuli, including the testing script, can be accessed here: https://osf.io/ts84j.

Warm-up task. In the warm-up task, children were shown images of animals with colors matching the ones in the experimental stimuli, and were asked to identify them by pointing (e.g., the experimenter asked "Can you show me where the horse is?") and to label the color of the animal. This also provided a way to disambiguate children's responses in the later experimental task. Children were then introduced, with the help of an image, to a character called a "kobo" (a green circle with eyes) who likes collecting apples, but to do so, he sometimes has to go around rocks lying in the environment.

Experimental blocks. Each experimental block contained a familiarization followed by two test trials. In the first experimental block (Block 1) we probed whether children would help an animated agent in such a way that the agent's action cost is minimized. At the beginning of the block, a familiarization video was played. This video contained the house of the kobo in the top center, and vertical walls on the left and right side of the scene. The walls were each interrupted by a short gap approximately at the height of the house. The top of the walls began at the boundaries of the display, while on the bottom, there was some space, such that the walls could be passed either through the gap or by moving around them at the bottom. In the first part of the familiarization video, the kobo exited his house, and an apple appeared in the scene behind the gap in one of the walls. The kobo looked at the apple, then approached it through the gap, paused when making contact with the apple, and carried it back into his house. The second part of the familiarization video started the same way, but just before the kobo reached the gap in the wall to approach the apple, a grey rock appeared in front of the opening. The kobo bumped against the rock twice as if attempting to move it, then detoured around the unobstructed, long end of the wall on the bottom of the screen, thus taking a longer path than before. He again took the apple back into his house along the same longer path.

After this familiarization, children were presented with two test trials. In each of the trials, children saw the kobo in a new layout with a single horizontal wall placed directly below it. This wall was interrupted by two gaps, one of which being closer to the location of the kobo. In one of the trials ("one goal"), a single apple appeared behind the closer gap. The kobo looked at it, but before he started moving to approach it, two rocks in different colors (orange and white) appeared in front of the gaps. At this point the experimenter told the child that she could help the kobo by moving aside one of the rocks, and asked which one should be moved. After the child responded, the experimenter played a video clip where the rock indicated by the child moved aside, and the kobo went to the apple.

In the other test trial ("two goals"), *two* apples appeared, one behind each of the gaps respectively. Again, after the kobo looked at both of them in turn, two rocks (in different colors: black and purple) appeared in front of the gaps. The child was asked which of them should be moved, and after she responded, the experimenter played the corresponding video clip where the rock moved aside and the kobo approached the apple.

The second experimental block (Block 2) also contained two test trials. Here we tested whether children could correctly identify which of two agents helped, that is, who acted in a way that reduced the Helpee's action cost. The test trials in this block were preceded by the same familiarization video that was used in Block 1. In the first test trial ("help vs. irrelevant"), the video's layout was identical to the one in the preceding familiarization clip. Here, two (grey) rocks were already present at the onset of the video, each blocking a gap in the wall. In this video, after the green kobo exited his house, two other kobos (pink and yellow) entered from the bottom of the screen. An apple appeared on one side of the screen, and all kobos looked at it (first the green kobo, then the other two simultaneously). After this, the pink and yellow kobo each moved towards one of the rocks and pulled it downward such that the gaps in the walls were now unobstructed, at which point the video ended. The video was repeated a second time, and the experimenter then asked the child which kobo helped. In this trial, one of the new kobos reduced the cost of the green kobo by removing an obstacle which freed a relatively shorter path to the apple. The other agent performed an action that looked similar (i.e. removed a rock covering a gap in the wall) but was irrelevant for the green kobo: It did not affect her utility, as there was no apple behind this gap.

The second test trial ("help vs. hinder") had a similar structure, but the layout of the scene was slightly different. In contrast to the first test trial, the two vertical walls extended to the top and bottom boundaries of the screen but were interrupted by two gaps, respectively. Here, the wall could only be crossed by passing through one of the gaps; there was no way to detour it. On one side, a rock covered the bottom gap, on the other side, a rock covered the top gap. In this video, two (rather than one) apples appeared, one behind each of the top gaps in the walls. When the novel kobos (blue and red) moved the rocks aside, one of them moved the rock covering the top gap so that it now covered the bottom gap, while the other moved the rock from the bottom to the top gap. This video was also repeated a second time and the child was then asked by the experimenter which kobo helped. In this trial, the helping action performed by one agent was contrasted with a hindering one: Although both novel kobos moved a rock from one gap to another, one agent made a shorter path to an apple passable, while the other agent obstructed the shortest path to the other apple.

Note that in Block 2 (in contrast to Block 1), children did not see the continuation of the video after the rocks were moved, and thus did not see the consequence of the novel kobos' actions on the behavior of the green kobo (i.e., which path he took, or to what extent that actual path was shorter than the counterfactual ones that would have been available to him without the novel kobos' intervention).

The key events in all videos were accompanied by different sound cues.

We counterbalanced the following factors in the stimuli: (1) the colors of the rocks in Block 1 and of the agents in Block 2; (2) the order of trials in Block 1 ("one goal" first vs. "two goals" first); (3) the location of the apple in the familiarization video (left vs. right); (4) the location of the correct response, relative to the location of the apple in the preceding familiarization video, in the first and second test trial in a block (same-same, same-different, different-same, different-different). This resulted in 32 sets of stimuli.

If in any of the four trials a child did not respond immediately, responded with "both", or indicated the green kobo, the experimenter repeated the question and encouraged the child to "pick one".

We excluded from analyses trials (a) on which a child did not provide a codable response (e.g., by responding "none", not responding, responding with "both", or referring to the protagonist, even after being prompted repeatedly to identify one option), (b) where a child's verbal response with a color label did not unambiguously pick out one of the two options, and (c) where their color labels from the warm-up task did not help disambiguate their response, and the child did not subsequently provide a response by pointing. One trial from Block 1 and seven trials from Block 2 were excluded. Finally, we excluded all participants who didn't provide a codeable response in at least two trials, one from each block; and who displayed a side bias by indicating the same side (left or right) across all four trials.

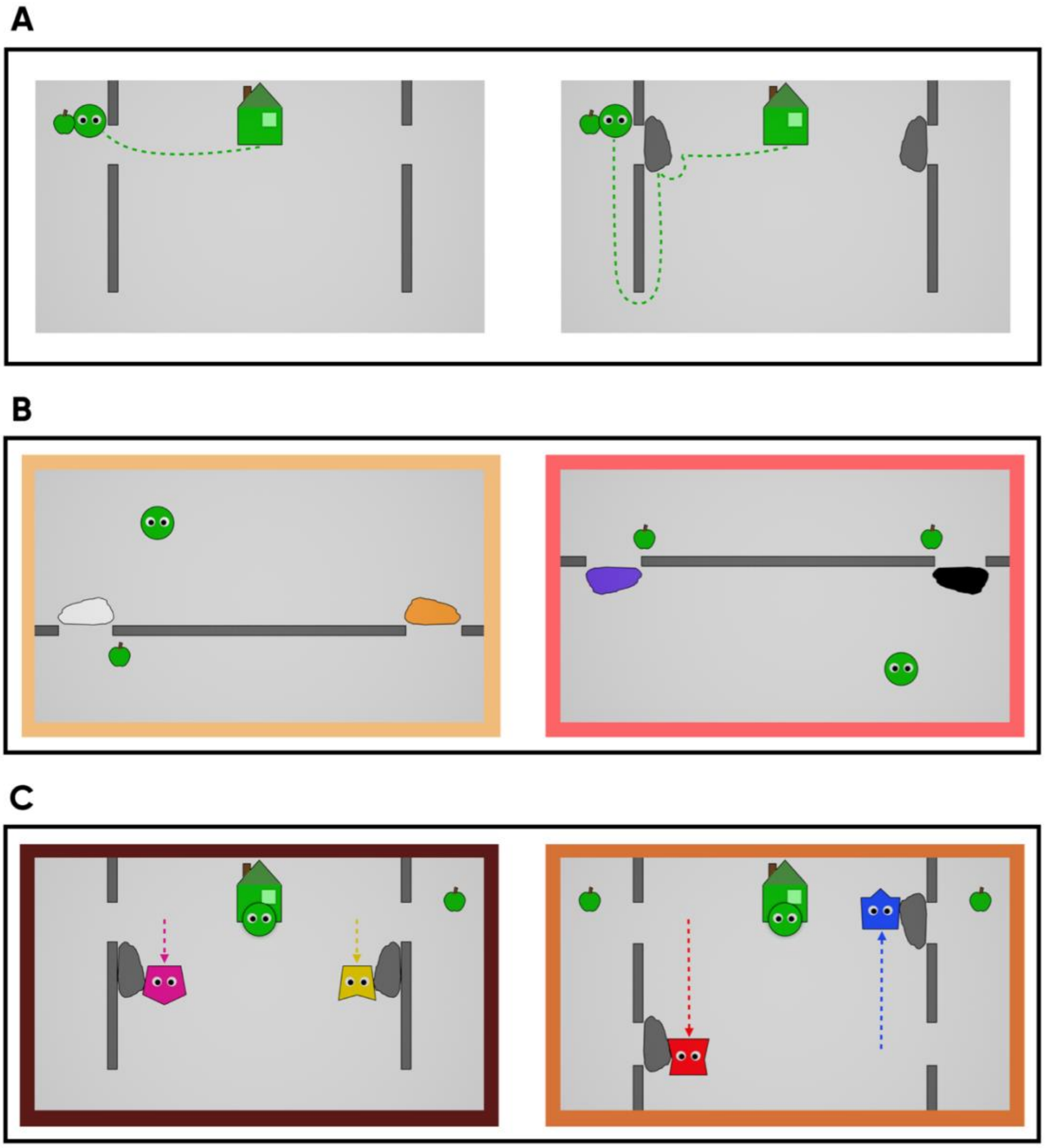

Figure 2.11. Stimuli used in Experiment 2.3. In the Familiarization video (A), the agent either approached the goal on the most direct path through a gap in the wall (left), or, when a rock obstructed this path, first attempted to move the obstacle, and then detoured around the bottom of the wall (right). In the Block 1 test events (B), children were asked to help by moving aside one of two rocks, blocking the paths to a single goal (left) or two possible goals (right). In the Block 2 test events (C), an agent who helped by allowing the green agent to take a shortcut (left: yellow, right: red) was contrasted with one who performed an action that did not affect the green agent's action plan (left: pink; "help vs. irrelevant" trial), or one who hindered by blocking the shortest path (right: blue; "help vs. hinder" trial).

Analysis

Our main question was whether children would above chance (1) choose to move an obstacle which lets the protagonist take a relatively shorter path to a goal (Block 1 test trials), and (2) respond that another agent helped who removed an obstacle which would let the protagonist take a shorter path to a goal (Block 2 test trials). The dependent variable was thus children's choice of the "H-NUC - help" option in each of the four trials (coded as the "correct" response).

We analyzed the data using a Bayesian logistic regression. As predictors, we included the trial (1-4) and subject. This allowed us to estimate (1) whether children responded more accurately on some trials than others, and (2) whether there was an effect of block. Further, we included a trial by order interaction for Block 1, to assess whether the order in which the trials in Block 1 were presented matters. Because the Bayesian model performs parameter estimation (and not hypothesis testing), we can characterize the entire posterior probability distribution for each of these estimates. We report the 89% credible intervals (CI) as well as the means of the parameter estimates for responding correctly (where 0 is incorrect, 1 is correct, and 0.5 is the chance level), or for the difference between blocks (where chance is excluded if the 89% CI does not include 0). These CIs specify that the true parameter value lies in this interval with 89% probability.

We conducted an exploratory analysis to determine whether the counterbalanced factors affected children's responses. A further exploratory analysis assessed whether children's rate of correct responses in Block 1 predicted their responses in Block 2.

Results

Hypothesis-driven results

In Block 1, 58.7% of responses in the "one goal" trial were correct, as were 45.3% of responses in the "two goals" trial. In Block 2, 61.4 % of responses in the "help vs. irrelevant" trial and 64.1% of responses in the "help vs. hinder" trial were correct.

We found that children were more accurate in Block 2 than Block 1. The parameter estimate for the difference between responses in the two blocks excluded the chance level (mean = 0.1, 89% CI: [0.0151, 0.187]).

In Block 1, the parameter estimates did not exclude chance, although for the "one goal" trial, the majority of estimates were above the chance value ("one goal" trial: mean = 0.597, 89% CI: [0.474, 0.724]; "two goals" trial: mean = 0.449, 89% CI: [0.325, 0.568]). In Block 2, the parameter estimates excluded chance, although for the "help vs. irrelevant" trial, this was just barely the case ("help vs. irrelevant" trial: mean = 0.621, 89% CI: [0.505, 0.74]; "help vs. hinder" trial: mean = 0.656, 89% CI: [0.538, 0.78]).

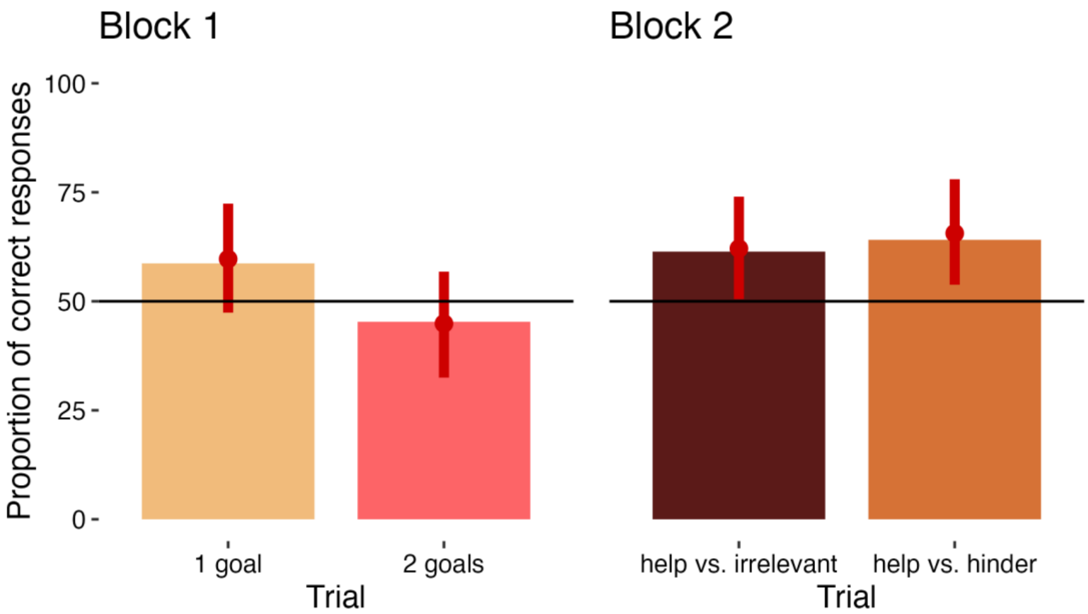

Figure 2.12. Bar plots representing the proportion of correct responses in Experiment 2.3. Red error bars indicate the 89% CI of the parameter estimates, with dots indicating the means.

Additional results

We assessed whether there was an order effect in Block 1, where the order of trials was counterbalanced (2 goals first vs. 1 goal first) and found no interaction of Trial and Order (mean = 0.039, 89% CI: [-0.058, 0.169]). Children's accuracy here was thus not affected by whether a trial came first or second in the sequence.

We also checked whether children's responses in Block 1 predicted those in Block 2, to see whether children who were more likely to be correct in one would also be better in the other. The correlation of children's responses in the two blocks did not exclude the chance level (mean = 0.166, 89% CI: [-0.488, 0.72]).

We recorded whether children in Block 2, as their first response to the question which agent helped, replied "both". In the "help vs. irrelevant" trial, 33 of 64 from the initial responses, and in the "help vs. hinder" trial, 4 of 64 of the responses were "both".

Finally, we looked into whether side congruency between familiarization and test (i.e., the location of the apple in a familiarization trial was the same as the side of the correct response in the immediately following test trial) biased children's responses. Children were somewhat more correct in side-congruent trials (66.12%) than in side-incongruent trials (50%), though this was not significant in a chi-squared test (χ^2 (1) = 2.578, p = .108). This effect is more pronounced in Block 1 (accuracy in congruent trials: 65.62%, in incongruent trials: 45.16%) than in Block 2 (accuracy in congruent trials: 66.67%, in incongruent trials: 55.56%)[8].

Discussion

The aim of the present study was to investigate how young children interpret the term "help", and whether they can leverage reasoning about agents' costs and benefits to ascribe this goal to others. To do so, we assessed whether preschoolers (1) would help in a way that maximized the utility of the agent by allowing him to take the shortest path available to the goal, and (2) would correctly identify the agent who increased the Helpee's utility in this way to have performed a helping action.

We found that children succeeded in task (2) but did not in task (1): When asked to help another agent, 3-year-olds in our study did not consistently do so in a way that minimized the Helpee's action cost, though they were somewhat above chance when there was only one goal available. When asked to identify who helped, they tended to select the agent who increased the Helpee's utility, especially when the helping agent was contrasted with one who hindered; in the trial where the non-helping agent performed an irrelevant action, children's responses were just barely above chance.

The different rates of correct responses in the two blocks may reveal features of preschoolers' concept of helping. Specifically, one possibility is that children consider any action helping which (potentially) contributes to a Helpee's goal-directed action by increasing her utility compared to an unassisted baseline state. In our experiment, each of the two possible interventions in the trials in Block 1 was helpful in this sense: Without aid, the kobo could not have passed the obstacles and reached her goal, thus even removing a rock that required the kobo to take a relatively longer path increased her utility. In contrast, in Block 2, only the Helpers', but not the other agents', actions lowered the Helpee's action cost and allowed him to take a shorter path than he could have on his own. Thus, under such a concept of helping, either of the actions in Block 1 can be considered helping, while in each trial of Block 2 only one of the agents helped.

A different version of H-NUC requires a Helper to *maximize* the Helpee's utility. Such a goal can be derived from a utility function of the Helper whose reward term is the Helpee's utility: By bringing about the best outcome for the Helpee, given fixed action costs for herself, the Helper also maximizes her own utility. It is possible that for children (and maybe adults as well), this construct contains a normative standard of helping *well* and the ways a Helper *ought to* act. They may still call an action helping which merely increased the Helpee's utility but was not as helpful as it could have been, but evaluate actions more positively in which a Helper generates relatively more utility. For instance,

[8] This is surprising, as the layout in the first test trial of Block 2 was almost identical to the familiarization layout, but was different in Block 1 (see figure 2.11).

in Bennett-Pierre et al. (2018), children were asked which Helpee *should* be helped, and Woo et al. (2022) measured which Helper children *preferred*. In the task we posed children in Block 1, either possible intervention contributed to the Helpee's goal, so when prompted to help, they may have selected the first action that satisfied this criterion and, in the absence of a reason to do so, may not have considered the alternative option.

Relatedly, we found that children responded most accurately in the trial where helping was contrasted with hindering. This suggests that they may be most certain that an action which increases an agent's action cost is *not* helping. Tentatively bolstering this interpretation is the fact that in the "help vs. irrelevant" trial, many more children first responded that both agents helped than in the "help vs. hinder" trial. (Though note that this could be due to the order of trials, as the experimenter always prompted children to pick just one, and the "help vs. hinder" trial was always presented second.) In this context it may be relevant that most studies researching infants' and young children's understanding of helping and preference of Helpers contrast this paradigmatically "prosocial" action with an "antisocial" one (i.e., hindering), which could contribute to highlighting the valence of the former (though see Hamlin et al., 2007 and Chae & Song, 2018, where 6- and 10-month-olds preferred a helpful over a neutral character). In a recent study, Wong and colleagues (2023) showed that children who were told about an agent acting prosocially, and subsequently generated selfish counterfactual alternatives for how the agent could have behaved, were more likely to positively evaluate the agent. This highlights the potential effects of thinking about action options with maximally distant (social) effects.

The pattern of results we obtained may also have been affected by the specific ways we implemented our research questions. For instance, there could be an effect of trial order: Block 2, in which children were asked to point out who helped, always came second, and the "help vs. hinder" trial, where children responded most accurately, was always presented last[9]. It is possible that children could have learned something over the course of the experiment, for example by observing the consequences of their own helping actions on the agent's action cost in Block 1. It should be noted, however, that they never received explicit feedback on their responses, and that there was no effect of trial order in Block 1, where this was counterbalanced.

Further, the differences in responses between blocks could be due to children having difficulties helping in a first-person task compared to reasoning in a third-person context. Children were asked to assist an animated agent in an animated grid-world, where the

[9] This procedure was fixed after piloting different versions of the experiment. During piloting, (1) children's responses in the "who helped" trials were better when they appeared in Block 2, but the accuracy in the "how to help" trials did not differ by whether they were in Block 1 or 2; (2) children were least accurate in the "help vs. hinder" trials; and (3) the order of the "how to help" trials seemed to matter (such that children improved from the first to the second trial).

mechanism by which they "moved" the obstacle aside remained opaque. In a more realistic context, helping a partner with a familiar task, impeded by known constraints and costs, children may have performed better. Finally, because children were not themselves represented by a character in the grid-world, it may be that some of them imagined themselves to be located not at the center of the screen, but instead on the side that was furthest away from the agent. In this case they may have attempted to minimize *their own* action cost by moving aside an obstacle that was closer to where they imagined themselves, rather than solely focusing on the Helpee's cost.

It should be noted here that 3-year-olds' failure in Block 1 echoes that found by Gönül and Paulus (2021), who interpreted their results as evidence of limitations in young preschoolers' reasoning about rational behavior. In light of the large body of literature demonstrating robust expectations of efficiency in infancy, reviewed in Section 1, it is plausible that task characteristics here masked children's competency. Perhaps using path length as a proxy for cost is not ideally suited to probe their intuitions, or they may struggle with generating explicit responses, and implicit measures may be more appropriate. A potential source of bias in our study was the location of goal rewards during familiarization, as children seemed somewhat driven to indicate the same side in a subsequent test trial. They may have assumed the location carried some relevance, and that helping constituted in assisting the agent reach that particular side. Prior studies have found that frequency information can affect children's behavior in goal prediction tasks (Gönül et al., 2023), and though we only showed a single familiarization trial, it is possible that a similar process affected the responses.

Taken together, however, the results provide support for our hypothesis that preschool-aged children have access to H-NUC, as they could otherwise not have responded as predicted in Block 2. They could only have identified who helped by realizing who of the agents contributed to the Helpee's goal-directed action by increasing her utility.

General discussion

To make sense of the world around them and learn from and about social partners, young children not only have to be able to comprehend behaviors that proximately yield rewards for an agent herself, but also those that target, or affect, the welfare of others. Among actions with social goals, helping is one that likely occurs frequently in the child's environment, as people often perform tasks collaboratively.

Developmental research on the ontogenetic origins of social and (proto-) moral cognition has amassed evidence for the idea that even young infants recognize helping interactions from a third-party perspective, and draw inferences about, as well as positively evaluate, agents who help. It is implied that the participants in these studies understand observed helping actions similarly to the adult experimenters. In this section, we have argued that this has not been conclusively demonstrated, and that some fundamental questions about children's and infants' understanding of helping actions remain open.

First, we pointed out that much of the research in this domain relies on the manual choice measure introduced by Hamlin et al. (2007). Here, infants watch representations of interactions between agents, some of whom help another character, some of whom do not, and are then prompted to reach for one of the figures. As there have been debates on the robustness of this measure (and the phenomenon of infant sociomoral evaluation it is meant to reflect), we conducted a replication of the original experiment by Hamlin and colleagues, which did not yield the same result (Chapter 2.1). This highlights firstly that further research is called for to clarify under which circumstances infants' preference for Helpers can be elicited (Lucca et al., 2024). Secondly, it points to the importance of designing tasks that rely on other measures to probe the early helping concept. While Hamlin's paradigm has been undoubtedly pivotal for probing infants' sociomoral evaluation across domains (Geraci & Surian, 2011; Hamlin et al., 2011; Kanakogi et al., 2013; Thomas et al., 2018), it appears less suited to examine issues surrounding the rationality of helping (i.e., the range of utility distributions that are compatible with its ascription) and the architecture of its representation (i.e., the way in which the two agents' goals are teleologically related, whether in a nested or means–end structure).

Here we have argued that a mature understanding of helping is grounded in a naive utility calculus. In this framework, the goal of helping is to increase the utility of another goal-directed agent by intervening on her action constraints. We designed a set of experiments (reported in Chapters 2.2 and 2.3) to test whether preverbal infants and preschoolers recruit such an understanding upon observing helping scenarios. These scenarios featured abstract agents and minimal cues of social interaction, so to construe them as helping, participants had to leverage utility reasoning. We found that one-year-old infants, while they successfully ascribed a first-order individual goal (i.e., to reach an object) to the Helpee in the scenario and expected him to act efficiently, did not seem to interpret the

Helper's behavior in terms of a second-order social goal. Three-year-olds, on the other hand, applied a hierarchical, utility-based concept when identifying an agent who helped. Surprisingly, when asked themselves to help, they did not choose the utility-maximizing option above chance. This may be due to methodological idiosyncrasies, or because preschoolers take any action to be helpful that leaves the Helpee better off than he would be without assistance (i.e., utility increase).

Future research will have to probe further at what age—and with what precise features—H-NUC emerges in development, and what, if any, its precursors are. In the introduction, we proposed two conceptual alternatives that are simpler and thus may be more accessible to young infants: helping as an enabling or as joint action. Helping as enabling, where the Helper's goal is to make goal completion possible for the Helpee, requires infants to nest the goal of the individual agents into one another but does not necessitate a comparison of graded utility differences. In contrast, helping as joint action, where the Helper's goal is to work with the Helpee toward a shared outcome, requires infants to compute aggregate utility functions but not to embed the agents' goals into one another. From these two conceptual precursors, a mature concept of helping can emerge when infants acquire the respective components: the capacity to generate graded utility comparisons (in the enabling case) or to embed individual goals into one another (in the joint-action case). Our results from Chapter 2.2 are consistent with 12-month-old infants possessing either of these candidates, but cannot disambiguate which one.

To adjudicate, future research ought to test empirical predictions that can be derived from them. Helping as enabling would allow infants to identify only those helping situations that involve the Helpee's (imminent) failure. Moreover, with this concept, they would not make possible fine-grained assessments of how much an agent helped, so long as the Helpee's goal was reached. On the other hand, with a concept of helping as joint action, infants should not be able to distinguish between instances of helping proper and other collaborative interactions. For instance, they may represent the outcome of helping events as a shared goal ("A or B obtain the object in the box"), rather than as the proprietary goal of the Helpee ("A obtains the object in the box"), and thus should not be surprised upon observing the Helper completing the goal that previously the Helpee was pursuing. Furthermore, under this account, infants' well-known preference for Helpers would be reframed as a broader affiliative stance toward collaborators—with the surprising prediction that infants would choose randomly if asked to choose between Helper and Helpee. Finally, unlike helping proper, the goal of helping as joint action is to bring about a shared outcome at the lowest aggregate costs attainable. This entails that, for instance, it may be perfectly rational for an agent to choose an action that increases her partner's cost if doing so allows the joint action to become co-efficient.

An open question we have thus far ignored concerns the understanding of hindering actions, which are often discussed as mirror opposites of helping. Under this interpretation, hindering is represented as the second-order social goal of lowering the utility of a Hinderee (i.e., as spite). An alternative option, however, is to represent hindering as the side-effect of conflicting first-order goals (e.g., two agents seeking the same object). The difference between these two construals cannot be overstated: In the former, lowering the Hinderee's utility is the goal of the intervening agent; in the latter, it is collateral. Notwithstanding these differences, tests aimed at telling these two interpretations apart are missing in the literature. Moreover, the adoption of the former construal is unlikely to be undergirded by the same assumptions that make the ascription of helping possible: While in the case of helping, an agent's costly investment in a partner's welfare can be justified by reference to potential future benefits obtained in a cooperative relationship, such an explanation would not be available for hindering. This makes the proposal that infants may default to interpreting antisocial behaviors in terms of spite rather than competition all the more question-begging.

Moreover, even when children have eventually mastered an adult-like concept of helping, further questions remain: For instance, how do they integrate different kinds of cost across agents (cf. Pomiechowska & Csibra, 2022)? Do they determine an optimal utility trade-off between the two agents' respective investments for helping to occur, or do they interpret any action that increases the utility of another agent as helping, even if the action is very effortful and brings about only a negligible cost reduction? Finally, a pivotal question for future research will be whether infants interpret the Helper's altruistic action as indicative of a partner-invariant individual trait (e.g., helpfulness) or of a specific relationship with the Helpee within which the costs of helping may be recouped (for instance, through reciprocation), which would not justify predictions of future prosocial behaviors directed to a novel partner.

Section 3:

Trait attribution and partner choice

Introduction

Information gleaned from third-party social interactions can be valuable to observers for multiple reasons. On the one hand, it can warrant inferences about agents' dispositions (e.g., liking bananas) or traits (e.g., being strong enough to push aside a boulder). On the other hand, it can reveal features of the agents' social relationships and the broad network they are embedded in (they could be friends, business partners, family members, and so on). The conclusions drawn from behavior observation can be used to guide observers' own social decision-making: For instance, one might choose to interact with a social partner who one surmises to possess a relevant trait.

Infants' early-emerging sensitivity to third-party helping interactions and their tendency to prefer Helpers over non-Helpers have been interpreted as reflecting an evolved psychological mechanism dedicated to recognizing and choosing advantageous cooperation partners (Kuhlmeier et al., 2014; Van de Vondervoort & Hamlin, 2016). Such a mechanism is hypothesized to have played a role in the evolution and sustainment of human cooperation (e.g. Baumard et al., 2013). According to this account, babies and young children would, upon observing an agent help another, ascribe a trait of being *a Helper* (or prosocial/nice/a "good guy") to the agent. They would subsequently favor the Helper because she is expected to continue acting helpfully even towards the infant. However, there is no direct evidence that trait attribution and partner choice underlies infants' preference for Helpers. A further problem for this account is that literature on children's trait reasoning has argued for limitations in this capacity until middle childhood (Heyman, 2009).

Research into the development of partner choice in young children has only recently begun to directly probe children's ability to identify and choose advantageous partners for cooperative interactions (Grueneisen et al., 2023; Hermes et al., 2016; J. W. Martin et al., 2022; Prétôt et al., 2020; A. M. Woodward et al., 2022). These studies suggest that young children can ascribe cooperation-relevant character traits from a third-party perspective, which in turn informs their partner-choice decisions. However, it is yet unknown whether they do so (a) when they have to infer traits from behavioral observation rather than verbal vignettes; (b) when they choose partners on the basis of these attributed traits without explicit prompts; and (c) when their partner selection has real consequences for their reward rather than being merely hypothetical.

In this section, we report a series of experiments which address these gaps. Using a novel tablet-based research game, we investigated how children and adults would choose partners when they had the opportunity to play the game together with their chosen partner. The structure of the game was such that the outcome for a participant depended partly on their partner's behavior. Participants first observed computer-animated agents from a third-party perspective. The behavior of these agents was guided by preset parameters that specified their traits: prosociality (here: helpfulness) and competence (here: speed). Then, participants selected one of the previously observed agents and played together with their chosen partner. We conducted a series of experiments with children (n = 436) and adults (n = 71) in both Central Europe (Hungary, Austria) and East Asia (Japan[10]). The aim was to investigate whether participants could identify differences between agents from observing their behavior without verbal prompts highlighting the concepts under investigation, whether they would prefer interaction partners who are more helpful and/or skilled, and whether they would systematically prioritize one trait over another in their partner choice.

Cooperation, partner choice, and trait attribution

The ubiquity of humans' altruistic and cooperative behaviors towards even genetically unrelated strangers long posed a puzzle for researchers studying behavior in an evolutionary framework: It is not immediately clear how actions that confer benefits to another at a cost to the agent could propagate. A variety of mechanisms have been proposed to explain this puzzle, including direct (Trivers, 1971) and indirect reciprocity (Alexander, 1987), and punishment (Boyd & Richerson, 1992).

In environments that provide opportunities for mutualistic collaboration (i.e., cooperators stand to gain from an interaction, either immediately or in the long run), it is beneficial for agents to cooperate. However, cooperation may nonetheless not emerge or persist, as it is vulnerable to exploitation from free-riders. Recently, theorists have emphasized that partner choice can be a viable mechanism for stabilizing cooperation (e.g. Barclay, 2016; Baumard et al., 2013; Bshary & Noë, 2003; Noë & Hammerstein, 1994). When there is a pool of potential collaborators with whom to enter into mutualistic cooperative interactions, agents can abandon cheaters and move on to a new partner. Bad cooperators are less likely to be chosen in the future and thus miss out on beneficial joint endeavors. This way, agents can be incentivized to cooperate and act fairly out of self-interest (Chiang, 2010; Sylwester & Roberts, 2010), without a need to enforce cooperation.

If partner choice played a crucial role for novel, more sophisticated forms of cooperation to emerge and persist in human evolution, present-day humans' social and moral cognition may have been shaped by the demands of partner choice dynamics (e.g. Baumard et al., 2013). According to this view, humans are equipped with evolved psychological mechanisms allowing them to successfully navigate social environments where choosing good partners confers benefits (J. Martin et al., 2019).

[10] Data collection in Japan was led by Kazuhide Hashiya and Hiromi Kobayashi from Kyushu University in Fukuoka.

One component of such a partner choice psychology is a motivation for potential partners to be seen by others as good cooperators, i.e., they should track and manage their reputation. The presence of mutualistic opportunities and many potential collaborators creates a "biological market" where individuals advertise themselves as high-value partners, for example via competitive displays of generosity (Barclay, 2013; Barclay & Willer, 2007; Roberts, 1998).

On the other hand, individuals tasked with choosing partners have an interest in recognizing and preferentially interacting with good collaborators. One such putative mechanism is concerned with drawing inferences about cooperation-relevant character traits from observing agents' behaviors in third-party interactions. The observer here has to distinguish between situational contributions to an outcome (such as luck) from agent-specific traits that are stable across variable circumstances and contribute causally to the agent's behavior (Delton & Robertson, 2012). Moreover, people should exercise epistemic vigilance in their social evaluation, to identify whether behaviors genuinely reflect traits, or are strategically deployed (Barclay & Willer, 2007; Heintz et al., 2016).

A wealth of research in social psychology has been dedicated to people's trait attribution (also under the header of "person perception" or "impression formation") (Ames et al., 2011; Gilbert & Malone, 1995; Trope & Gaunt, 2007). An ongoing adjacent debate in psychology and philosophy concerns the metaphysical reality of character traits (see C. B. Miller, 2016). However, this debate is orthogonal to whether people behave *as if* traits are real: Mental concepts, independent of their scientific validity, can be useful abstractions and feature in folk-psychological inference (Westra, 2018). In the context of partner choice, the focus is thus on whether, how, and why people attribute traits to others, and how they use such ascriptions for explaining and predicting behavior.

The value of a partner is generally thought to depend on the partner's *willingness* to confer benefits to the self (e.g. generosity, fairness, loyalty, or helpfulness) and *ability* to do so (e.g. skills, intelligence, or access to resources) (Barclay, 2013, 2016; Fiske et al., 2007). Research suggests that people attend to both of these domains and use them in partner choice decisions (Barclay & Willer, 2007; Bliege Bird & Power, 2015; Cottrell et al., 2007; Landy et al., 2016; Macfarlan & Lyle, 2015; Sylwester & Roberts, 2010). Some of these studies suggest that participants prioritize traits that indicate willingness to cooperate over those indicating ability to do so (Bliege Bird & Power, 2015; Cottrell et al., 2007; Delton & Robertson, 2012; Eisenbruch & Krasnow, 2022; Eisenbruch & Roney, 2017; Landy et al., 2016; Raihani & Barclay, 2016).

However, some cross-cultural investigations point to the possibility that trait attribution may not operate the same way across human populations, and that there may be influences of culture and socialization on the types of traits that matter in partner choice (Apicella et al., 2012; Macfarlan & Lyle, 2015; Smith & Apicella, 2020a, 2020b). For instance, research by Apicella and colleagues on the social structure and interactions of the Hadza, a group of hunter-gatherers in Tanzania, points to the possibility that WEIRD[11] concepts are not universally applicable. In several studies, they found that the Hadza did not unconditionally prefer to form ties with cooperative partners (Apicella et al., 2012), though this may be mediated by higher exposure to and knowledge of outside cultural institutions (including formal schooling or a job) (Smith & Apicella, 2020b). Moreover, the Hadza agreed less on ratings of group members' moral traits than ratings of others' hunting ability and hard work (Smith & Apicella, 2020a). Developmental research on whether young children prioritize either a partner's willingness or ability to provide benefits has the potential to offer important insights in this debate, and studying the trajectory of partner choice across childhood in different cultural contexts can clarify the influence of socialization on social preferences.

The development of trait attribution and partner choice in children

If some form of partner choice psychology is part of humans' evolved social cognition, it should emerge early in ontogeny. However, there are distinct bodies of research that have produced mixed evidence on whether the necessary skills are already possessed by young children. Specifically, it is unclear whether children draw conclusions from agents' previous behaviors and use them to predict how the agents will behave towards themselves.

On the one hand, research on young children's reasoning about others' behavior in terms of stable character traits has yielded mixed results. Some studies suggest that this ability is limited in early childhood and develops gradually over the elementary school years, with more robust trait reasoning emerging around 7-8 years of age (Alvarez et al., 2001; Boseovski & Lapan, 2021; Kalish, 2002; Kalish & Shiverick, 2004; Livesley & Bromley, 1973; Rholes & Ruble, 1984; for a review, see Heyman, 2009). Other research has emphasized that preschoolers and young school-age children may require multiple instances of consistent behavioral information to generate trait inferences or behavioral predictions (Aloise, 1993; Boseovski & Lee, 2006; Cain et al., 1997; Ferguson et al., 1984; Heller & Berndt, 1981). A study by Liu, Gelman and Wellman (2007) suggested that preschoolers succeed at inferring traits from observed behaviors, and generating behavioral predictions from trait labels, but struggle with behavior-to-behavior

[11] The acronym WEIRD stands for Western, Educated, Industrialized, Rich, and Democratic; it has been argued to describe the majority of human research subjects in academic psychology, although this population is not particularly representative for humanity (Henrich et al., 2010).

predictions (cf. Fitneva & Dunfield, 2010; but see Surian et al., 2018), which is precisely what is needed in partner choice.

On the other hand, recent findings of children's and infants' social selectivity after observing agents in interactions are taken as evidence that from an early age, children already can recognize and choose good cooperators. From the second year of life, young children preferentially engage with those who previously acted in a prosocial manner towards another person (for a review Kuhlmeier et al., 2014). For instance, in a study by Vaish and colleagues (Vaish et al., 2010), 3-year-olds withheld help from an adult who performed harmful actions toward a third party. 4.5-year-olds (Kenward & Dahl, 2011) and 3-year-olds (Olson & Spelke, 2008) favored an agent who previously helped and shared, respectively, when distributing a scarce resource. 25-month-olds helped (Surian & Franchin, 2017), 15-month-olds provided a resource to (Burns & Sommerville, 2014), and 13- to 17-month-olds accepted a toy from (Lucca et al., 2018; see also Tasimi & Wynn, 2016) an agent who distributed resources equally over one who distributed unequally. In making such social decisions, recent findings suggest that children also take agents' intentions into consideration (Jara-Ettinger, Tenenbaum, et al., 2015; J. W. Martin et al., 2022).

In even younger infants, research demonstrating infants' preferential reaching or looking towards a prosocial (e.g., a Helper) over an antisocial character (e.g., a Hinderer), as reviewed earlier in this dissertation (e.g. Hamlin et al., 2007, 2010; Hamlin & Wynn, 2011), has also been interpreted in terms of an early-emerging partner choice psychology (Kuhlmeier et al., 2014; Van de Vondervoort & Hamlin, 2016).

Overall, these studies purport to show that young children, and possibly even infants, can extrapolate prospective partner quality from observing other agents by ascribing dispositional character traits to them, and tend to prefer helpful, fair, and generous cooperators. Infants and children do so even when they themselves were not the recipients of the previous prosocial behavior.

Finally, recent research explicitly investigating children's choice of partners has demonstrated that they can track an agent's past behavior, e.g., whether he possesses a certain skill (Grueneisen et al., 2023; Hermes et al., 2016), or how he behaves in a social scenario (Grueneisen et al., 2023; J. W. Martin et al., 2022; Prétôt et al., 2020; A. M. Woodward et al., 2022). In these studies, children as young as 3 years of age let this information guide their partner choice decisions.

Co-Collectors: Aims of the present project

One limitation of previous studies is that children were generally asked to make hypothetical assessments about fictional characters, and were not choosing an actual partner for a cooperative endeavor (Droege & Stipek, 1993; Feldman & Ruble, 1988; Hermes et al., 2015, 2016; Prétôt et al., 2020). Thus, unlike in the real world, partner choice in these tasks was low-stakes and picking a “bad” partner had no tangible consequences for the child.

Another limitation is that studies often relied on verbal vignettes as stimuli and experimenter prompts for eliciting responses. This may have affected the results of these studies, as the framing could have (inadvertently or deliberately, see Rholes & Ruble, 1984) highlighted to children the socially relevant features of the situation and the concept under investigation (Fitneva & Dunfield, 2010; Heyman & Gelman, 1999). For example, if an experimenter asks a question like “Who is *daxier*?”, this provides the participant with a relevant trait label, conveys that the agents potentially possess the trait of being *daxy*, and implies that there is a difference in *daxiness* between agents. In many real-world partner choice scenarios, agents have to identify important trait dimensions directly from behavior observation, and little research has dealt with the question of whether and how children (and adults) would infer traits and choose good partners for cooperation merely on this basis.

Moreover, if the behaviors or traits are highly desirable and normatively prescribed in children’s social environment (e.g., it is good to play nice, share, and help), designs in which an experimenter elicits a response could bias children towards responding in a norm-consistent way due to reputational concerns (Engelmann et al., 2012; Rapp et al., 2019). In real-world partner choice, it can be advantageous to interact with an agent who violates these norms (but see Dhaliwal et al., 2022) because she possesses other traits that are relevant for providing benefits in a specific context (e.g., selecting a ruthless, aggressive lawyer to defend oneself). We do not know whether children’s responses would vary accordingly in a task where they choose partners independently of adult supervision.

In collaboration with a software company, we devised an iPad application called Co-Collectors, which can be used in research on partner choice as well as other aspects of social cognition (Schlingloff-Nemecz et al., 2022). The app consists of a foraging game, in which computer-animated agents as well as the player herself collect resources, either alone or together, and the player’s payoff depends in part on the behavior of their partner. Here, children can gather information about social partners by observing their actions. Instead of an experimenter priming concepts of interest by verbal descriptions, children have to infer agents’ traits from observed behavior patterns—insofar as they interpret these patterns as indicating stable dispositions.

We focused on the character traits of helpfulness and skill, which are commonly used in the literature as examples for a partner’s willingness and ability to confer rewards to the agent. By design, these traits can be varied continuously in our game. This means that agents could be relatively more or less helpful, and relatively more or less competent.

Moreover, the trait dimensions are orthogonal to one another: Someone can be high in prosociality and in competence, high in only one of them, or high in neither, making it possible to investigate both the individual and the interaction effects of these traits.

Our study targeted the age range around the onset of formal schooling, from late preschool through the elementary school years (i.e., 5- to 10-year-old children), as prior studies have found development in trait attribution during this time. We also wanted to explore whether there might be systematic differences in children's preferences and relative weighting of prospective partners' traits that could be driven by cultural influences or socialization (e.g., schooling). Extant literature has compared trait reasoning and behavior explanation in subjects from Western (U.S.) and Eastern (Asian) cultural contexts, generally finding that the latter rely less on traits and internal dispositions and more on situational and relational causes for behavior, and that this difference emerges in childhood (Chen et al., 2016; I. Choi et al., 1999; Lockhart et al., 2008; J. G. Miller, 1984; Morris & Peng, 1994; Na & Kitayama, 2011; Y. Shimizu et al., 2017). For this reason, we conducted experiments in both Central Europe (Hungary, Austria) as well as East Asia (Japan). The experimental materials we used in these two contexts were as similar as possible.

The Co-Collectors game

Our aim was to investigate children's partner choice in a situation where

(1) potential partners' collaboration-relevant traits have to be inferred from observing their nonverbal behavior,

(2) partner choice has a concrete consequence for children's payoff in the task at hand, rather than being merely abstract or hypothetical, and

(3) children can choose a partner without being prompted or supervised by an experimenter.

In our game, participants can both play themselves and watch other agents play. After observing animated characters whose 'personality' profiles vary in prosociality and competence, the participant can choose which of the characters she wants as a partner for a subsequent round. The participant thus has to infer from the agents' previous behavior who would make a better partner (Goal 1). How well the participant does in the game depends in part on whether she identifies and chooses the relatively more prosocial or more competent partner (Goal 2). After an introduction and practice phase, children can play independently without supervision; researchers' intervention is not necessary for recording responses as the app automatically logs participants' choices and foraging actions (Goal 3).

The game has a particular payoff structure, where players collect resources for themselves, but there are opportunities for mutualistic collaboration, as well as for altruistic helping. Participants in the game should be primarily motivated to maximize their own payoff, and select partners that aid in the pursuit of this aim, i.e., who positively affect the participant's outcome. This means on the one hand that they should choose a high-skilled partner. Such a partner is advantageous even if he is not particularly altruistic, because in a mutualistic cooperative interaction he will generate windfall benefits for his partner even if he otherwise pursues a selfish benefit-maximizing strategy. On the other hand, participants should also choose highly prosocial agents who act in an altruistic manner. This is beneficial because when the collaborative option does not always yield the highest payoff, choosing an altruistic partner who sacrifices his own resources on behalf of another is a good strategy.

It should be noted that it is a unique feature of the game that an agent's degree of prosociality—and, more generally, the fact that prosociality is a relevant dimension—must be inferred by the observer from the agent's pattern of approaching specific types of resources, a behavior that is novel and specific to the game context. The observer cannot simply rely on verbal cues by the experimenter (e.g., indicating that an agent may be "nice" or "mean"), or on familiar prototypical features of prosocial actions. Instead, observers have to infer that one agent investing effort (or even forgoing her own rewards) to contribute to another agent's rewards constitutes helping.

Gameplay, agents, and resources

In the Co-Collectors game, small, colorful, monster-like autonomous agents called "kobos" collect berries. The berries are contained inside "coconuts", which fall from a tree and must be cracked open by the kobos using little hammers (Figure 3.1). If a kobo manages to hit a coconut enough times, the coconut opens and the berries it holds fly into the kobo's basket. If, however, the kobo does not deliver enough hits within a certain amount of time, the coconut disappears and its berries are lost.

Kobos can move left and right on the ground in horizontal directions to reach the coconuts, which fall from a large tree at random locations. In each round where there are two kobos, they can move simultaneously and independently from one another and don't display any explicit cues of social interaction.

The coconuts that fall from the tree can be either colorful ones, which match in color a kobo in the scene, or brown ones. Colorful coconuts are "proprietary": They contain berries of the same color, which are awarded to the matching kobo, irrespective of who cracks them open. In contrast, brown coconuts are "shared": They contain berries of two different colors, corresponding to the two kobos present in the scene. When the coconut is cracked, the berries inside automatically fly into the basket of the kobo with matching color, such that stealing berries is not possible. The height of the berry piles in the baskets represents the number of berries collected.

Kobos can crack coconuts alone or jointly. In joint cracking, they deliver hits on opposite sides of the coconut (see Figure 3.1). When one of the kobos is operated by a player (always a yellow one), she can move the kobo by dragging it left or right on the screen. The player can crack coconuts by positioning her kobo next to a coconut (at which point a hammer appears in the kobo's hand) and tapping on it; for each tap, the kobo hits the coconut once with its hammer. At the end of a round, the number of berries collected by each kobo is displayed in front of their houses. In rounds where the participant herself plays, she can enter her kobo's house (by tapping on it), where the berries collected throughout the game accumulate in a pile on the ground.

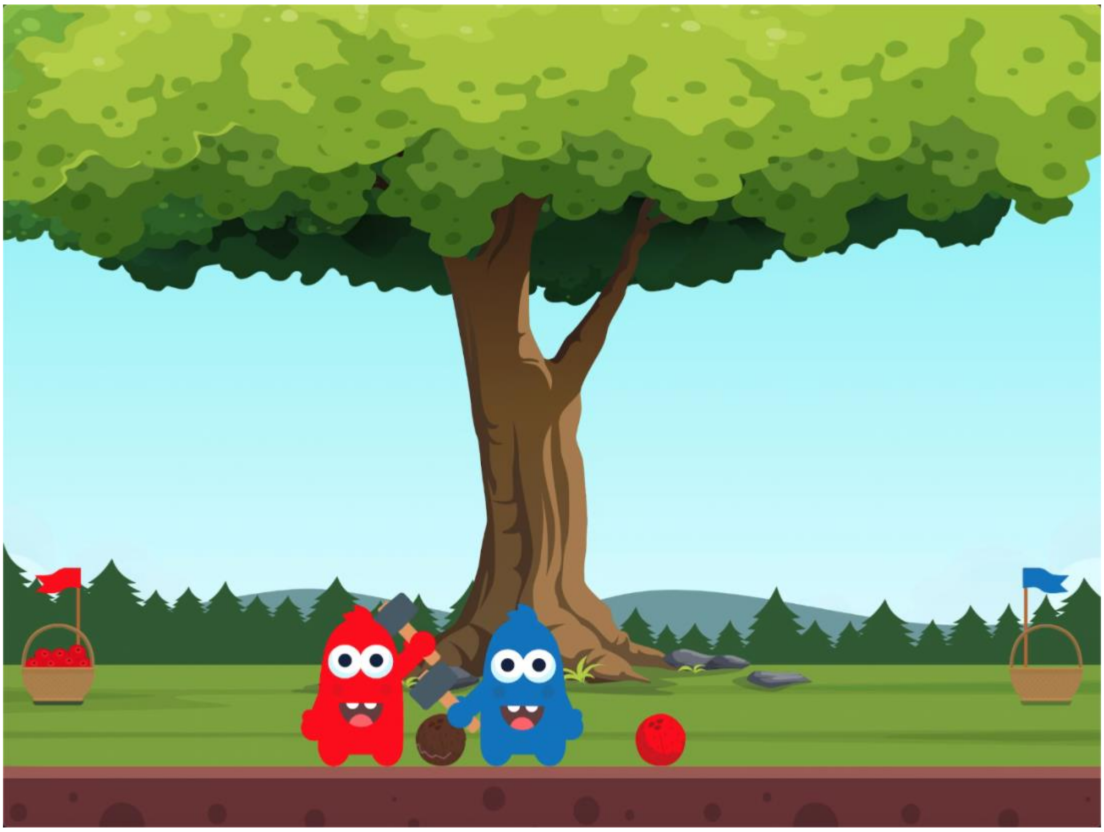

Figure 3.1. Screenshot of an Observation round.

A game is composed of a sequence of rounds specified by the researcher in advance. There are different types of rounds, aimed at introducing participants to the game (Practice rounds), allowing participants to observe other agents collect berries (Observation rounds, Figure 3.1), to choose a partner (Partner Choice rounds, Figure 3.2), and to play together with this partner (Cooperation rounds).

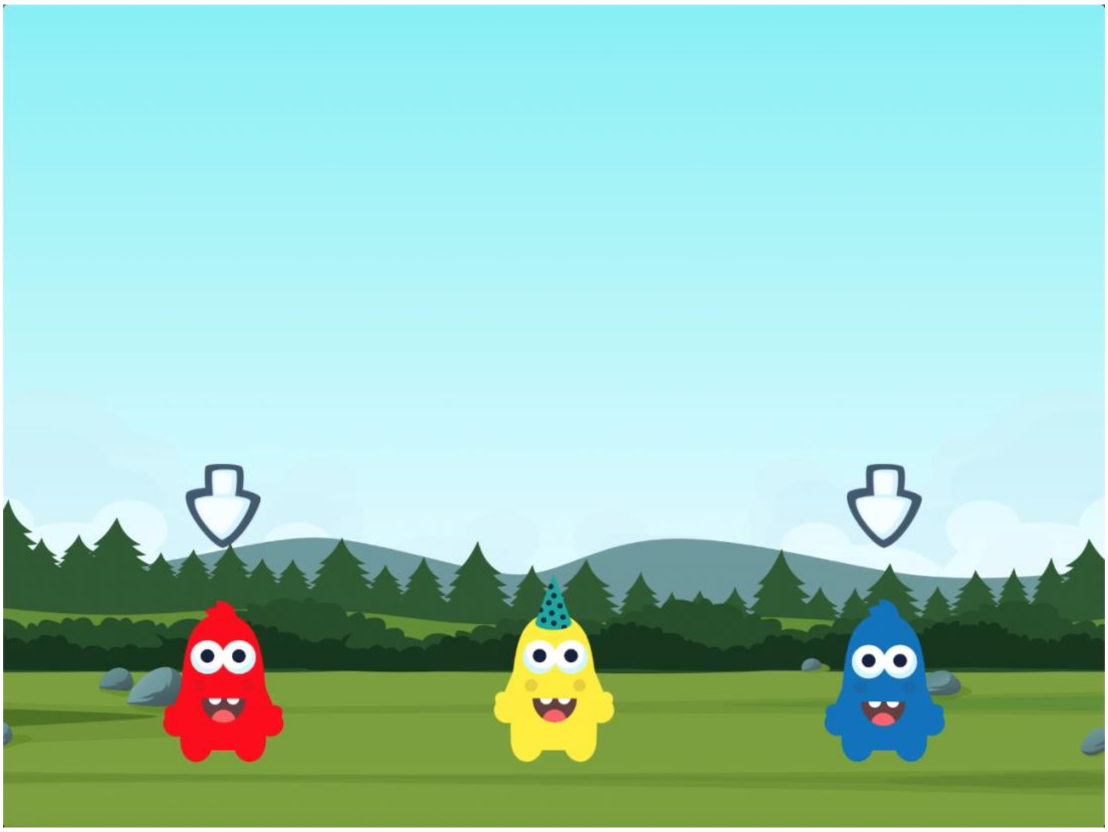

Figure 3.2. Screenshot of a Partner Choice round.

The structure of the game, as well as its specific parameter values (described in the following section), are set through an online configuration interface. For a detailed description of the parameters and structure of the configuration, see Appendix D.1. During play, the game records the actions of the participant (e.g., partner choices, tapping coconuts) as well as the timing of these actions, and saves them to disk. In Appendix D.2, we describe the log files that are thus generated.

Game settings and parameters

The game was designed such that its progression and its parameters for the environment and the agents can be flexibly set by the experimenter. Many of these parameters are stochastic: They are either defined as a value selected randomly from a uniform distribution over a parameter range or represent the probability of certain events to occur.

The environment parameters define the sets of coconuts available for the players in various rounds. These include the number of coconuts in a round, the probability distribution of coconuts between the two colorful coconuts and the brown coconuts, the coconuts' hardness (how many hits are required to crack them before they spoil), and the duration that a coconut exists before it spoils.

The agent parameters specify behavioral characteristics along two dimensions: skill and prosociality.

Skill. In the context of our game, a highly skilled kobo is one who, all else being equal, is more successful at cracking coconuts and gathering berries. This is implemented by defining the speed at which an agent can approach coconuts, and the speed at which he can deliver hits. Because each coconut has internal parameters that specify how much time can elapse until it spoils and how many times it needs to be hit before cracking open, a kobo who is faster is more likely to successfully crack more coconuts and gather berries. To infer that the higher success rate of one kobo is a sign of its skill, the observer has to assume that this outcome is driven not by random variation in, for example, coconut hardness (such that one kobo fortuitously happens to interact with easily crackable nuts), but that the outcome is caused by its faster response and hit frequency.

Prosociality. Social behaviors can be categorized by who pays a cost and who gains in an interaction. In our game, an agent's prosociality is implemented by a set of parameters that guide which types of coconuts a kobo will likely approach. Importantly, because it is irrelevant whether the hits are delivered by one or two kobos, the chances of successfully cracking a nut tend to be higher when two kobos work together, compared to when a kobo attempts to do so alone. Cracking brown coconuts is a mutualistic behavior (as both kobos receive berries when it is opened), while cracking the partner's coconut is altruistic (as only the partner receives berries). Conversely, not helping the partner with its coconuts is selfish, and not cracking brown coconuts is spiteful (because the kobo is forgoing a reward for itself in order to prevent the partner from benefiting). An additional cue for the agent's prosociality is provided by the likelihood with which an agent switches from one coconut to a new one that appears. This is indicative of whose welfare the kobo values *more*, and whether and how much of an opportunity cost a kobo is willing to pay in order to help. Thus, for instance, a very altruistic kobo will abandon its own as well as brown coconuts to help its partner, sacrificing its own berries for the sake of the other. A moderately helpful but ultimately selfish kobo on the other hand will abandon the other's and the brown coconuts when one of its own color becomes available.

By adjusting the values of the skill and prosociality parameters, one can create profiles for the kobos, which specify how they behave in the game. For example, there might be an "altruistic, somewhat skilled", or a "selfish, highly skilled" kobo. This way, one can systematically vary the target traits between kobos, flexibly generate Observation and Cooperation rounds that implement the desired kobo personalities, and test whether participants are sensitive to these traits and use them for selecting partners. Kobos chosen as partners after an Observation round always have the same personality profile in a subsequent Cooperation round (e.g., a kobo who acted prosocially towards a third party also helps the participant collect resources with the same probabilities).

The present experiments

Our main aims in the present study were to investigate (1) whether children would prefer partners who are relatively more prosocial and more skilled when playing the game by themselves, (2) whether they would systematically prioritize one trait over the other, (3) whether there would be differences between the choices made by the Central European and Japanese participants, and (4) whether adult participants would behave differently than children in the same task.

In Experiment 3.1, we tested whether 5- to 10-year-old children would recognize differences in the agents' behaviors and identify the kobos who helped more and who were faster when asked by an experimenter. This served as a comparison baseline for Experiment 3.2, where children of the same age, as well as adults, played the Co-Collectors game with the same stimuli, but now had to choose partners with whom to play. In Experiment 3.3, we slightly modified the experimental procedure to test children's choice of potential partners who varied in prosociality or skill, but not both. Finally, in Experiment 3.4, we probed whether elementary-school-aged children would prioritize one of these traits when the two were in contrast with one another. All materials (stimuli, experimental scripts, data, models, and their output) can be accessed at https://osf.io/p8uhj/.

Experiment 3.1a

Before conducting the experiment with the iPad paradigm, we wanted to ensure that children attend to the behaviors of the kobos when observing them in a third-party context; i.e., that they can recognize, when watching clips from the Co-Collectors game, the differences between kobos in prosociality and skill. This served (1) to establish that they understand the relevant aspects of the stimuli, and (2) as a comparison with the task used in Experiment 3.2, where children chose partners from among the same kobos. The experiment was run during the Covid-19 pandemic; therefore, testing took place online on Zoom, in a live call with an experimenter.

Methods

The experiment was preregistered on the OSF (https://osf.io/abr98).

Stimuli

The test phase stimuli were 8 video clips of kobos collecting berries, each roughly 1 minute in length. These 8 trials were of 4 different types (2 trials per type): Prosociality trials, in which agents differed only in how much they helped the other; Skill trials, in which agents varied in how competent they were at cracking coconuts; Contrast trials, in which agents varied in both of these traits such that one was highly competent but selfish and the other was helpful but incompetent; and Success trials, where agents did not differ in their traits (they were both moderately competent and somewhat helpful), but the distribution of resources was skewed such that one agent received more berries than the other.

Procedure

We ran the experiment online, using the Zoom video chat software for interacting with participants, and Slides.com for presenting stimuli. The child first played a warm-up game with the experimenter, where she was asked to identify colorful animal figures. This helped establish whether a child has problems with color vision or color naming, which could have affected the way she points out agents with color labels.

After a calibration clip, the experimenter introduced the premise of the stimuli, and explained the structure of the game. Importantly, this was done without any explicit reference to agents' traits or individual differences. The child was asked to respond to several comprehension questions, to confirm that they understood the reward structure and contingencies of the game. If a child answered incorrectly to a comprehension question, the experimenter corrected her and repeated the question up to 2 times.

At this point, caregivers were asked to close their eyes or turn away from the screen, so as not to bias the child's responses. The experimenter moved on to the test phase, where children watched 8 test trial videos. After each video, children were asked a question. Depending on the trial type, the question was "Which one helped the other more?" (Prosociality trials and one of the two Contrast trials), "Which one was faster?" (Skill trials and the other Contrast trial), or "Which one collected more berries?" (Success trials). Trials were presented in pseudo-random order. Thus, when they watched the videos, children did not know what question they would be asked afterwards. Children answered by stating the color of the kobo. Their responses (correct vs. incorrect) were recorded by the experimenter.

We counterbalanced trial order (one group received the reverse order compared to the other) and agent identity (i.e., the color of the kobos). To be included in the sample, children had to respond correctly to all comprehension questions during the introduction phase after a maximum of three prompts and provide a valid response to at least one trial per trial type. Moreover, if a child selected the agent on the same side across all trials, that child was excluded.

Participants

Sixty 5- to 10-year-old children participated in Experiment 3.1a. They were recruited such that their ages spread approximately evenly across the prespecified range (~10 per year cohort, mean age: 7.8 years). Additionally, two participants were tested for the study, but had to be excluded due to caregiver intervention (n = 1) or experimenter error (n = 1). The participants were Hungarian children, recruited via Facebook advertisement. Informed consent from caregivers and children was obtained before the experiment. The experiment, as well as all the following ones that were conducted in Hungary (3.2a, 3.2b, 3.3a), received full ethical approval from the United Ethical Review Committee for Research in Psychology (EPKEB) in Hungary and was conducted according to the principles of the Declaration of Helsinki. Participants received a gift voucher for their participation.

Coding and analyses

Children's responses were recorded by the experimenter. We analyzed the data from this and all following experiments by fitting Bayesian logistic regression models to estimate the response parameter. This approach has the advantage that it allows us to describe the full posterior distribution of parameter estimates and avoids the issue of having to correct for multiple comparisons. As described in the preregistration, we wanted to assess whether children correctly identified the target agent above chance. More specifically, we tested whether the proportion of correct responses varied by trial category (i.e., trials where only one of the agents' traits differed (Prosociality/Skill), both (Contrast), or only the outcome differed (Success)), whether children were more accurate in identifying who helped or who was faster, and whether responses differed as a function of the serial position of the trial. Children's age, gender, and the counterbalanced stimulus group were added as predictors. We report 89% credible intervals (CI) of parameter estimates for responding correctly (where 0 is incorrect, 1 is correct, and 0.5 is the chance level), or for the difference between trials or groups (where chance is excluded if the 89% CI does not include 0). These credible intervals specify that the true parameter value lies in this interval with 89% probability. In this and all following experiments, we report primarily only the findings that answer our main questions or where chance is excluded; for all results, see the model code and outputs on the OSF (https://osf.io/p8uhj/). The analyses were performed in R (version 4.2.2, R Core Team, 2023), using the rethinking package (version 2.0.1, McElreath, 2020).

Results

Children responded correctly above chance in all four trial types (Prosociality: mean = 0.801, 89% CI: [0.741, 0.857]; Skill: mean = 0.805, 89% CI: [0.747, 0.86]; Contrast: mean = 0.876, 89% CI: [0.824, 0.924]; Success: mean = 0.674, 89% CI: [0.599, 0.746]; see Figure 3.3). They were more accurate in Contrast trials, where both traits varied, compared to trials in which agents differed only in one trait (difference between trial categories: mean = 0.071, 89% CI: [0.012, 0.132]), This was because they were better at answering which agent was faster in Contrast trials compared to Skill trials (mean = 0.099, 89% CI: [0.02, 0.175]), but slightly less so for the question who helped more in Contrast vs. Prosociality trials (mean = 0.034, 89% CI: [-0.05, 0.117]). Children were more accurate in both of these trial categories compared to Success trials, in which only the distribution of resources varied (Contrast vs. Success: mean = 0.202, 89% CI: [0.111, 0.294]; Prosociality/Skill vs. Success: mean = 0.13, 89% CI: [0.051, 0.213]). They were equally likely to identify who helped more and who was faster, whether agents varied only in one trait dimension (mean = -0.004, 89% CI: [-0.081, 0.071]) or in both, i.e., in Contrast trials, though here there was a tendency to be more accurate in recognizing speed differences (mean = -0.069, 89% CI: [-0.154, 0.01]).

Children's accuracy tended to improve with age, though only in Prosociality trials did the age parameter estimates exclude chance (mean = 0.462, 89% CI: [0.097, 0.834]; see Figure D.1, Appendix D.3).

Finally, we examined whether children responded more accurately in later trials, after having gained more experience with the stimuli. A regression analysis indicated that accuracy somewhat increased across trials, although the chance value was not excluded (mean = 0.112, 89% CI: [-0.071, 0.293]). Children's accuracy was, on average, already above chance in the beginning (see Figure D.2, Appendix D.3).

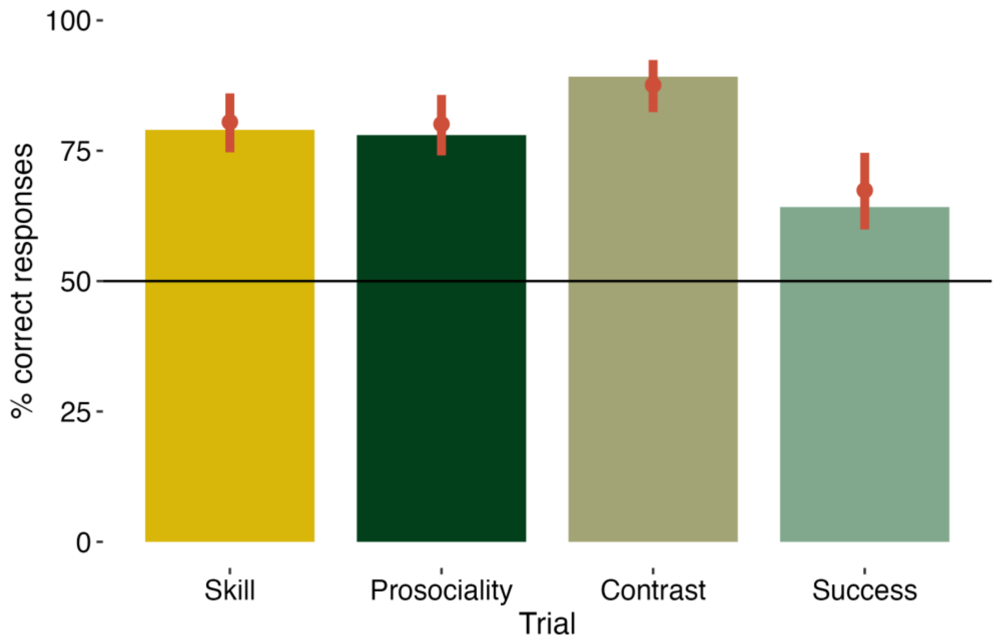

Figure 3.3. Bar plots representing the proportion of correct responses in Experiment 3.1a.

Discussion

The results from Experiment 3.1a provided evidence that children in our target age group can recognize differences in agents' behaviors along the predetermined trait-relevant dimensions in the stimuli generated with our Co-Collectors game. Even at a young age, from 5 years, and already from the first test trial on, participants attended to these behaviors without being explicitly cued about them. We also found that children's responses were the least accurate in the Success trials, which suggests that they did not solve the task posed to them in the other trials by merely attending to outcomes and comparing which agent obtained more resources.

It is important to note that in this task we asked children about behaviors rather than traits. Concluding that behavior differences were caused by variation in underlying stable character traits is not a given, but requires an additional inferential step. Thus, children may have thought that a kobo was fast either because he was skilled or strong, or because he happened to be momentarily more motivated to collect berries. Analogously, a kobo may have helped because he is generally nice, or because he had a special affinity for the kobo he was partnered with. However, for partner choice, trait inference is necessary because only stable traits could mediate between observed past behaviors and expected future behaviors in a novel context and with a different partner. We turn to the question whether children attribute traits to kobos, and whether they choose partners on the basis of these attributions, in Experiments 3.2-4.

First, however, we replicated Experiment 3.1a with Japanese children.

Experiment 3.1b

The same experimental procedure and stimuli were used to test a sample of children in Fukuoka (Japan).

Methods

Stimuli

The stimuli were the same as in Experiment 3.1a. The testing script was translated into Japanese.

Procedure

The experimental procedure was the same as in Experiment 3.1a.

Participants

60 children in the same age range as Experiment 3.1a participated in Experiment 3.1b (mean age: 8 years). Additionally, 1 participant was tested for the study, but was excluded due to dropping out early. The participants were Japanese children, recruited through the database of "Kyushu University Infant/child Scientist Project", whose caretakers had volunteered to participate in infant studies at Kyushu University. All participants gave informed consent prior to participating in this study after receiving a full description of the nature of the procedures. The experimental protocols were approved in advance of the study by the ethics committee of Kyushu University and were conducted in accordance with the Declaration of Helsinki. Participants received a gift voucher for their participation.

Coding and analyses

Responses were coded and data analyzed the same way as in Experiment 3.1a. Additionally, we directly compared the results from the Hungarian and the Japanese sample.

Results

Similar to Hungarian children, Japanese children's responses were also correct above chance in all four trial types (Prosociality: mean = 0.727, 89% CI: [0.661, 0.788]; Skill: mean = 0.8, 89% CI: [0.742, 0.853]; Contrast: mean = 0.838, 89% CI: [0.784, 0.887]; Success: mean = 0.77, 89% CI: [0.709, 0.828]; see Figure 3.4). They were overall more accurate in Contrast trials compared to those where only one trait varied (mean = 0.071, 89% CI: [0.009, 0.134]), though when comparing Contrast to Prosociality and Skill trials separately, chance was not excluded for either (Contrast vs. Skill: mean = 0.054, 89% CI: [-0.025, 0.133]; Contrast vs. Prosociality: mean = 0.088, 89% CI: [-0.004, 0.179]. In contrast to the Hungarian children, Japanese children's accuracy in the Success trials was similar to that other trial types (Contrast vs. Success: mean = 0.068, 89% CI: [-0.006, 0.145]; Prosociality/Skill vs. Success: mean = -0.004, 89% CI: [-0.069, 0.066]).

In a direct comparison, Hungarian children were less accurate in Success trials than their Japanese peers (mean = -0.096, 89% CI: [-0.189, -0.007]); in no other trial did responses differ between the groups.

Japanese children's responses changed with age only in Skill trials (mean = 0.35, 89% CI: [0.004, 0.7]) and when asked about agents' helpfulness in Contrast trials—here, older children responded correctly more often than younger children (mean = 0.475, 89% CI: [0.015, 0.942]) (see Figure D.1, Appendix D.3). We also assessed whether the trial position played a role and found that children's accuracy increased over the course of the experiment (mean = 0.244, 89% CI: [0.058, 0.428]), though Japanese children's responses, too, were already above chance in the first trial (see Figure D.2, Appendix D.3).

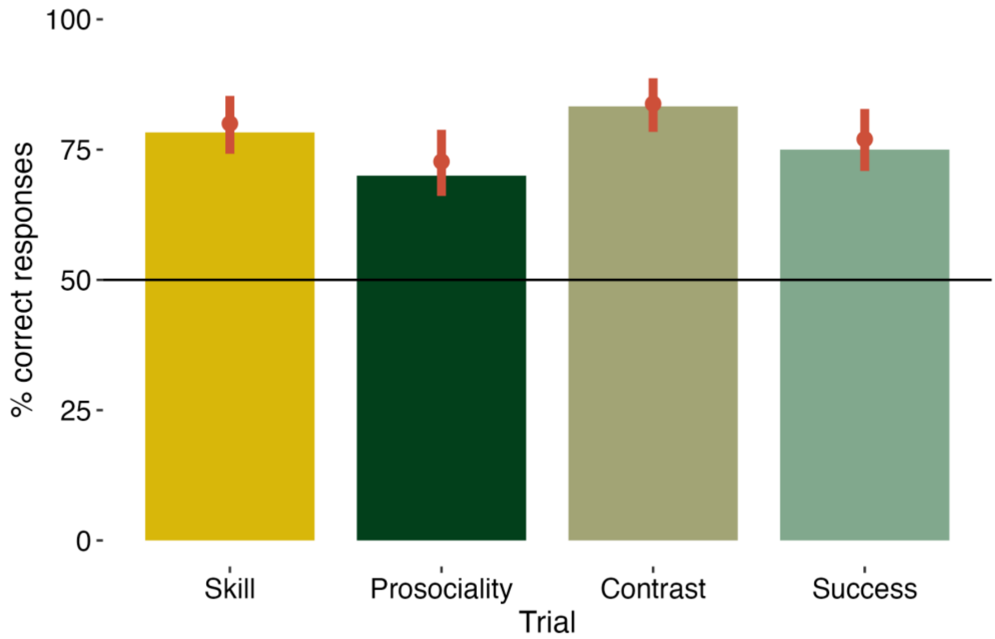

Figure 3.4. Bar plots representing the proportion of correct responses in Experiment 3.1b.

Discussion

Just like their Hungarian counterparts, Japanese children were successful at identifying behavior differences between the agents. The main difference we found between the two groups was that children in the Japanese group paid more attention to differences in payoff that were not driven by agents' behaviors (Success trials).

Experiment 3.2a

In the next step, we aimed to test whether children would use the information obtained by observing kobos in a third-party context to infer context-invariant traits and guide their partner choice when playing the Co-Collectors game.

Methods

The experiment was preregistered on the OSF (https://osf.io/y5sbm).

Stimuli

The stimuli used in Experiment 3.2 were largely the same as in Experiment 3.1. In the introduction phase, children additionally participated in trials where they got to play the game themselves. First, children could choose their avatar from a set of different yellow kobos; they were told that they would play as this kobo for the rest of the game. In a first Practice round, children played by themselves, to familiarize them with the task. Subsequently, children played through a brief version of an experimental trial sequence, consisting of an Observation round with two kobos collecting berries, a Partner Choice round, and a Cooperation round where the child was partnered with the kobo they selected. Having been familiarized with this structure, children could then play the game by themselves, without instructions from the experimenter. The Observation and Cooperation rounds during this training were short and contained only two brown coconuts each, so as to not bias children towards expecting helpfulness (or lack thereof) from a partner.

The test phase featured the same video clips used in Experiment 3.1 as Observation rounds. Each clip was followed by a Partner Choice round, in which the child's yellow kobo avatar appeared in the middle between the two previously seen kobos and children could choose one of them (Figure 3.2), and finally a Cooperation round.

Procedure

This experiment was conducted in-person, using an iPad. Data were collected at different locations: at the Zoo Lab and Babylab of the CEU Cognitive Development Center, and at various Budapest preschools.

The experimenter introduced children to the game, using the same script and comprehension questions as in Experiment 3.1, with the addition of a practice phase to familiarize children with the game.

After reminding children of the trial structure during the test phase, the experimenter told them that they would now play by themselves, then turned away and pretended to work. Caregivers accompanying the child could attend for the introduction phase if they wanted to, but were asked not to intervene or comment; for the test phase, the experimenter asked them to fill out some paperwork so that they would not supervise children's gameplay.

The order of stimuli and counterbalancing of factors (order and kobo identity) were the same as in Experiment 3.1.

We excluded participants who failed to respond correctly to the comprehension questions within three attempts, selected the agent on the same side of the screen (right or left) on all eight experimental trials, did not attend to the stimuli during the test phase, or did not provide data from at least one trial per type before quitting study participation.

Figure 3.5. A child plays the Co-Collectors game.

Participants

Sixty-five participants took part in Experiment 3.2a. The age range was the same as in Experiment 3.1a[12], as was the approximate distribution of participants across this range (mean age: 7.8 years). An additional 13 children participated but had to be excluded for failing the comprehension questions (n = 2), inattentiveness (n = 2), dropping out before having contributed a minimum of 4 trials (n = 4), experimenter error (n = 4), or technical failure (n = 1).

Participants were recruited through various means, depending on the testing location: by approaching visitors at the Budapest Zoo (Zoo Lab), by contacting caregivers through information sheets brought home by children (preschools), and through the lab's participant database (Babylab). Informed consent from caregivers and children was obtained before the testing session. Children received stickers for their participation.

Coding and analyses

Children's behaviors in the app—in particular, their partner choices—were automatically logged by the game. The analysis was similar to that in Experiment 3.1: Now we assessed

[12] One child was tested who had not turned 5 yet but was kept in the sample nonetheless.

the probability of children choosing a specific partner, in particular, one who was more helpful and more competent (we considered these choices as ‘correct’). We tested whether children had a relatively stronger preference for helpful or competent partners, whether their preference was more pronounced in trials where agents varied only in one trait (Prosociality, Skill) compared to both (Contrast), and whether responses differed as a function of the serial position of the trial. To rule out that children’s choices were determined by the agents’ success in gathering resources (the relatively more competent and the less helpful agent were more successful), we tested whether this factor could explain children’s behavior across all trial types, and further, whether children’s age, gender, and the counterbalanced stimuli group affected the estimates.

Results

The only above-chance preference children had was that for a helpful kobo in Prosociality trials (mean = 0.607, 89% CI: [0.531, 0.678]); in the other trials, choices did not differ from chance (Figure 3.6). A direct comparison of children’s responses in Prosociality with those in Skill trials showed that the estimates for the difference between the two did not exclude chance (mean = 0.057, 89% CI: [-0.047, 0.16]), and neither in the comparison with Contrast trials, where children could choose a helpful but incompetent partner (mean = 0.088, 89% CI: [-0.018, 0.191]). Children’s preference for helpful partners increased with age: older children reliably picked the more prosocial kobo (mean = 0.715, 89% CI: [0.418, 1.027]). In other trials, there was no substantial change in choice patterns with age (see Figure D.3 in Appendix D.3). When comparing the choices children made in the first trial of a type they encountered to those in the second, their preference for a helpful partner in Prosociality trials increased over the course of the experiment (mean = 0.143, 89% CI: [0.009, 0.277]); in Skill trials, chance was not excluded for this comparison, but children had a similar tendency and the majority of the parameter estimate distribution for this comparison was above chance (mean = 0.111, 89% CI: [-0.023, 0.246]). Finally, children did not show a systematic bias for a more or less successful partner across all trial types (mean = 0.488, 89% CI: [0.317, 0.663]).

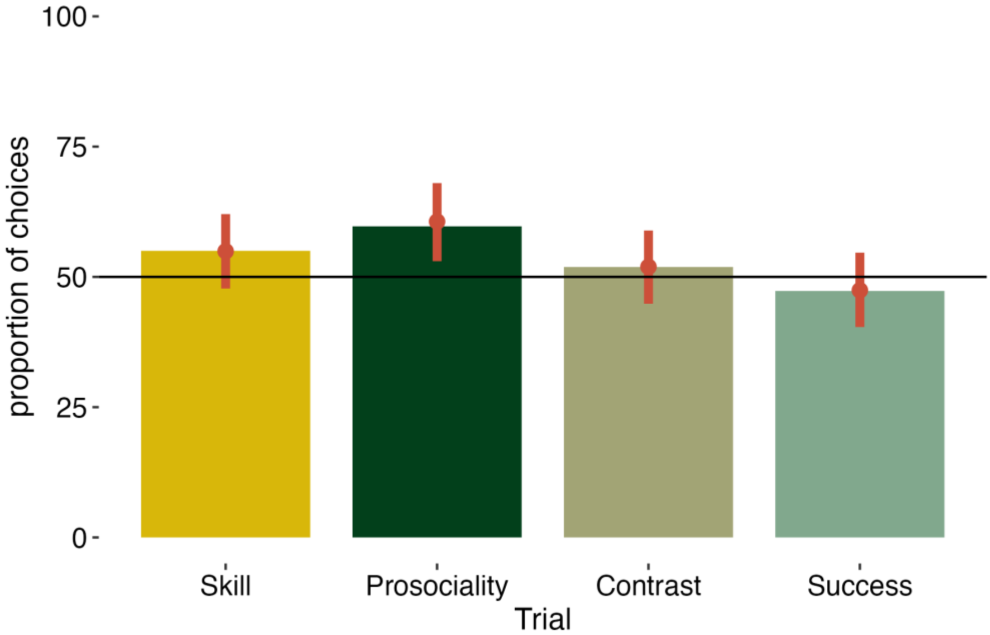

Figure 3.6. Bar plots representing the proportion of correct responses in Experiment 3.2a. Red error bars indicate the 89% CI of the parameter estimates, with dots indicating the means.

Discussion

Children tended to somewhat prefer a more helpful over a selfish partner after observing the two agents in Prosociality trials, and this tendency increased with age (consistent with the developmental pattern in Experiment 3.1a). Children did not have a preference in Success trials, which, together with the result from Experiment 3.1a, suggests that mere difference in outcome was not particularly salient for children and did not constitute sufficient grounds to prefer an agent who acquired more (or fewer) resources (though some previous research has found that young children positively evaluate lucky agents and expect them to perform prosocial actions; see Olson et al., 2006). Children did not prefer a skilled but selfish or a helpful but incompetent partner in Contrast trials, suggesting that they did not prioritize one trait over the other when they were pitted against each other. However, contrary to what we predicted, participants did not prefer a competent over an incompetent partner in Skill trials. This is in spite of the fact that in Experiment 3.1a, children were highly accurate at detecting this difference between agents.

There are multiple explanations for this pattern of results. One option is that children recognized that agents varied in how they behaved (Experiment 3.1), but did not spontaneously attribute this variation to underlying traits of the agents. Because of this, they might not have generated an expectation for how the agents would behave in the future, specifically in a context where agents were paired with the children themselves. In other words, children may have failed at behavior-to-trait inference (e.g., recognizing that one kobo helped more, but not concluding that this was due to an enduring disposition).

Another option is that children failed to generate trait-to-behavior predictions. That is, they did not use the information about agent differences they derived from the stimuli to inform partner choice. Perhaps older children inferred traits and even engaged in sociomoral evaluation by choosing the helpful kobo, but did not understand how these traits would affect their own performance in the game. Recent studies by Woodward et al. (2022) and Hwang and Markson (2020) found that sensitivity to third-party behaviors and sociomoral evaluation on the one hand, and partner choice on the other, can diverge. A similar divergence may have played a role in our task.

Relatedly, it is also possible that at least some children interpreted the foraging task in the game as a competitive one. During or after testing sessions, some children commented that they wanted to "beat" their partner or "get more" than the other kobo. If this was the aim they pursued, they may have chosen a partner who did not appear to pose a threat (i.e., the helpful kobo) and avoided the skilled kobo, who would prove to be a more difficult competitor (Baer & Odic, 2022; Grueneisen et al., 2023; Rule et al., 2023).

A final possibility is that the task was too demanding for children: In Experiment 3.2a, they had to not only comprehend the structure of the game and attend to agents' behaviors, but had to learn how to play the game themselves and infer in what ways the variabilities in agents' behaviors were relevant for their own payoff. This, in combination with the complexity of the stimuli structure (children received four different trial types, only two trials per type, and all trials were presented in a pseudo-random order so that one could not predict which characteristics would be relevant in a subsequent trial), may have hindered children's performance. In fact, we found that the proportion of children's choices for helpful and skilled partners in Prosociality and Skill trials, respectively, increased in the second trial of a type they encountered (Prosociality: from 52.3% to 67.2%, Skill: from 49.2% to 61%; though note that for the latter difference, the analysis did not exclude chance). This suggests that children learned to choose a better partner from the experience of playing the game.

We attempted to address some of these potential shortcomings in Experiment 3.3. However, first we wanted to compare the behaviors of children in our task with that of adults, to confirm that adults would respond according to our predictions, see whether their choices would differ from those of the children, and test whether adults would systematically prioritize one trait over another in Contrast trials. We therefore tested adult participants in Hungary (Experiment 3.2b) and Japan (Experiment 3.2c) with the same procedure as in Experiment 3.2a.

Experiment 3.2b

Methods

The experiment was preregistered on the OSF (https://osf.io/jgs35).

Stimuli

The stimuli were the same as in Experiment 3.2a.

Procedure

The experimental procedure was the same as in Experiment 3.2a. Data were collected in our laboratory.

Participants

Thirty-two adults took part in the experiment (mean age: 29.0 years, age range: 20-42 years). Participants were recruited through the University's Research Participation System (SONA Systems). They gave informed consent and received vouchers in exchange for their participation. All participants that were tested met the inclusion criteria.

Coding and analyses

The logging of responses and analysis were the same as in Experiment 3.2a. Additionally, we compared adults' choices directly to those of the children in Experiment 3.2a.

Results

Participants preferred a more competent partner in Skill trials (mean = 0.746, 89% CI: [0.651, 0.833]) and a more helpful one in Prosociality trials (mean = 0.789, 89% CI: [0.702, 0.865]). In Success trials, they chose randomly between an agent who collected more and one who collected fewer berries (mean = 0.454, 89% CI: [0.351, 0.557]), and in Contrast trials, they also did not show a preference for either partner (mean = 0.486, 89% CI: [0.378, 0.596]; see Figure 3.7). The strength of partner preference in Prosociality and Skill trials did not differ (mean = 0.042, 89% CI: [-0.088, 0.174]).

In comparison to the choices of children in Experiment 3.2a, adults' proclivities to choose the more helpful partner in Prosociality and the more skilled partner in Skill trials were more pronounced (Prosociality: mean = 0.191, 89% CI: [0.067, 0.308]; Skill: mean = 0.203, 89% CI: [0.073, 0.327]).

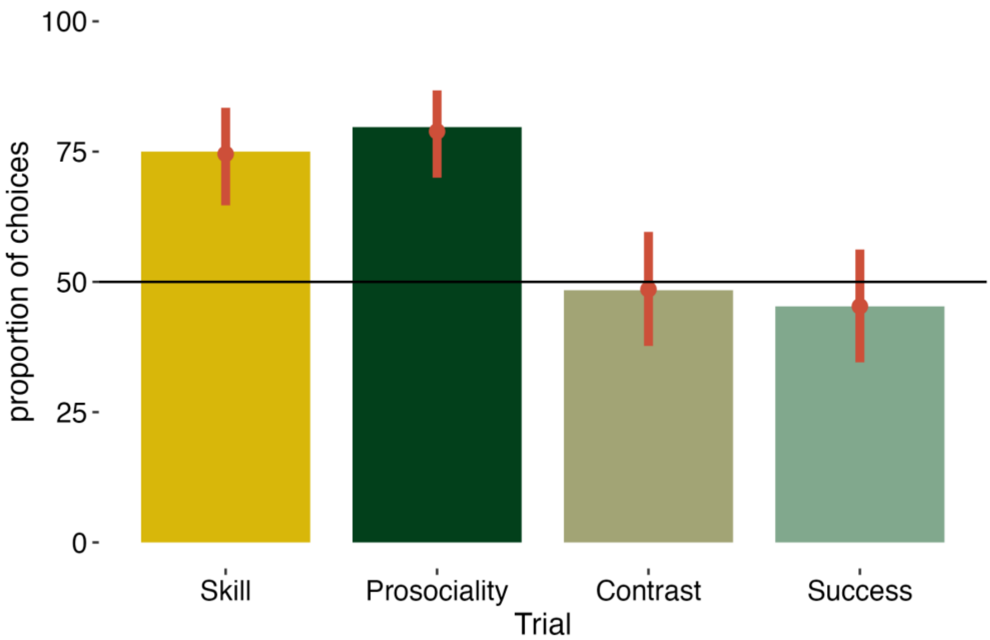

Figure 3.7. Bar plots representing the proportion of correct responses in Experiment 3.2b. Red error bars indicate the 89% CI of the parameter estimates, with dots indicating the means.

Discussion

Adult participants chose partners as we predicted: They preferred a better partner in both Prosociality and Skill trials. However, it should be noted that they did not perform at ceiling, suggesting that the task was not trivial for them, or that they may have engaged in random exploration.

They did not have a systematic preference in Contrast trials at the group level, which might also indicate that they engaged in exploration to see whether one or the other trait would yield higher payoff, or could reflect individual priorities.

We subsequently ran the same experimental design with a sample of participants in Japan.

Experiment 3.2c

Methods

Stimuli

The stimuli were the same as in Experiment 3.2a.

Procedure

The experimental procedure was the same as in Experiment 3.2a.

Participants

Thirty-nine adults took part in the experiment (mean age: 25.1 years, age range: 19-39 years). One additional adult participated but was excluded for displaying a side bias (choosing the agent on the same side on each of the 8 trials). Participants were recruited on the campus of Kyushu University, or through the database of "Kyushu University Infant/child Scientist Project" (in which case not the infants or children, but their caregivers participated). All participants gave informed consent prior to participating in the study. The experimental protocols were approved in advance of the study by the ethics committee of Kyushu University and were conducted in accordance with the Declaration of Helsinki. Participants received a gift voucher for their participation.

Coding and analyses

The logging of responses and analysis were the same as in Experiment 3.2a and 3.2b. Additionally, we compared the results directly to those obtained with the sample of Hungarian adults in Experiment 3.2b.

Results

Adult participants in Japan, like those in Hungary, preferred a helpful partner in Prosociality trials (mean = 0.875, 89% CI: [0.809, 0.929]) and a skilled partner in Skill trials (mean = 0.863, 89% CI: [0.796, 0.92]), and the relative strength of these preferences did not differ (mean = 0.012, 89% CI: [-0.077, 0.102]). They did not show a preference in Success trials (mean = 0.447, 89% CI: [0.348, 0.548]). Unlike the Hungarian sample, however, Japanese participants were not at chance in Contrast trials, but preferred the skilled, selfish kobo over the incompetent, helpful one (mean = 0.643, 89% CI: [0.545, 0.735]; see Figure 3.8). Even so, this preference was less strong than that in Skill trials (mean = -0.22, 89% CI: [-0.328, -0.114]) as well as that in Prosociality trials (mean = -0.232, 89% CI: [-0.351, -0.117]).

When comparing the Hungarian and Japanese groups' responses directly, participants in Japan were relatively more likely to choose a skilled agent in both Skill (mean = 0.117, 89% CI: [0.01, 0.225]) and Contrast trials (mean = 0.158, 89% CI: [0.016, 0.302]).

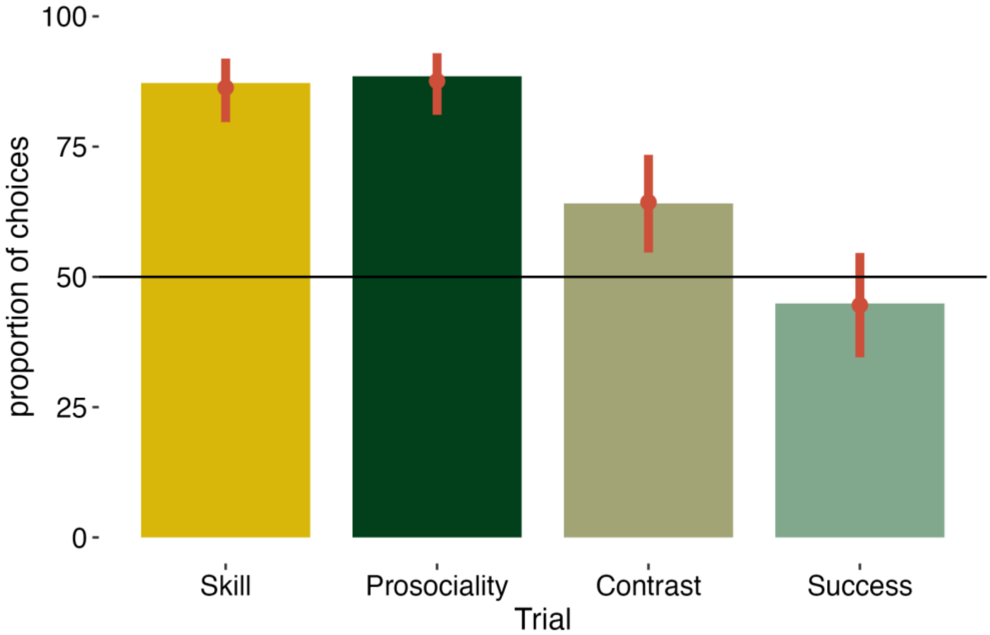

Figure 3.8. Bar plots representing the proportion of correct responses in Experiment 3.2c. Red error bars indicate the 89% CI of the parameter estimates, with dots indicating the means.

Discussion

Overall, Japanese adults responded similarly to the Hungarian participants, preferring prosocial and skilled partners above chance. A main difference between the two groups was that their choice of competent partners was relatively more pronounced in the Japanese sample. In fact, they selected a highly skilled but selfish partner who only pursued his own resources over one who helped but was less competent.

Experiment 3.3a

In Experiment 3.2, children did not show the same pattern as adults: Although they also preferred the relatively more helpful agent in Prosociality trials, they did not show a preference in Skill trials, despite recognizing the relevant behavior variable when asked about it in Experiment 3.1a. There are a number of potential reasons that could account for this discrepancy, which we laid out in the discussion of Experiment 3.2a: They may not have ascribed character traits to agents despite tracking their behavior differences, they may not have understood in what way these traits were relevant for their own payoff, they may have interpreted the task to be competitive, and/or the structure of the game, where children had to attend to a different character trait on each subsequent trial, may have been too demanding for them. To address these issues, we ran another Experiment that was similar to Experiment 3.2a, but in which we implemented some key changes.

We simplified the task in a number of ways: First, children received four trials each of only two trial types (Prosociality and Skill). The trials of the same type were presented together in blocks, which made it more predictable to participants which features of the stimuli to attend to. Moreover, instead of having scenes that contained all types of coconuts, Skill trials now featured only brown coconuts and Prosociality trials contained only colorful coconuts. This was also meant to simplify the stimuli, so that children would not have to learn about and focus on all coconut types at the same time. In Skill trials, it was sufficient to have only brown coconuts such that kobos always cooperated by cracking nuts together, as the kobos here did not differ in helpfulness. In Prosociality trials, agents could act altruistically by assisting their partner with her resources while forgoing their own rewards, or could act selfishly by only pursuing their own rewards. Therefore, these trials featured only colorful (proprietary) coconuts, which allowed disambiguating whether an agent prioritized her own or the partner's payoff.

Further, to diminish the possibility that children would think of the game as competitive such that they stood to gain from outperforming or "beating" their "opponent", we introduced some changes to the introduction phase to convey that cooperating (i.e., hitting a nut together) is more likely to lead to successful nut-cracking compared to working alone. For this purpose, children participated in two additional practice trials during the introduction: one in which they acted alone, and another in which they played together with a preset partner and where they collected more berries. Crucially, we did not want to highlight the possibility of altruistic or selfish behavior, and we did not want to convey the fact that agents in the game could differ in where they fell along this trait dimension. Therefore, the additional practice trials only featured brown coconuts which yielded a payoff for both the participant and the confederate. (It should be noted that the information concerning the advantageousness of collaborating, which we thus provided, should not by itself influence children towards choosing a more helpful partner: If they interpreted this element of the introduction phase in terms of a simple heuristic of choosing agents who are more likely to "work together" on cracking nuts, this heuristic should actually lead them to select the selfish agent as a partner in the subsequent Prosociality test trials. In such trials, children would observe that the selfish kobo receives assistance from his helpful confederate and thus collaborates on all his coconuts, while the helpful kobo does not receive help and often acts alone while attempting to crack her nuts. The selfish kobo is thus seen "working together" more often than the helpful kobo, who works alone more than her partner. Therefore, to choose the latter as a partner, children must infer what these behavior patterns reveal about the agents' dispositions.)

Methods

The experiment was preregistered on the OSF (https://osf.io/7pf8u).

Stimuli

As in Experiment 3.2, children received 8 test trials. However, now there were trials of only 2 types (4 each): Prosociality and Skill trials, grouped in blocks. The trials in the Prosociality block only contained colorful coconuts, and the trials in the Skill block only brown coconuts.

In comparison to the stimuli of Experiments 3.1 and 3.2, we set the trait parameters of the agents to more extreme values, to make the differences between them even more apparent.

We modified the introduction phase by adding some additional trials to clarify that the game did not have a competitive structure. In Observation trials, the experimenter highlighted that kobos are less successful at cracking coconuts alone compared to jointly with another kobo. Subsequently, children could experience this themselves: Playing alone in a Practice trial, it was difficult for the child to open coconuts, while playing with a randomly assigned partner in a Cooperation trial, they succeeded much more frequently. This effect was enhanced by setting the coconut hardness parameter to a higher level in this Practice trial than the Cooperation trial.

Procedure

Data collection took place in our laboratory. The experimental procedure was similar to that of Experiment 3.2a, aside from a few modifications. Firstly, because the trials were now presented in blocks (a "Skill" block and a "Prosociality" block), each of the two blocks was preceded by an introduction phase and corresponding comprehension questions which made participants acquainted with the coconut type that was to appear in the subsequent stimuli (brown coconuts in the Skill block and colorful coconuts in the Prosociality block). A general introduction to the game that did not highlight differences in coconuts was presented to all participants at the very beginning of the experiment.

A further small modification to the procedure was to tell children that the amount of rewards (stickers) they would receive at the end of the experiment would depend on how many berries they collected in the game. This was also meant to incentivize participants more strongly to pursue a benefit-maximizing rather than a contrastive (competitive) strategy.

The order of blocks (Skill first or Prosociality first) and the identity of the kobos within trials were counterbalanced across the participants.

To be included in the analysis, participants had to contribute at least two trials of each type. Further, if a child failed to respond correctly to the comprehension questions preceding the two blocks after a maximum of three prompts, or chose the kobo on the same side across all trials, this child's data was excluded.

Participants

60 children of the same age range as in Experiments 3.1 and 3.2a, i.e. 5- to 10-year-olds, participated in Experiment 3.3a (mean age: 7.8 years). An additional 3 participants were tested but had to be excluded from the sample for displaying a side bias (n = 1), technical failure (n = 1), or quitting the game before participating in the minimum number of test trials (n = 1). Participants were Hungarian children who were recruited from the lab's database. Informed consent from caregivers and children was obtained before the experiment. Children received stickers for their participation.

Coding and analyses

The logging of responses was the same as in Experiment 3.2. As before, we analyzed the data with a Bayesian logistic regression to estimate the probability of children choosing a more helpful and more competent partner. We tested whether either of these preferences was relatively more pronounced. Moreover, we assessed whether there was an order effect: specifically, whether there was an effect of block and an interaction of block with trial type, and further, whether choices changed across the four trials within a block. We also tested whether children's age, gender, and the counterbalanced stimuli group had an effect.

Results

Children chose the more helpful partner in Prosociality trials (mean = 0.622, 89% CI: [0.567, 0.674]) and the more skilled partner in Skill trials (mean = 0.687, 89% CI: [0.635, 0.738]) above chance (Figure 3.9). There was no difference between the strengths of these preferences (mean = -0.065, 89% CI: [-0.139, 0.009]). In both trial types the tendency to choose a better partner increased with age, such that older children were more likely to pick the more helpful and skilled partner (Prosociality: mean = 0.497, 89% CI: [0.259, 0.738]; Skill: mean = 0.766, 89% CI: [0.511, 1.031]; see Figure D.4 in Appendix D.3).

There was an effect of block, such that children's tendency to choose the better partner was higher in the second compared to the first block (mean = 0.096, 89% CI: [0.025, 0.172]; see Figure D.5 in Appendix D.3). In particular, children who encountered Prosociality trials in the first block subsequently chose the more competent partner at a higher rate in the Skill trials of the second block (mean = 0.151, 89% CI: [0.051, 0.26]); for children who first encountered Skill and then Prosociality trials, this effect did not exclude chance (mean = 0.041, 89% CI: [-0.059, 0.144]). There was no interaction with the trial type, suggesting the increase from the first to the second block was similar for Skill and for Prosociality trials (mean = -0.02, 89% CI: [-0.36, -0.311]).

Moreover, children's choice of the better partner increased across the four trials within a block for both trial types (Prosociality: mean = 0.391, 89% CI: [0.128, 0.657]; Skill: mean = 0.475, 89% CI: [0.203, 0.749]).

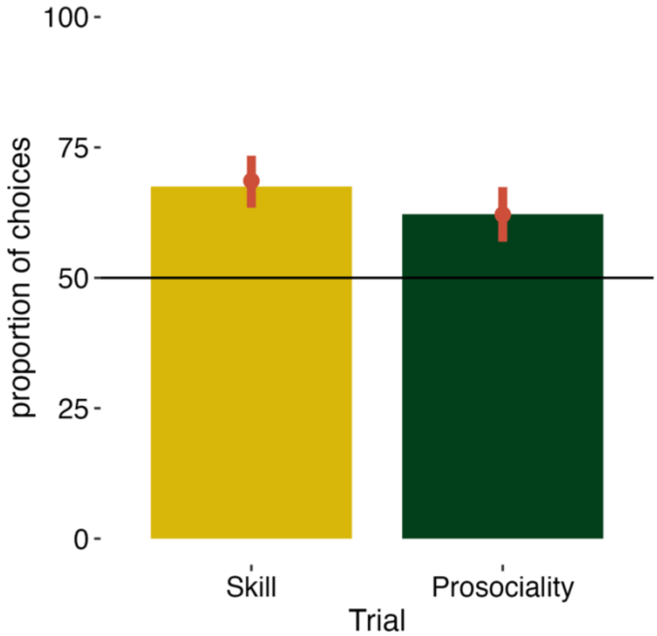

Figure 3.9. Bar plots representing the proportion of correct responses in Experiment 3.3a. Red error bars indicate the 89% CI of the parameter estimates, with dots indicating the means.

Discussion

In Experiment 3.3a, children from around 7 years responded as we predicted, choosing a partner that would help them maximize their payoff both when potential partners' skill and prosociality was varied. This is approximately the age at which, according to prior research, more robust abilities to reason about traits and generate predictions from past behaviors emerge (Fitneva & Dunfield, 2010; Kalish, 2002; Rholes & Ruble, 1984).

This result differed from that of Experiment 3.2a, where children had a slight preference for a helpful, but not a more competent collaborator. The multiple changes to the stimuli and experimental procedure we made may have contributed to this outcome to varying degrees.

In the present experiment, we collected an additional measure from children after the test phase proper. We showed children one video each from the Skill and Prosociality observation trials they had seen earlier, and asked children (1) who from the video was faster or helped more, respectively, and (2) whether it would be better to play with someone who was faster vs. slower, or helpful vs. selfish, respectively. We found that children were extremely accurate: In (1), 58 of 60 children correctly identified which agent in a Skill trial was faster, and 55 of 60 recognized who helped more in a Prosociality trial. In (2), 57 of 60 answered that it would be better to play with a faster, and all 60 children said it would be better to play with a more helpful kobo partner. These accuracy rates likely reflect the fact that the questions were posed at the end of the procedure, after children had gained extensive experience with the game and had opportunities to learn about agents' traits and their effect on one's own payoff. Nonetheless, it is possible that the format of the prompt, whereby an experimenter explicitly asked children and used trait labels, played a role too.

We replicated Experiment 3.3a with a Japanese sample of the same age in Experiment 3.3b.

Experiment 3.3b

Methods

Stimuli

The stimuli were the same as in Experiment 3.3a.

Procedure

The experimental procedure was the same as in Experiment 3.3a.

Participants

Seventy children of the same age range as in the previous experiments participated in Experiment 3.3b (mean age: 7.7 years). An additional 5 participants were tested but had to be excluded from the sample for interference by a parent or sibling (n = 2), dropping out before contributing the minimum number of trials (n = 1), experimenter error (n = 1), and color blindness (n = 1). Participants were Japanese children who were recruited through the database of the "Kyushu University Infant/child Scientist Project", whose caregivers had volunteered to participate in infant studies at Kyushu University. All participants gave informed consent prior to participating in this study. The experimental protocols were approved in advance of the study by the ethics committee of Kyushu University and were conducted in accordance with the Declaration of Helsinki. Participants received a gift voucher for their participation.

Coding and analyses

The coding and analysis were the same as in Experiment 3.3a. Additionally, we compared the data directly to that collected with Hungarian children in Experiment 3.3a.

Results

Just like their Hungarian peers, Japanese children preferred helpful and skilled partners above chance (Prosociality: mean = 0.703, 89% CI: [0.654, 0.751]; Skill: mean = 0.766, 89% CI: [0.722, 0.808]; see Figure 3.10). The comparison between preferences in the two trial types did not exclude chance, though only just barely: children had a somewhat higher rate of choosing a competent partner (mean = -0.063, 89% CI: [-0.129, 0.003]. The choice patterns also changed with age, such that older children were more likely to select helpful and skilled partners (Prosociality: mean = 0.818, 89% CI: [0.589, 1.052]; Skill: mean = 0.452, 89% CI: [0.232, 682]; see Figure D.4 in Appendix D.3).

The order effects mirrored closely those found with Hungarian children in Experiment 3.3a. There was also an effect of block (mean = 0.118, 89% CI: [0.051, 0.192]), similarly driven by those participants who encountered Prosociality trials in the first and Skill trials in the second block (mean = 0.208, 89% CI: [0.103, 0.325]; Skill-first participants: mean = 0.028, 89% CI: [-0.053, 0.112]). Here, too, there was no interaction with the trial type (mean = -0.023, 89% CI: [-0.31, -0.247]). In contrast to the Hungarian sample, Japanese children's choice of the better partner did not increase across the four trials within a block for either trial type (Prosociality: mean = 0.075, 89% CI: [-0.176, 0.323]; Skill: mean = 0.187, 89% CI: [-0.082, 0.458]).

Finally, compared to the Hungarian sample, the preference of Japanese children was relatively more pronounced in both Skill and Prosociality trials (Prosociality: mean = 0.081, 89% CI: [0.009, 0.153]; Skill: mean = 0.079, 89% CI: [0.014, 0.146]).

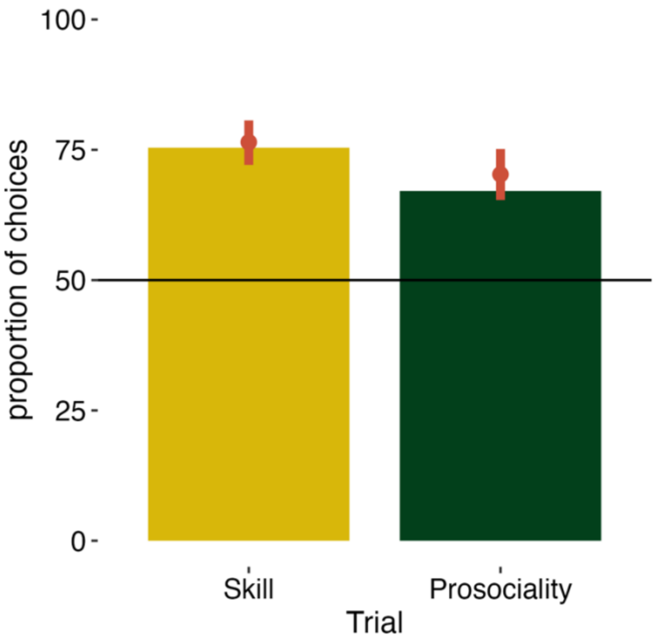

Figure 3.10. Bar plots representing the proportion of correct responses in Experiment 3.3b. Red error bars indicate the 89% CI of the parameter estimates, with dots indicating the means.

Discussion

The results with Japanese children not only replicated the finding from Hungary that participants chose collaboration partners who were relatively more likely to help them collect resources and were more adept at doing so, but in fact both of these preferences were stronger than in Hungarian children. Therefore, in addition to the stronger proclivity to select highly competent partners which we already found in Japanese compared to Hungarian adults in Experiments 3.2b and 3.2c, children between the two cultures also differed in how much they favored a helpful kobo over a selfish one.

We found that participants in both samples chose the better partner at a higher rate after getting practice with the game. In particular, children who observed and interacted with agents who differed in helpfulness were more likely to choose more competent partners in the subsequent block, despite having to attend to a different trait dimension and encountering novel resource types (colorful/brown coconuts). One way to interpret this result is that in the first block, children learned that agents in this game possess character traits that remain stable across contexts, and were able to generalize this knowledge when agents varied along another trait dimension. Participants who first participated in Skill trials did not show a similar improvement, which may suggest that children have different prior expectations about types of traits: The fact that agents have varying levels of skill may not warrant the inference that they can differ in prosociality.

Experiment 3.4

After finding that children, as predicted, preferred helpful and skilled cooperation partners in Experiment 3.3 (at least from a certain age), we returned to the question of whether they would prioritize prosociality or skill when the two traits are in contrast with each other. In Experiment 3.2, neither children nor Hungarian adults had shown such a preference, while Japanese adults chose a skilled but selfish over a helpful but incompetent partner. The chance result in children in Experiment 3.2a was difficult to interpret, as they also did not show a preference for a highly skilled partner in Skill trials, when agents only differed in this trait dimension. We therefore applied the changes made in Experiment 3.3, where we successfully elicited partner choice, to investigate children's behaviors in Contrast trials in a further experiment.

We made some further modifications to the experimental design. First, we tested only children between the ages of 7 and 10, who had chosen the better partner in Experiment 3.3. Second, Experiment 3.4 had a between-subject design, such that each participant only received four trials of the same type. This was because we wanted to present participants with the same number of trials of the same type as in Experiment 3.3, but wanted to compare participants' partner choices across Prosociality, Skill, as well as Contrast trials, and presenting four trials of all three trial types would have made the testing procedure too long. Third, in order to make the different trial types as similar as possible to one another, and to present the same introduction phase to all participants regardless of condition, we included all coconut types in all stimuli, as had been the case in Experiments 3.1 and 3.2. Finally, we shortened the introduction phase and removed the comprehension questions. Participants were recruited from the visitors of a local museum and we wanted to reduce the duration of testing sessions to help with recruitment; we assumed that since we were testing older children, who in our previous experiments did not struggle with understanding the instructions, we would not have to verify that they did so here.

Methods

The experiment was preregistered on the OSF (https://osf.io/9xt6y).

Stimuli

In this experiment, children were assigned to one of three conditions. In each condition, children received four test trials of a single type: Prosociality, Skill, or Contrast trials.

In order to make the trials in the three conditions as similar to each other as possible, all of them again included all types of coconuts (brown and colorful).

The trait parameters of the agents were the same as in Experiment 3.3. Agents in Contrast trials had the same prosociality and skill parameter values, crossed with each other.

Procedure

Data collection took place at the Natural History Museum Vienna (Austria). The experimental procedure was similar to that of Experiment 3.3, except that we made the instructions more concise and removed the comprehension questions.

The identity of the kobos was counterbalanced.

In order to be included, participants had to contribute a minimum of two trials. If a child chose the agent on the same side on all trials, this child's data was excluded from the analysis. A further exclusion criterion was interference from others, i.e., if bystanders made comments that related to the concepts under investigation or attempted to influence the child's choice.

Participants

One hundred twenty-one 7-10-year-old children participated in Experiment 3.4 (age: 9 years). We reached this sample size despite preregistering a sample of 120 participants due to an experimenter error in condition assignment. An additional 14 participants took part but had to be excluded from the sample for displaying a side bias (n = 8), technical failure (n = 3), parental interference (n = 1), color-blindness (n = 1), and a language barrier preventing the experimenter from conveying the instructions to the participant (n = 1). A higher number of children were excluded for a side bias because now there were only 4 test trials (in comparison to 8 in the previous experiments), but we retained the same conservative exclusion criterion.

Participants were children of varying nationalities who spoke German, English, or Hungarian and were recruited while on a visit to the Natural History Museum Vienna. Informed consent from caregivers and children was obtained before the experiment. The experiment received full ethical approval from the University's Psychological Research Ethics Board (PREBO) in Austria. Children received stickers for their participation.

Coding and analyses

The logging of responses was the same as in Experiments 3.2 and 3.3. Using a similar model as before, we tested whether children chose partners who were relatively more helpful and/or skilled. We also tested whether their preference for helpful partners in the Prosociality condition differed from that for fast partners in the Skill condition, and whether the tendencies to choose partners in these conditions differed from that in the Contrast condition. Again, we assessed possible order effects (i.e., whether responses differed as a function of the serial position of the trial), and whether children's age, gender, and the counterbalanced stimuli group had an effect.

Results

Participants chose the more skilled partner in the Skill condition (mean = 0.691, 89% CI: [0.616, 0.761]; see Figure 3.11). However, the rate of choosing a helpful partner in the Prosociality condition did not exclude chance (mean = 0.529, 89% CI: [0.45, 0.608]). When comparing the two conditions directly, the preference was stronger in the Skill than in the Prosociality condition (mean = 0.162, 89% CI: [0.052, 0.271]). In the Contrast condition, children chose the more skilled, selfish over a helpful, less skilled kobo above chance (mean = 0.595, 89% CI: [0.517, 0.673]). The preference for a highly skilled partner did not differ between the Skill and the Contrast conditions (mean = -0.065, 89% CI: [-0.177, 0.045]), while the preference for a helpful partner was higher in the Prosociality compared to the Contrast condition (mean = 0.123, 89% CI: [0.011, 0.233]). We found an effect of participants' age only for children in the Skill condition, such that older children were more likely than younger ones to prefer a competent agent (mean = 0.443, 89% CI: [0.129, 0.763]; see Figure D.6 in Appendix D.3).

Finally, we found that for children in the Prosociality condition, the proportion of choosing a helpful partner increased across the four test trials (mean = 0.457, 89% CI: [0.156, 0.767]), while in the other conditions the estimates for this parameter did not exclude chance (Skill: mean = 0.117, 89% CI: [-0.195, 0.437]; Contrast: mean = -0.159, 89% CI: [-0.459, 0.136]).

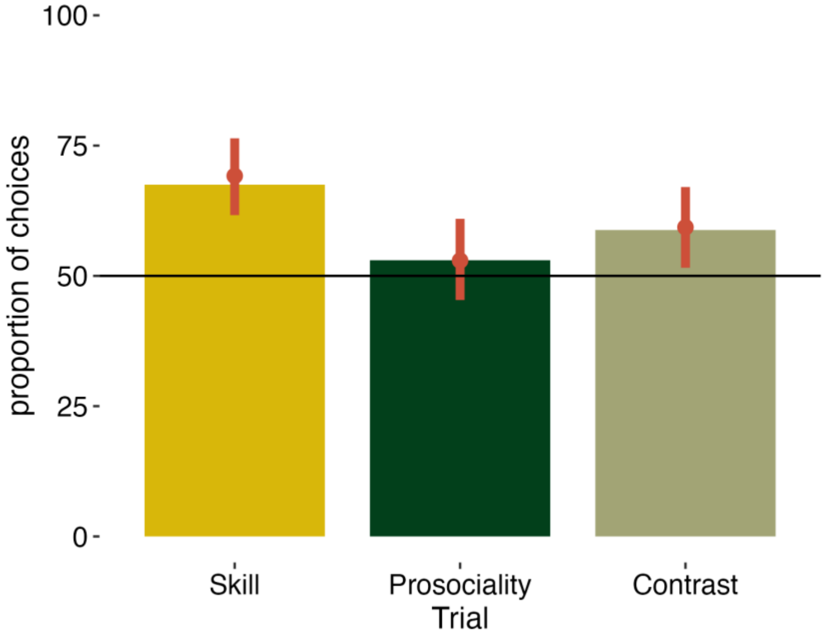

Figure 3.11. Bar plots representing the proportion of correct responses in Experiment 3.4. Red error bars indicate the 89% CI of the parameter estimates, with dots indicating the means.

Discussion

In Experiment 3.4, we found that children preferred to play with a more skilled agent, and even prioritized a partner's skill over prosociality when the two were in contrast. However, they did not select a relatively more helpful over a selfish agent as a partner. The findings thus differ from those in Experiment 3.2a, where children only showed a preference in Prosociality trials and were at chance in Skill and Contrast trials, and 3.3a, where children selected both helpful and skilled partners.

This pattern of results need not constitute a discrepancy. In Experiment 3.2a, as we discussed above, children might not have interpreted the task as we intended (e.g., they may have positively evaluated the helpful agent's behavior without ascribing a prosocial trait or engaging in genuine partner choice, or they may have pursued a competitive strategy). In contrast, in Experiment 3.3 and 3.4, children received more experience with the task, which may have helped them recognize that agents behaved consistently across trials, and that collaborating with certain partners yielded higher rewards for them.

If such an explanation is on the right track, why did children choose a helper above chance in Experiment 3.3 but not in Experiment 3.4? First, children's preference for helpful partners in Experiment 3.3 increased from 58.3% when Prosociality trials were presented in the first block to 66.1% when they came second, and in Experiment 3.4, this preference increased across the four test trials. Thus, if children had received even more practice with a second block of trials in Experiment 3.4, they might have also shown an above-chance preference for prosocial agents. Moreover, the task in Experiment 3.3 was simpler in one potentially crucial aspect: Here children only encountered proprietary colorful coconuts in Prosociality trials, whereas in Experiment 3.4, they had to learn about all coconut types from the beginning and apply this knowledge to attribute dispositions to the agents. For this reason, it is plausible to assume that it was easier for children to learn about an agent's skill, as every instance of hitting a coconut, regardless of type, provides evidence about how fast he can hit them, whereas inferring the prosociality of an agent requires interpreting a pattern of coconut-approach and -avoidance behaviors. Note, however, that children in Experiment 3.1, when asked explicitly about agents' behaviors, recognized which agent helped more, although for this task they also had to track all types of coconuts. This discrepancy in responses between Experiments 3.1 and 3.4 may be due to the fact that in Experiment 3.4 children had to go beyond interpreting past behavior and infer a character trait which would exert stable influence over agents' behaviors across settings and partners, and thus directly affect their own payoff. The fact that children's choices of a helpful partner increased across the four test trials supports the idea that such learning played a crucial role here.

General discussion

The aim of the present project was to study children's and adults' trait attribution and partner choice in a context where (1) prospective partners' traits had to be inferred from behavioral observations, (2) the selection of an interaction partner had tangible consequences on participants' payoff on the task, and (3) responses were not elicited through an experimenter prompt. For this purpose, we used the Co-Collectors game, which we had developed as a research tool suitable to meet these criteria. In the game, participants had to infer prosociality and skill from observing the behavior of agents in a novel setting (i.e., they had to understand what it means to be prosocial and skilled in the context of the game), conclude that these are enduring and stable traits, and use this information to select a partner for a subsequent cooperative interaction. This partner choice affected how well they did in the game (how many berries they collected) and, insofar as amassing resources in a game (or, as in Experiments 3.3 and 3.4, collecting stickers) is intrinsically rewarding, had real consequences. Moreover, participants' decisions of who to play with were not made in response to an explicit prompt by an experimenter, and we avoided explicitly referring to the concepts in question (e.g., helping, skill) or conveying value judgments, as we wanted to probe how they would choose partners when they had to rely solely on behavior observation and strategic consideration of which agent would generate more benefits for them. To address our research questions, we tested 5- to 10-year-old children and adults in two different cultural settings (Central Europe and East Asia).

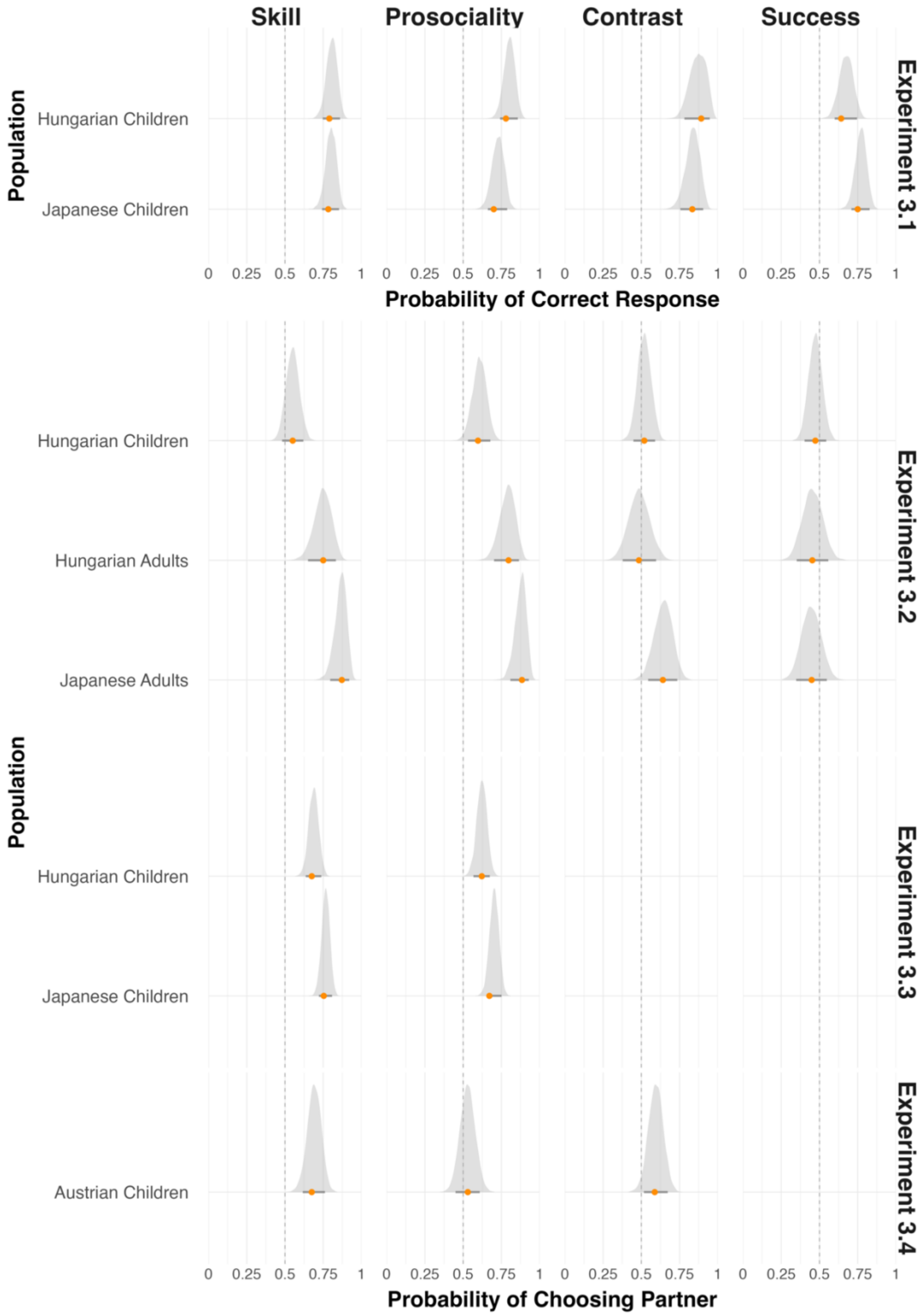

Figure 3.12. Overview of the results from Experiments 3.1-4 across the four trial types (Skill, Prosociality, Contrast, Success). Orange dots indicate the observed proportion of a response in a trial. Density plots indicate the distribution of parameter estimates generated by the Bayesian model, while horizontal lines mark the 89% credible intervals. On the *x*-axes are the proportion of correct responses (Experiment 3.1) or proportion of choices for one of two agents (Experiments 3.2-4).

We found that children across our age range were successful at identifying which agents were faster and helped more, and that they likely did not generate these responses by merely tracking differences in payoff, but actually attended to the behavior of the agents and its consequences in the game (Experiment 3.1). However, when provided with the same introduction and tested with the same stimuli in a task where they actually had to play the game and choose a partner, children did not seem to successfully apply this information: They preferred a more helpful partner slightly above chance, but not a more skilled one (Experiment 3.2a). One possible explanation for this pattern of results is that children interpreted the task as a competitive one and attempted to beat the confederate, and as prior research has shown, children's strategy and choices in a game can be affected by subtle changes in their motivation or goal (Rule et al., 2023). Another possibility is that children did not spontaneously explain the differences in behavior among agents in terms of stable character traits, or that they failed to use this information to strategically select partners. Instead, children may have preferred the prosocial character due to an affective tagging mechanism that leads to general valence-based evaluation (Dunfield et al., 2023).

In a modified procedure, where we simplified the task, gave children more practice, and highlighted the collaborative structure of the game, children succeeded: From around 7 years, they chose helpful and skilled partners (Experiment 3.3). In a between-subject design where children only received half the overall number of trials as before, children still had a preference for skilled partners, but did not select helpful partners above chance (Experiment 3.4). When children could choose between a helpful, but not very competent and a skilled, but selfish agent, they showed a preference for the latter (Experiment 3.4).

In Experiments 3.3 and 3.4, the opportunity to encounter agents differing along the same domain(s) on multiple successive trials provided information that could support a crucial inferential leap: namely, that agents always behave the same way in a third-party interaction as with the child herself; i.e., that behavioral variability between agents is caused by enduring, context-independent character traits. This inference may have only been carried out to a limited extent by children in Experiments 3.1 and 3.2 (though here too, children were more likely to choose the better partner on the second trial of a type they encountered), but allowed them to successfully select more helpful and skilled partners in Experiments 3.3 and 3.4. The fact that in Experiment 3.4 children had a stronger preference in Skill compared to Prosociality trials, and prioritized skill in Contrast trials, possibly indicates that children are more willing to infer that agents will behave consistently in domains that are less dependent on social factors (an idea also supported by the fact that children's preferences for a competent partner did not change across the four test trials in Experiment 3.4); or they may have valued the ability of a partner to provide benefits more than the willingness to do so. Interestingly, Experiment 3.3 showed that children performed better in the second compared to the first block they received, despite the fact that they had to attend to different domains of agent traits in the two blocks. This suggests that children—particularly those who initially had to attend to the domain of prosociality—generalized from one to the next block, having derived the overhypothesis that kobos' behaviors in the game are likely stable and generalizable.

We also tested samples of adults in Hungary (Experiment 3.2b) and Japan (Experiment 3.2c). They successfully chose helpful and skilled agents to play with, implying that they spontaneously interpreted the kobos' behaviors in terms of traits. However, their choices for better partners were not at ceiling, the task was thus not trivially easy for them. The result from adult participants is in line with the overall developmental trends we found, suggesting that children at the lower end of our age range (5 years) initially do not show robust preferences, but gradually come to show adult-like choice behaviors.

When comparing the data from the two different cultural contexts in which we tested, some interesting patterns emerge. First, Japanese children seemed to have paid relatively more attention than their Hungarian counterparts to payoff differences between agents that were driven by circumstance, rather than agents' behaviors. In Experiment 3.1b, when responding to experimenter questions about the stimuli, Japanese children were as accurate in tracking which kobo got more berries when this was a result of a skewed resource distribution as they were in identifying who was faster and who helped more.

Second, both Japanese adults and children showed a stronger preference for skilled agents: They preferred a high-skilled over a low-skilled partner more than the corresponding Hungarian samples (Experiments 3.2b and 3.2c, and 3.3a and 3.3b, resp.). Further, in Contrast trials (Experiment 3.2), Japanese adults picked a fast, selfish over a helpful, slow partner, unlike Hungarians, who did not prioritize either trait. Moreover, Japanese children's preference for a more helpful partner was also higher than that of the Hungarians'. It is possible that the increased attention to outcomes and payoff differences in Japanese participants that we found in Experiment 3.1 contributed to this divergence.

Overall, the present study highlights that researchers aiming to study social cognition in children should aim to develop methodologies that go beyond vignette stimuli and explicit experimenter prompts. Our results point to the possibility that there may be distinct mechanisms involved in action understanding, trait attribution, moral evaluation and partner choice, and that responses children give to a question posed by an experimenter needn't reflect their spontaneous social reasoning or behavior in a more realistic context with tangible consequences.

The series of experiments reported here leave many questions about the development of trait reasoning and partner choice psychology unanswered, some of which we believe the Co-Collectors application has the potential to address. One advantage of the application is the parametrization of the character traits that can be implemented therein. Thus, future research could test in a more fine-grained way the role these parameters play in observers' responses. For instance, it could be investigated at which parameter settings behavior differences are recognized, and at which trade-off between helpfulness and skill a preference for either emerges. Research could also probe further children's overhypotheses about traits, and how broadly they generalize across novel contexts and partners. Taking the application in a further direction, it could also be studied whether children themselves act more prosocially towards certain partners, whether they reciprocate altruistic behavior, or even engage in reputation management (that is, whether they would be strategically more helpful or invest more effort under conditions where they themselves might be chosen as partners). The Co-Collectors app can be used to address these and other research questions on trait reasoning, partner choice and other aspects of social cognition. It is freely available to researchers and, as we demonstrate with the present project, particularly well-suited for replications of experiments across different testing sites and international collaborations.

Summary, conclusions, outlook

The aim of this thesis was to explore how infants and children make sense of third-party helping interactions, the cognitive mechanisms this understanding relies on, and the inferences about agents' character traits children draw on the basis of observed helping events. We addressed these questions in 12 experiments, using a range of different methods—from recording looking times, gaze behavior, and manual choice, to collecting responses to verbal vignettes and partner choices in a custom research game—and testing children from 10 months up to 10 years. A common theme shared by these studies is that they address issues which may seem uncontroversial in light of extant literature, but which we argue have not been conclusively settled.

In Section 1, we asked whether infants would expect a goal-directed agent to choose a better goal option when multiple alternatives became available. We found no evidence for such a capacity in looking-time studies with 10-month-olds, who did not look longer when an agent chose a relatively smaller amount of goal objects (Experiment 1.1), or one of two identical-looking targets that was more effortful to reach (Experiment 1.2). However, when testing 14- to 16-month-olds in an eye-tracking design that posed a similar task as the latter experiment, we found that toddlers directed a higher proportion of looks to the less costly goal option, and this did not seem to be driven by low-level visual saliency. These results point to the possibility that representing different potential courses of action, and reasoning counterfactually to compare the utility of an observed behavior with that of non-chosen alternatives, may be challenging for young infants.

This project raises fundamental questions about the nature of infants' goal attributions; specifically, whether infants' concept of a goal-directed action entails an assumption of utility maximization, and whether their concept of "agent" entails "choice". In teleological reasoning (Gergely & Csibra, 2003), a goal—to naive observers—is a state of the world towards which an agent directs her efforts in an efficient manner. Consider a bustling market hall: Someone walks in the front door and goes straight to the cheese stand. You can infer that cheese is what she wanted to buy, unlike another shopper, who roams around aimlessly, browsing the treats on display. NUC goes beyond this: Here, a goal is an outcome that an agent acts towards because she anticipates the greatest utility from it, given the available options. As Jara-Ettinger and colleagues explain: "the naïve utility calculus expands on the teleological stance by explaining how agents select their goals" (Jara-Ettinger et al., 2016), and perhaps means, too (see Appendix A). Back to the market: You can tell that a visitor clearly has a preferred dairy purveyor because she walked past another cheesemonger next to the entrance. With NUC, the observer can represent the relevant choice set and consider not only endeavors pursued, but all the other eventualities an agent may choose or could have selected instead[13,14].

A pioneering study by Liu and colleagues (2017) indicated that NUC is present in the first year of life, showing that not only can infants infer the value of goals from the cost an agent incurred to reach them, but they expect the agent to approach the goal that is relatively more valuable. Our findings, in contrast, point to limitations in NUC at this age. This discrepancy cannot be resolved by the present work. However, it substantiates the possibility that comprehensive NUC, including counterfactual reasoning and mentalistic concepts (beliefs, desires), arises later than an initial, more limited capacity to reason about the efficiency of goal-directed action.

Section 2 pursued the question whether an early understanding of helping actions, demonstrated by studies such as the one by Hamlin et al. (2007), is undergirded by a hierarchical NUC-based concept. We first gave a detailed account of such a concept (H-NUC), according to which a helping action has the goal of increasing or maximizing the utility another agent obtains in a goal-directed action. With H-NUC, observers should ascribe the goal of helping whenever this outcome (i.e., another agent incurs relatively lower costs, or reaps higher rewards as a result of the intervention) best explains an agent's behavior. We argued that such a concept is what adults have in mind when thinking of helping, and investigated its emergence in development. Just as teleological reasoning or NUC are invaluable tools for naive learners, H-NUC may support the ascription of other-directed goals in novel contexts and even for unfamiliar actions.

In a set of three looking-time experiments with infants (Experiments 2.2.1-3), we found no evidence that twelve-month-olds apply H-NUC to ascribe a goal of helping to an agent in a third-party interaction. Three-year-olds, on the other hand, identified the agent who contributed to another agent's utility as the one who helped, suggesting that they acquired a version of the mature concept (Experiment 2.3). Surprisingly, they did not themselves help by maximizing a Helpee's utility, which could point to distinct features of their understanding of helping. For instance, for children, acting helpfully may not necessarily mean generating the best possible option for the Helpee.

Alternatively, even at three years of age, children may struggle in certain contexts to identify what the best possible option is, i.e., to consider and compare alternatives.

[13] In discussing inverse planning approaches to action understanding (Baker et al., 2009; Ullman & Tenenbaum, 2020), Spelke writes: "When we see a person who engages in a specific series of actions in an environment that affords many other actions, we may infer both her action plans and her goals [...] by imagining actions that she could have taken but didn't, and by imagining how her actions would vary if her goal changed [...]" (Spelke, 2022).

[14] Note that in Bayesian models of NUC reasoning, the available alternative options and their respective utilities are explicitly represented in the hypothesis space (e.g., Baker et al., 2017; Liu et al., 2017).

Preschoolers as well as infants may have failed here because they don't possess, or didn't consistently apply, the concept of choice. We pointed out in Section 2 that counterfactual reasoning is required to compare the Helpee's actual utility—which is causally linked to the Helper's intervention—with how the Helpee would have fared had she not been helped. The action options available to the Helpee thus depend on the behavior of the Helper: For instance, she cannot take a shortcut through a door if that door is closed and she is unable to open it. These two alternatives (the most efficient action with and without help, respectively) have to be explicitly represented as such to recognize the impact of helping.

As for the Helper, there are two types of choice that may be relevant for assessing her behavior. First, corresponding to the design of Experiment 2.2.2 and the "help vs. irrelevant" trial in Experiment 2.3, an agent with the goal of helping faces a choice of whether or not to intervene, and (according to H-NUC) should rationally only act if doing so increases the Helpee's utility. This can be derived from the principle of rationality: Because the Helpee's utility increase is embedded as the reward term in the Helper's utility function, the Helper's cost of acting is only offset if this value reaches a certain positive threshold. Second, as implemented in Experiment 2.2.1 and Block 1 of Experiment 2.3, if there are multiple different options for intervening (for instance, opening a close vs. a far-away door), the Helper can choose whether to help by selecting the option that will maximize the Helpee's utility, or a sub-optimal alternative. In this case, too, a rational Helper should choose the former option (provided that the cost of both actions is equal): By doing what is best for the Helpee, she can maximize her own utility. Thus, the rationality of helping can be assessed by evaluating both the Helper's action cost, and comparing the consequences for the Helpee as a result of (1) the Helper's action vs. inaction, and (2) different means to help. If young children don't apply a concept of choice, they might not be able to perform either comparison.

It may be relevant here that most studies evidencing sophisticated NUC have been with children 4 years and older (Aboody et al., 2021, 2022; Ahl et al., 2023; Bridgers, Jara-Ettinger, et al., 2020; Gerdin & Dunham, 2022; Huh & Friedman, 2019; Jara-Ettinger, Gweon, et al., 2015; Jara-Ettinger et al., 2020; Jara-Ettinger, Tenenbaum, et al., 2015). This discrepancy may parallel that found in other domains between infants' early competency in implicit tasks, and much later success in explicit tasks that toddlers and younger preschoolers fail (modal reasoning: Leahy & Carey, 2020; theory of mind: Rakoczy, 2022).

Finally, Section 3 of this thesis was concerned with the inferences children draw from action observation; specifically, whether they interpret helping behaviors as indicative of a situation-independent prosocial character trait. It is often assumed that people spontaneously and overzealously attribute behaviors to underlying dispositions or traits of agents ("fundamental attribution error" or "correspondence bias"; Gawronski, 2004;

Ross, 1977). Developmental research has often found that until middle childhood, children don't show similar tendencies (Heyman, 2009). On the other hand, research with younger participants—particularly using the manual choice measure—has been interpreted as showing that infants attribute pro- or antisocial traits to agents upon observing their actions in a third-party context, and therefore preferentially interact with "good guys".

We addressed this gap with the experiments discussed in this section. We developed a tablet-based research game in which participants observed third-party interactions of agents who varied in prosociality and skill. Similar as in designs probing infants' action understanding, participants had to make sense of these behaviors without being provided with a trait label by the experimenter. We used partner choice as an indirect measure of trait attribution, hypothesizing that if children ascribed a skill or prosocial disposition that agents possess across contexts and partners, they should choose to play with one whose behavior would generate more rewards for them. In Experiments 3.1-2a, we found that 5- to 10-year old children were adept at recognizing which agent was faster, and which agent helped more than another. Yet, this did not translate to choosing a more advantageous partner. In a modified procedure, where we simplified the task and highlighted that it was not competitive, and where children had more opportunities to learn about the cross-trial consistency of agents' behaviors, their responses overall indicated that they recognized who would be a better partner (3.3-4). The results suggest that children were faster or more ready to attribute competence, while their preference for a prosocial partner seemed to depend more on experience—children may have learned while playing the game that agents who helped a third party would also help them.

It should be noted that to correctly interpret the agents' behaviors as helping, children could not rely on prototypical features of this action, but had to recruit H-NUC. In the Co-Collectors game, helping consisted in an agent directing his efforts towards the coconuts which contained the partner's berries. Thus, a Helper (1) reduced the action cost of the Helpee (because the latter had to deliver fewer hits to crack a coconut), and (2) increased the Helpee's rewards (by contributing to a relatively higher amount of berries than the Helpee could have gathered by herself).

With the project from Section 3, we argued that goal ascription is not the same as trait inference (though the two are naturally not independent of one another; e.g. it has been argued that they are at different levels in a hierarchical system of social cognition: Malle & Holbrook, 2012; Read et al., 1990; Reeder, 2009; Westra, 2018): Children may understand who helped, but not prefer a Helper. They might thus not commit the fundamental attribution error of defaulting on trait inference from behavior observation. This has potential implications for interpreting research with infants. It is often assumed that their purported preference for Helpers is indicative of an early-emerging moral sense, rudimentary and less sophisticated than mature moral cognition, but persistent across development (Baillargeon et al., 2015; Woo et al., 2022). This account relies on the assumptions that young infants (1) interpret helping interactions similarly as adults, and (2) on this basis, like (Western) adults, ascribe traits to agents who help. Our results question the validity of the moral continuity account, and invite revisiting its theoretical foundations. Instead of ascribing cooperative character traits early on, infants may initially conceive of social interactions in terms of the relations of the agents engaged in them (Mascaro & Csibra, 2012; Pomiechowska et al., 2022; Powell, 2022; Tatone et al., 2015).

Taken together, the work described in this thesis warrants some methodological conclusions. First, we emphasize that researchers studying the origins of naive psychology and sociomoral reasoning should ensure to spell out as precisely as possible the concepts under investigation, the theoretical commitments they entail, and the assumptions underlying the methods used to probe them. Second, our findings highlight the importance of studying how children interpret observed actions without verbal prompts and vignettes. In social cognition research with infants, this approach is used by necessity; with older children, it is less common. As we have argued, linguistic scaffolding may introduce concepts and categorizations that children would otherwise not have spontaneously applied, or may induce a social desirability bias. On the other hand, examining children's interpretation of a term can be a tool to investigate the corresponding concept they recruit (see Chapter 2.3). Finally, we want to underscore the value of conducting (conceptual and direct) replications of studies to assess their effects, despite the fact that and especially because data collection in developmental research is effortful and time-consuming, and this field often relies on conclusions drawn from small sample sizes.

To conclude: This dissertation has discussed fundamental questions concerning the nature, origins, and development of young children's attributions of goals and traits upon observing others' actions, particularly helping. We have asked whether infants possess a concept of choice and apply it to compare different possible goal alternatives an agent faces; how young children represent helping actions, and whether they rely on similar mechanisms to ascribe social as non-social goals; and whether children tend to spontaneously attribute enduring character traits to agents upon observing their actions, and choose interaction partners on this basis. Our research has highlighted that some of the processes purported to be involved in early action understanding (e.g., NUC, counterfactual reasoning, hierarchical representations) demand further scrutiny. We hope that our work and the approaches reported here constitute a step towards clarifying these issues, which has potential implications for theories of naive psychology as well as cognitive development more broadly.

References

Aboody, R., Huey, H., & Jara-Ettinger, J. (2022). Preschoolers decide who is knowledgeable, who to inform, and who to trust via a causal understanding of how knowledge relates to action. *Cognition*, *228*, 105212. https://doi.org/10.1016/j.cognition.2022.105212

Aboody, R., Zhou, C., & Jara-Ettinger, J. (2021). In Pursuit of Knowledge: Preschoolers Expect Agents to Weigh Information Gain and Information Cost When Deciding Whether to Explore. *Child Development*, *92*(5), 1919–1931. https://doi.org/10.1111/cdev.13557

Adam, M., & Elsner, B. (2020). The impact of salient action effects on 6-, 7-, and 11-month-olds' goal-predictive gaze shifts for a human grasping action. *PLoS ONE*, *15*(10), e0240165. https://doi.org/10.1371/journal.pone.0240165

Adam, M., Reitenbach, I., Papenmeier, F., Gredebäck, G., Elsner, C., & Elsner, B. (2016). Goal saliency boosts infants' action prediction for human manual actions, but not for mechanical claws. *Infant Behavior and Development*, *44*, 29–37. https://doi.org/10.1016/j.infbeh.2016.05.001

Ahl, R. E., Cook, E., & McAuliffe, K. (2023). Having less means wanting more: Children hold an intuitive economic theory of diminishing marginal utility. *Cognition*, *234*, 105367. https://doi.org/10.1016/j.cognition.2023.105367

Alexander, R. D. (1987). *The biology of moral systems* (pp. xx, 301). Aldine de Gruyter.

Allen, M., Perry, C., & Kaufman, J. (2018). Toddlers prefer to help familiar people. *Journal of Experimental Child Psychology*, *174*, 90–102. https://doi.org/10.1016/j.jecp.2018.05.009

Aloise, P. A. (1993). Trait Confirmation and Disconfirmation: The Development of Attribution Biases. *Journal of Experimental Child Psychology*, *55*(2), 177–193. https://doi.org/10.1006/jecp.1993.1010

Alvarez, J. M., Ruble, D. N., & Bolger, N. (2001). Trait Understanding or Evaluative Reasoning? An Analysis of Children's Behavioral Predictions. *Child Development*, *72*(5), 1409–1425. https://doi.org/10.1111/1467-8624.00356

Ames, D. L., Fiske, S. T., & Todorov, A. T. (2011). Impression formation: A focus on others' intents. In *The Oxford handbook of social neuroscience* (pp. 419–433). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780195342161.001.0001

Anderson, J. R., Bucher, B., Chijiiwa, H., Kuroshima, H., Takimoto, A., & Fujita, K. (2017). Third-party social evaluations of humans by monkeys and dogs. *Neuroscience & Biobehavioral Reviews*, *82*, 95–109. https://doi.org/10.1016/j.neubiorev.2017.01.003

Apicella, C. L., Marlowe, F. W., Fowler, J. H., & Christakis, N. A. (2012). Social networks and cooperation in hunter-gatherers. *Nature*, *481*(7382), 497–501. https://doi.org/10.1038/nature10736

Augusti, E.-M., Melinder, A., & Gredebäck, G. (2010). Look Who's Talking: Pre-Verbal Infants' Perception of Face-to-Face and Back-to-Back Social Interactions. *Frontiers in Psychology*, *1*, 161.

Bach, P., & Schenke, K. C. (2017). Predictive social perception: Towards a unifying framework from action observation to person knowledge. *Social and Personality Psychology Compass*, *11*(7), e12312. https://doi.org/10.1111/spc3.12312

Baer, C., & Odic, D. (2022). Mini managers: Children strategically divide cognitive labor among collaborators, but with a self-serving bias. *Child Development*, *93*(2), 437–450. https://doi.org/10.1111/cdev.13692

Baillargeon, R. (1987). Object permanence in 3½- and 4½-month-old infants. *Developmental Psychology*, *23*(5), 655–664. https://doi.org/10.1037/0012-1649.23.5.655

Baillargeon, R., Scott, R. M., He, Z., Sloane, S., Setoh, P., Jin, K.-S., Wu, D., & Bian, L. (2015). Psychological and Sociomoral Reasoning in Infancy. In *M. Mikulincer, P. R. Shaver, E. Borgida, & J. A. Bargh (Eds.), APA handbook of personality and social psychology, Vol. 1. Attitudes and social cognition: Vol. Vol. 1. Attitudes and social cognition* (pp. 79–150). American Psychological Association.

Baker, C. L., Goodman, N. D., & Tenenbaum, J. B. (2008). Theory-based Social Goal Inference. *Proceedings of the 30th Annual Meeting of the Cognitive Science Society*.

Baker, C. L., Saxe, R., & Tenenbaum, J. B. (2009). Action understanding as inverse planning. *Cognition*, *113*(3), 329–349. https://doi.org/10.1016/J.COGNITION.2009.07.005

Baker, C.L., Jara-Ettinger, J., Saxe, R. & Tenenbaum, J. B. (2017). Rational quantitative attribution of beliefs, desires and percepts in human mentalizing. *Nature Human Behavior, 1*, 0064. https://doi.org/10.1038/s41562-017-0064

Barclay, P. (2013). Strategies for cooperation in biological markets, especially for humans. *Evolution and Human Behavior*, *34*(3), 164–175. https://doi.org/10.1016/j.evolhumbehav.2013.02.002

Barclay, P. (2016). Biological markets and the effects of partner choice on cooperation and friendship. *Current Opinion in Psychology*, *7*, 33–38. https://doi.org/10.1016/j.copsyc.2015.07.012

Barclay, P., & Willer, R. (2007). Partner choice creates competitive altruism in humans. *Proceedings of the Royal Society B*, *247*, 749–753. https://doi.org/10.1098/rspb.2006.0209

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*, 1–48. https://doi.org/10.18637/jss.v067.i01

Baumard, N., André, J.-B., & Sperber, D. (2013). A mutualistic approach to morality: The evolution of fairness by partner choice. *Behavioral and Brain Sciences*, *36*(01), 59–78. https://doi.org/10.1017/S0140525X11002202

Bear, A., Bensinger, S., Jara-Ettinger, J., Knobe, J., & Cushman, F. (2020). What comes to mind? *Cognition*, *194*, 104057. https://doi.org/10.1016/j.cognition.2019.104057

Beck, S. R., Robinson, E. J., Carroll, D. J., & Apperly, I. A. (2006). Children's Thinking About Counterfactuals and Future Hypotheticals as Possibilities. *Child Development*, *77*(2), 413–426. https://doi.org/10.1111/j.1467-8624.2006.00879.x

Begus, K., Curioni, A., Knoblich, G., & Gergely, G. (2020). Infants understand collaboration: Neural evidence for 9-month-olds' attribution of shared goals to coordinated joint actions. *Social Neuroscience*, *15*(6), 655–667. https://doi.org/10.1080/17470919.2020.1847730

Behne, T., Carpenter, M., Call, J., & Tomasello, M. (2005). Unwilling Versus Unable: Infants' Understanding of Intentional Action. *Developmental Psychology*, *41*(2), 328–337. https://doi.org/10.1037/0012-1649.41.2.328

Beier, J. S., & Spelke, E. S. (2012). Infants' Developing Understanding of Social Gaze. *Child Development*, *83*(2), 486–496. https://doi.org/10.1111/j.1467-8624.2011.01702.x

Bennett-Pierre, G., Asaba, M., & Gweon, H. (2018). Preschoolers consider expected task difficulty to decide what to do & whom to help. *Proceedings of the 40th Annual Meeting of the Cognitive Science Society*, 1359–1364.

Benton, D. T., & Lapan, C. (2022). Moral masters or moral apprentices? A connectionist account of sociomoral evaluation in preverbal infants. *Cognitive Development*, *62*, 101164. https://doi.org/10.1016/j.cogdev.2022.101164

Biro, S. (2013). The role of the efficiency of novel actions in infants' goal anticipation. *Journal of Experimental Child Psychology*, *116*(2), 415–427. https://doi.org/10.1016/j.jecp.2012.09.011

Biro, S., Verschoor, S., & Coenen, L. (2011). Evidence for a unitary goal concept in 12-month-old infants. *Developmental Science*, *14*(6), 1255–1260. https://doi.org/10.1111/j.1467-7687.2011.01042.x

Bliege Bird, R., & Power, E. A. (2015). Prosocial signaling and cooperation among Martu hunters. *Evolution and Human Behavior*, *36*(5), 389–397. https://doi.org/10.1016/j.evolhumbehav.2015.02.003

Boseovski, J. J., & Lapan, C. (2021). Traits or circumstances? Children's explanations of positive and negative behavioral outcomes. *Cognitive Development*, *58*, 101019. https://doi.org/10.1016/j.cogdev.2021.101019

Boseovski, J. J., & Lee, K. (2006). Children's use of frequency information for trait categorization and behavioral prediction. *Developmental Psychology*, *42*(3), 500–513. https://doi.org/10.1037/0012-1649.42.3.500

Boyd, R., & Richerson, P. J. (1992). Punishment allows the evolution of cooperation (or anything else) in sizable groups. *Ethology and Sociobiology*, *13*(3), 171–195. https://doi.org/10.1016/0162-3095(92)90032-Y

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*, 433–436.

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

Bridgers, S., Jara-Ettinger, J., & Gweon, H. (2020). Young children consider the expected utility of others' learning to decide what to teach. *Nature Human Behaviour*, *4*(2), 144–152. https://doi.org/10.1038/s41562-019-0748-6

Bridgers, S., Yang, C., Gerstenberg, T., & Gweon, H. (2020). Whom will Granny thank? Thinking about what could have been informs children's inferences about relative helpfulness. *Proceedings of the 42nd Annual Meeting of the Cognitive Science Society*.

Bshary, R., & Noë, R. (2003). The Ubiquitous Influence of Partner Choice on the Dynamics of Cleaner Fish – Client Reef Fish Interactions. In *Genetic and Cultural Evolution of Cooperation*. MIT Press.

Buresh, J. S., & Woodward, A. L. (2007). Infants track action goals within and across agents. *Cognition*, *104*(2), 287–314. https://doi.org/10.1016/j.cognition.2006.07.001

Burns, M., & Sommerville, J. (2014). “I pick you”: The impact of fairness and race on infants’ selection of social partners. *Frontiers in Psychology*, *5*, 93.

Cain, K. M., Heyman, G. D., & Walker, M. E. (1997). Preschoolers’ Ability to Make Dispositional Predictions Within and Across Domain. *Social Development*, *6*(1), 53–75. https://doi.org/10.1111/j.1467-9507.1997.tb00094.x

Cannon, E. N., & Woodward, A. L. (2012). Infants generate goal-based action predictions. *Developmental Science*, *15*(2), 292–298. https://doi.org/10.1111/j.1467-7687.2011.01127.x

Carey, S. (2009). *The Origin of Concepts*. Oxford University Press.

Cesana-Arlotti, N., Kovács, Á. M., & Téglás, E. (2020). Infants recruit logic to learn about the social world. *Nature Communications*, *11*(1), 5999. https://doi.org/10.1038/s41467-020-19734-5

Cesana-Arlotti, N., Martín, A., Téglás, E., Vorobyova, L., Cetnarski, R., & Bonatti, L. L. (2018). Precursors of logical reasoning in preverbal human infants. *Science*, *359*(6381), 1263–1266. https://doi.org/10.1126/science.aao3539

Cesana-Arlotti, N., Varga, B., & Téglás, E. (2022). The pupillometry of the possible: An investigation of infants’ representation of alternative possibilities. *Philosophical Transactions of the Royal Society B*, *377*(1866), 20210343. https://doi.org/10.1098/rstb.2021.0343

Chae, J. J. K., & Song, H. (2018). Negativity bias in infants’ expectations about agents’ dispositions. *British Journal of Developmental Psychology*, *36*(4), 620–633. https://doi.org/10.1111/bjdp.12246

Chen, E. E., Corriveau, K. H., & Harris, P. L. (2016). Person Perception in Young Children Across Two Cultures. *Journal of Cognition and Development*, *17*(3), 447–467. https://doi.org/10.1080/15248372.2015.1068778

Chiang, Y.-S. (2010). Self-interested partner selection can lead to the emergence of fairness. *Evolution and Human Behavior*, *31*(4), 265–270. https://doi.org/10.1016/j.evolhumbehav.2010.03.003

Chijiiwa, H., Kuroshima, H., Hori, Y., Anderson, J. R., & Fujita, K. (2015). Dogs avoid people who behave negatively to their owner: Third-party affective evaluation. *Animal Behaviour*, *106*, 123–127. https://doi.org/10.1016/j.anbehav.2015.05.018

Chijiiwa, H., Takagi, S., Arahori, M., Anderson, J. R., Fujita, K., & Kuroshima, H. (2021). Cats (Felis catus) Show No Avoidance of People Who Behave Negatively to Their Owner. *Animal Behavior and Cognition*, *8*(1), 23–35. https://doi.org/10.26451/abc.08.01.03.2021

Choi, I., Nisbett, R. E., & Norenzayan, A. (1999). Causal attribution across cultures: Variation and universality. *Psychological Bulletin*, *125*(1), 47–63. https://doi.org/10.1037/0033-2909.125.1.47

Choi, Y., & Luo, Y. (2023). Understanding preferences in infancy. *WIREs Cognitive Science*, *14*(4), e1643. https://doi.org/10.1002/wcs.1643

Choi, Y., Luo, Y., & Baillargeon, R. (2022). Can 5-month-old infants consider the perspective of a novel eyeless agent? New evidence for early mentalistic reasoning. *Child Development*, *93*(2), 571–581. https://doi.org/10.1111/cdev.13707

Choi, Y., Mou, Y., & Luo, Y. (2018). How do 3-month-old infants attribute preferences to a human agent? *Journal of Experimental Child Psychology*, *172*, 96–106. https://doi.org/10.1016/j.jecp.2018.03.004

Cohen, J. D., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: An interactive graphic system for designing and controlling experiments in the psychology laboratory using Macintosh computers. *Behavior Research Methods, Instruments & Computers*, *25*(2), 257–271. https://doi.org/10.3758/BF03204507

Colaizzi, J. M. (2016). *Empathy and prosocial behaviors in infancy* [Oklahoma State University]. https://shareok.org/bitstream/handle/11244/48793/Colaizzi_okstate_0664D_14781.pdf?sequence=1

Cortes Barragan, R., & Dweck, C. S. (2014). Rethinking natural altruism: Simple reciprocal interactions trigger children's benevolence. *Proceedings of the National Academy of Sciences*, *111*(48), 17071–17074. https://doi.org/10.1073/pnas.1419408111

Cottrell, C. A., Neuberg, S. L., & Li, N. P. (2007). What do people desire in others? A sociofunctional perspective on the importance of different valued characteristics. *Journal of Personality and Social Psychology*, *92*(2), 208–231. https://doi.org/10.1037/0022-3514.92.2.208

Cowell, J. M., & Decety, J. (2015). Precursors to morality in development as a complex interplay between neural, socioenvironmental, and behavioral facets. *Proceedings of the National Academy of Sciences*, *112*(41), 12657–12662. https://doi.org/10.1073/pnas.1508832112

Csibra, G. (2008a). Action mirroring and action interpretation: An alternative account. In *Sensorimotor Foundations of Higher Cognition: Attention and Performance XXII* (pp. 435–459). Oxford University Press. http://doi.org/10.1093/acprof:oso/9780199231447.003.0020

Csibra, G. (2008b). 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/10.1207/s15516709cog2701_4

Csibra, G., & Gergely, G. (2007). "Obsessed with goals": Functions and mechanisms of teleological interpretation of actions in humans. *Acta Psychologica*, *124*(1), 60–78. https://doi.org/10.1016/j.actpsy.2006.09.007

Csibra, G., Gergely, G., Bíró, S., Koós, 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), 521–536. https://doi.org/10.1037/dev0000083

Dahl, A. (2020). Early Developments in Acts and Evaluations of Helping: When and How to Help? In *The Oxford Handbook of Moral Development: An Interdisciplinary Perspective* (pp. 288–305). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190676049.013.16

Dahl, A., Gross, R. L., & Siefert, C. (2020). Young children's judgments and reasoning about prosocial acts: Impermissible, suberogatory, obligatory, or supererogatory? *Cognitive Development*, *55*, 100908. https://doi.org/10.1016/J.COGDEV.2020.100908

Dahl, A., & Paulus, M. (2019). From Interest to Obligation: The Gradual Development of Human Altruism. *Child Development Perspectives*, *13*(1), 10–14. https://doi.org/10.1111/cdep.12298

Davidson, D. (1963). Actions, Reasons, and Causes. *The Journal of Philosophy*, *60*(23), 685–700. https://doi.org/10.2307/2023177

de Cooke, P. A. (1992). Children's understanding of indebtedness as a feature of reciprocal help exchanges between peers. *Developmental Psychology*, *28*(5), 948–954. https://doi.org/10.1037/0012-1649.28.5.948

Delton, A. W., & Robertson, T. E. (2012). The social cognition of social foraging: Partner selection by underlying valuation. *Evolution and Human Behavior*, *33*(6), 715–725. https://doi.org/10.1016/j.evolhumbehav.2012.05.007

Delton, A. W., & Robertson, T. E. (2016). How the mind makes welfare tradeoffs: Evolution, computation, and emotion. *Current Opinion in Psychology*, *7*, 12–16. https://doi.org/10.1016/j.copsyc.2015.06.006

Dennett, D. C. (1987). *The intentional stance* (pp. xi, 388). The MIT Press.

DesChamps, T. D., Eason, A. E., & Sommerville, J. A. (2016). Infants Associate Praise and Admonishment with Fair and Unfair Individuals. *Infancy*, *21*(4), 478–504. https://doi.org/10.1111/infa.12117

Dhaliwal, N. A., Martin, J. W., Barclay, P., & Young, L. L. (2022). Signaling benefits of partner choice decisions. *Journal of Experimental Psychology: General*, *151*(6), 1446–1472. https://doi.org/10.1037/xge0001137

Droege, K., & Stipek, D. (1993). Children's use of dispositions to predict classmates' behavior. *Developmental Psychology*, *29*(4), 646.

Duh, S., & Wang, S. (2019). Infants Detect Patterns of Choices Despite Counter Evidence, but Timing of Inconsistency Matters. *Journal of Cognition and Development*, *20*(1), 96–106. https://doi.org/10.1080/15248372.2018.1528976

Dunfield, K. A., Isler, L., Chang, X. M., Terrizzi, B., & Beier, J. (2023). Helpers or halos: Examining the evaluative mechanisms underlying selective prosociality. *Royal Society Open Science*, *10*(4), 221188. https://doi.org/10.1098/rsos.221188

Eason, A. E., Doctor, D., Chang, E., Kushnir, T., & Sommerville, J. A. (2018). The choice is yours: Infants' expectations about an agent's future behavior based on taking and receiving actions. *Developmental Psychology*, *54*(5), 829–841. https://doi.org/10.1037/dev0000482

Egyed, K., Király, I., & Gergely, G. (2013). Communicating Shared Knowledge in Infancy. *Psychological Science*, *24*(7), 1348–1353. https://doi.org/10.1177/0956797612471952

Eisenberg-Berg, N., & Neal, C. (1981). Children's Moral Reasoning about Self and Others: Effects of Identity of the Story Character and Cost of Helping. *Personality and Social Psychology Bulletin*, *7*(1), 17–23. https://doi.org/10.1177/014616728171004

Eisenbruch, A. B., & Krasnow, M. M. (2022). Why Warmth Matters More Than Competence: A New Evolutionary Approach. *Perspectives on Psychological Science*, *17*(6), 1604–1623. https://doi.org/10.1177/17456916211071087

Eisenbruch, A. B., & Roney, J. R. (2017). The Skillful and the Stingy: Partner Choice Decisions and Fairness Intuitions Suggest Human Adaptation for a Biological Market of Cooperators. *Evolutionary Psychological Science*, *3*(4), 364–378. https://doi.org/10.1007/s40806-017-0107-7

Engelmann, J. M., Herrmann, E., & Tomasello, M. (2012). Five-Year Olds, but Not Chimpanzees, Attempt to Manage Their Reputations. *PLOS ONE*, *7*(10), e48433. https://doi.org/10.1371/journal.pone.0048433

Falck-Ytter, T., Gredebäck, G., & von Hofsten, C. (2006). Infants predict other people's action goals. *Nature Neuroscience*, *9*(7), 878–879. https://doi.org/10.1038/nn1729

Fawcett, C., & Gredebäck, G. (2013). Infants use social context to bind actions into a collaborative sequence. *Developmental Science*, *16*(6), 841–849. https://doi.org/10.1111/desc.12074

Fawcett, C., & Gredebäck, G. (2015). Eighteen-Month-Olds, but not 14-Month-Olds, Use Social Context to Bind Action Sequences. *Infancy*, *20*(1), 115–125. https://doi.org/10.1111/infa.12062

Fawcett, C., & Liszkowski, U. (2012). Infants Anticipate Others' Social Preferences: Social Preference Anticipation. *Infant and Child Development*, *21*(3), 239–249. https://doi.org/10.1002/icd.739

Feigenson, L., Carey, S., & Hauser, M. (2002). The Representations Underlying Infants' Choice of More: Object Files Versus Analog Magnitudes. *Psychological Science*, *13*(2), 150–156. https://doi.org/10.1111/1467-9280.00427

Feiman, R., Carey, S., & Cushman, F. (2015). Infants' representations of others' goals: Representing approach over avoidance. *Cognition*, *136*, 204–214. https://doi.org/10.1016/j.cognition.2014.10.007

Feldman, N. S., & Ruble, D. N. (1988). The Effect of Personal Relevance on Psychological Inference: A Developmental Analysis. *Child Development*, *59*(5), 1339–1352. https://doi.org/10.2307/1130496

Ferguson, T. J., Olthof, T., Luiten, A., & Rule, B. G. (1984). Children's Use of Observed Behavioral Frequency versus Behavioral Covariation in Ascribing Dispositions to

Others. *Child Development*, *55*(6), 2094–2105. https://doi.org/10.2307/1129782

Fiske, S. T., Cuddy, A. J. C., & Glick, P. (2007). Universal dimensions of social cognition: Warmth and competence. *Trends in Cognitive Sciences*, *11*(2), 77–83. https://doi.org/10.1016/j.tics.2006.11.005

Fitneva, S. A., & Dunfield, K. A. (2010). Selective information seeking after a single encounter. *Developmental Psychology*, *46*(5), 1380–1384. https://doi.org/10.1037/a0019818

Fortin, C. (2019). *L'évaluation sociale et l'apprentissage en contexte social des nourrissons* [Université de Montréal]. https://papyrus.bib.umontreal.ca/xmlui/bitstream/handle/1866/23841/Fortin_Camille_2019_Essai_Doctoral.pdf?sequence=3&isAllowed=y

Franchin, L., Savazzi, F., Neira-Gutierrez, I. C., & Surian, L. (2019). Toddlers map the word 'good' to helping agents, but not to fair distributors. *Journal of Child Language*, *46*(1), 98–110. https://doi.org/10.1017/S0305000918000351

Ganglmayer, K., Attig, M., Daum, M. M., & Paulus, M. (2019). Infants' perception of goal-directed actions: A multi-lab replication reveals that infants anticipate paths and not goals. *Infant Behavior and Development*, *57*, 101340. https://doi.org/10.1016/j.infbeh.2019.101340

Gao, T., Newman, G. E., & Scholl, B. J. (2009). The psychophysics of chasing: A case study in the perception of animacy. *Cognitive Psychology*, *59*(2), 154–179. https://doi.org/10.1016/j.cogpsych.2009.03.001

Gawronski, B. (2004). Theory-based bias correction in dispositional inference: The fundamental attribution error is dead, long live the correspondence bias. *European Review of Social Psychology*, *15*(1), 183–217. https://doi.org/10.1080/10463280440000026

Geraci, A., & Surian, L. (2011). The developmental roots of fairness: Infants' reactions to equal and unequal distributions of resources. *Developmental Science*, *14*(5), 1012–1020. https://doi.org/10.1111/j.1467-7687.2011.01048.x

Gerdin, E., & Dunham, Y. (2022). Do children interpret costs as signals of commitment to groups? *Proceedings of the Annual Meeting of the Cognitive Science Society*, *44*.

Gergely, G., & Csibra, G. (2003). Teleological reasoning in infancy: The naïve theory of rational action. *Trends in Cognitive Sciences*, *7*(7), 287–292. https://doi.org/10.1016/S1364-6613(03)00128-1

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/10.1016/0010-0277(95)00661-H

Gerstenberg, T. (2022). What would have happened? Counterfactuals, hypotheticals and causal judgements. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *377*(1866), 20210339. https://doi.org/10.1098/rstb.2021.0339

Gilbert, D. T., & Malone, P. S. (1995). The correspondence bias. *Psychological Bulletin*, *117*(1), 21–38. https://doi.org/10.1037/0033-2909.117.1.21

Gönül, G., Kammermeier, M., & Paulus, M. (2023). What is in an action? Preschool children predict that agents take previous paths and not previous goals. *Developmental Science*, e13466. https://doi.org/10.1111/desc.13466

Gönül, G., & Paulus, M. (2021). Children's reasoning about the efficiency of others' actions: The development of rational action prediction. *Journal of Experimental Child Psychology*, *204*, 105035. https://doi.org/10.1016/j.jecp.2020.105035

Grueneisen, S., Török, G., Wathiyage Don, A., & Ruggeri, A. (2023). Young children's adaptive partner choice in cooperation and competition contexts. *Child Development*. https://doi.org/10.1111/cdev.14036

Hamlin, J. K. (2013a). Failed attempts to help and harm: Intention versus outcome in preverbal infants' social evaluations. *Cognition*, *128*(3), 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*(3), 186–193. https://doi.org/10.1177/0963721412470687

Hamlin, J. K. (2014). Context-dependent social evaluation in 4.5-month-old human infants: The role of domain-general versus domain-specific processes in the development of social evaluation. *Frontiers in Psychology*, *5*, 614. https://doi.org/10.3389/fpsyg.2014.00614

Hamlin, J. K. (2015). The case for social evaluation in preverbal infants: Gazing toward one's goal drives infants' preferences for Helpers over Hinderers in the hill paradigm. *Frontiers in Psychology*, *5*, 1563. https://doi.org/10.3389/fpsyg.2014.01563

Hamlin, J. K., Hallinan, E. V., & Woodward, A. L. (2008). Do as I do: 7-month-old infants selectively reproduce others' goals. *Developmental Science*, *11*(4), 487–494. https://doi.org/10.1111/j.1467-7687.2008.00694.x

Hamlin, J. K., Newman, G. E., & Wynn, K. (2009). Eight-Month-Old Infants Infer Unfulfilled Goals, Despite Ambiguous Physical Evidence. *Infancy*, *14*(5), 579–590. https://doi.org/10.1080/15250000903144215

Hamlin, J. K., 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*(2), 209–226. https://doi.org/10.1111/desc.12017

Hamlin, J. K., & Wynn, K. (2011). Young infants prefer prosocial to antisocial others. *Cognitive Development*, *26*(1), 30–39. https://doi.org/10.1016/j.cogdev.2010.09.001

Hamlin, J. K., Wynn, K., & Bloom, P. (2007). Social evaluation by preverbal infants. *Nature*, *450*(7169), 557–559. https://doi.org/10.1038/nature06288

Hamlin, J. K., Wynn, K., & Bloom, P. (2010). Three-month-olds show a negativity bias in their social evaluations. *Developmental Science*, *13*(6), 923–929. https://doi.org/10.1111/j.1467-7687.2010.00951.x

Hamlin, J. K., Wynn, K., Bloom, P., & Mahajan, N. (2011). How infants and toddlers react to antisocial others. *Proceedings of the National Academy of Sciences*, *108*(50), 19931–19936. https://doi.org/10.1073/pnas.1110306108

He, X., Yang, Y., Lin, J., Wu, X., & Yin, J. (2020). Attributions of Social Interaction Depend on the Integration of the Actor's Simple Goal and the Influence on Recipients. *Social Cognition*, *38*(3), 266–286. https://doi.org/10.1521/soco.2020.38.3.266

Heider, F. (1958). *The psychology of interpersonal relations* (pp. ix, 326). John Wiley & Sons Inc. https://doi.org/10.1037/10628-000

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

Heintz, C., Karabegovic, M., & Molnar, A. (2016). The Co-evolution of Honesty and Strategic Vigilance. *Frontiers in Psychology*, *7*(7), 3389–1503. https://doi.org/10.3389/fpsyg.2016.01503

Heller, K. A., & Berndt, T. J. (1981). Developmental Changes in the Formation and Organization of Personality Attributions. *Child Development*, *52*(2), 683–691. https://doi.org/10.2307/1129190

Henderson, A. M. E., Wang, Y., Matz, L. E., & Woodward, A. L. (2013). Active Experience Shapes 10-Month-Old Infants' Understanding of Collaborative Goals. *Infancy*, *18*(1), 10–39. https://doi.org/10.1111/j.1532-7078.2012.00126.x

Henderson, A. M. E., & Woodward, A. L. (2011). "Let's work together": What do infants understand about collaborative goals? *Cognition*, *121*(1), 12–21. https://doi.org/10.1016/j.cognition.2011.05.008

Henderson, A. M. E., & Woodward, A. L. (2012). Nine-month-old infants generalize object labels, but not object preferences across individuals. *Developmental Science*, *15*(5), 641–652. https://doi.org/10.1111/j.1467-7687.2012.01157.x

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, *33*(2–3), 61–83. https://doi.org/10.1017/S0140525X0999152X

Hepach, R., Vaish, A., Grossmann, T., & Tomasello, M. (2016). Young Children Want to See Others Get the Help They Need. *Child Development*, *87*(6), 1703–1714. https://doi.org/10.1111/cdev.12633

Hepach, R., Vaish, A., & Tomasello, M. (2012). Young Children Are Intrinsically Motivated to See Others Helped. *Psychological Science*, *23*(9), 967–972. https://doi.org/10.1177/0956797612440571

Hepach, R., & Warneken, F. (2018). Editorial overview: Early development of prosocial behavior: Revealing the foundation of human prosociality. *Current Opinion in Psychology*, *20*, iv–viii. https://doi.org/10.1016/j.copsyc.2018.02.001

Hermes, J., Behne, T., & Rakoczy, H. (2015). The role of trait reasoning in young children's selective trust. *Developmental Psychology*, *51*(11), 1574–1587. https://doi.org/10.1037/dev0000042

Hermes, J., Behne, T., Studte, K., Zeyen, A.-M., Gräfenhain, M., & Rakoczy, H. (2016). Selective Cooperation in Early Childhood – How to Choose Models and Partners. *PLOS ONE*, *11*(8), e0160881. https://doi.org/10.1371/journal.pone.0160881

Hernik, M., & Southgate, V. (2012). Nine-months-old infants do not need to know what the agent prefers in order to reason about its goals: On the role of preference and persistence in infants' goal-attribution. *Developmental Science*, *15*(5), 714–722. https://doi.org/10.1111/j.1467-7687.2012.01151.x

Heyman, G. D. (2009). Children's reasoning about traits. In P. Bauer (Ed.), *Advances in Child Development and Behavior* (Vol. 37, pp. 105–143). JAI. https://doi.org/10.1016/S0065-2407(09)03703-3

Heyman, G. D., & Gelman, S. A. (1999). The Use of Trait Labels in Making Psychological Inferences. *Child Development*, *70*(3), 604–619. https://psycnet.apa.org/doi/10.1111/1467-8624.00044

Hobbs, K., & Spelke, E. (2015). Goal attributions and instrumental helping at 14 and 24 months of age. *Cognition*, *142*, 44–59. https://doi.org/10.1016/j.cognition.2015.03.014

Holvoet, C., Scola, C., Arciszewski, T., & Picard, D. (2019). Infants' social evaluation abilities: Testing their preference for prosocial agents at 6, 12 and 18 months with different social scenarios. *Early Child Development and Care*, *189*(6), 1018–1031. https://doi.org/10.1080/03004430.2017.1361415

Howard, R. M., Spokes, A. C., Mehr, S., & Krasnow, M. (2018). *Welfare tradeoff psychology is present in children and adults*. PsyArXiv. https://doi.org/10.31234/osf.io/6daeg

Hudson, M., McDonough, K. L., Edwards, R., & Bach, P. (2018). Perceptual teleology: Expectations of action efficiency bias social perception. *Proceedings of the Royal Society B: Biological Sciences*, *285*(1884), 20180638. https://doi.org/10.1098/rspb.2018.0638

Huh, M., & Friedman, O. (2019). Young children use supply and demand to infer desirability. *Developmental Psychology*, *55*(12), 2483–2490. https://doi.org/10.1037/dev0000833

Hunnius, S., & Bekkering, H. (2010). The early development of object knowledge: A study of infants' visual anticipations during action observation. *Developmental Psychology*, *46*(2), 446–454. https://doi.org/10.1037/a0016543

Hwang, H. G., & Markson, L. (2020). The Development of Social Exclusion Detection in Early Childhood: Awareness of Social Exclusion Does Not Always Align with Social Preferences. *Journal of Cognition and Development*, *21*(2), 166–190. https://doi.org/10.1080/15248372.2019.1706521

Jara-Ettinger, J., Floyd, S., Huey, H., Tenenbaum, J. B., & Schulz, L. E. (2020). Social Pragmatics: Preschoolers Rely on Commonsense Psychology to Resolve Referential Underspecification. *Child Development*, *91*(4), 1135–1149. https://doi.org/10.1111/cdev.13290

Jara-Ettinger, J., Floyd, S., Tenenbaum, J. B., & Schulz, L. E. (2017). Children understand that agents maximize expected utilities. *Journal of Experimental Psychology: General*, *146*(11), 1574–1585. https://doi.org/10.1037/xge0000345

Jara-Ettinger, J., Gweon, H., Schulz, L. E., & Tenenbaum, J. B. (2016). The Naïve Utility Calculus: Computational Principles Underlying Commonsense Psychology. *Trends in Cognitive Sciences*, *20*, 589–604. https://doi.org/10.1016/j.tics.2016.05.011

Jara-Ettinger, J., Gweon, H., Tenenbaum, J. B., & Schulz, L. E. (2015). Children's understanding of the costs and rewards underlying rational action. *Cognition, 140*, 14–23. https://doi.org/10.1016/j.cognition.2015.03.006

Jara-Ettinger, J., Tenenbaum, J. B., & Schulz, L. E. (2015). Not So Innocent: Toddlers' Inferences About Costs and Culpability. *Psychological Science*, *26*(5), 633–640. https://doi.org/10.1177/0956797615572806

Jin, K.-S., & Baillargeon, R. (2017). Infants possess an abstract expectation of ingroup support. *Proceedings of the National Academy of Sciences*, *114*(31), 8199–8204. https://doi.org/10.1073/pnas.1706286114

Kahneman, D., & Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, *47*(2), 263–291. https://doi.org/10.2307/1914185

Kalish, C. W. (2002). Children's predictions of consistency in people's actions. *Cognition*, *84*(3), 237–265. https://doi.org/10.1016/S0010-0277(02)00052-5

Kalish, C. W., & Shiverick, S. M. (2004). Children's reasoning about norms and traits as motives for behavior. *Cognitive Development*, *19*(3), 401–416. https://doi.org/10.1016/j.cogdev.2004.05.004

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

Kampis, D., Somogyi, E., Itakura, S., & Király, I. (2013). Do infants bind mental states to agents? *Cognition*, *129*(2), 232–240. https://doi.org/10.1016/j.cognition.2013.07.004

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*(2), 0037. https://doi.org/10.1038/s41562-016-0037

Kanakogi, Y., & Itakura, S. (2011). Developmental correspondence between action prediction and motor ability in early infancy. *Nature Communications*, *2*(1), 341. https://doi.org/10.1038/ncomms1342

Kanakogi, Y., Okumura, Y., Inoue, Y., Kitazaki, M., & Itakura, S. (2013). Rudimentary Sympathy in Preverbal Infants: Preference for Others in Distress. *PLOS ONE*, *8*(6), e65292. https://doi.org/10.1371/journal.pone.0065292

Kenward, B., & Dahl, M. (2011). Preschoolers distribute scarce resources according to the moral valence of recipients' previous actions. *Developmental Psychology*, *47*(4), 1054–1064. https://doi.org/10.1037/a0023869

Kibbe, M. M., & Leslie, A. M. (2013). What's the object of object working memory in infancy? Unraveling 'what' and 'how many.' *Cognitive Psychology*, *66*(4), 380–404. https://doi.org/10.1016/j.cogpsych.2013.05.001

Killen, M., & Turiel, E. (1998). Adolescents' and Young Adults' Evaluations of Helping and Sacrificing for Others. *Journal of Research on Adolescence*, *8*(3), 355–375. https://doi.org/10.1207/s15327795jra0803_4

Kim, E. Y., & Song, H. (2015). Six-month-olds actively predict others' goal-directed actions. *Cognitive Development*, *33*, 1–13. https://doi.org/10.1016/j.cogdev.2014.09.003

Kim, S., Sodian, B., & Paulus, M. (2014). Does he need help or can he help himself? Preschool children's expectations about others' instrumental helping versus self-helping. *Frontiers in Psychology*, *5*, 430. https://doi.org/10.3389/fpsyg.2014.00430

Kishimoto, R., Itakura, S., Fujita, K., & Hashiya, K. (2018). Evaluation of "Calculating" Helpers Based on Third-Party Observation in Adults and Children. *Psychologia*, *61*(3), 185–199. https://doi.org/10.2117/psysoc.2019-A008

Kleiman-Weiner, M., Austerweil, J. L., & Tenenbaum, J. B. (2016). Coordinate to cooperate or compete: Abstract goals and joint intentions in social interaction. *Proceedings of the 38th Annual Meeting of the Cognitive Science Society*.

Kleiner, M., Brainard, D., & Pelli, D. (2007). What's new in Psychtoolbox-3? *Perception*, *36*(14), 1–16.

Kochukhova, O., & Gredebäck, G. (2010). Preverbal Infants Anticipate That Food Will Be Brought to the Mouth: An Eye Tracking Study of Manual Feeding and Flying Spoons. *Child Development*, *81*(6), 1729–1738. https://doi.org/10.1111/j.1467-8624.2010.01506.x

Kominsky, J. F. (2019). PyHab: Open-source real time infant gaze coding and stimulus presentation software. *Infant Behavior & Development*, *54*, 114–119. https://doi.org/10.1016/j.infbeh.2018.11.006

Kominsky, J. F., Gerstenberg, T., Pelz, M., Sheskin, M., Singmann, H., Schulz, L., & Keil, F. C. (2021). The trajectory of counterfactual simulation in development. *Developmental Psychology*, *57*(2), 253–268. https://doi.org/10.1037/dev0001140

Köster, M., Itakura, S., Omori, M., & Kärtner, J. (2019). From understanding others' needs to prosocial action: Motor and social abilities promote infants' helping. *Developmental Science*, *22*(6), e12804. https://doi.org/10.1111/desc.12804

Köster, M., Ohmer, X., Nguyen, T. D., & Kärtner, J. (2016). Infants Understand Others' Needs. *Psychological Science*, *27*(4), 542–548. https://doi.org/10.1177/0956797615627426

Krogh-Jespersen, S., & Woodward, A. L. (2014). Making Smart Social Judgments Takes Time: Infants' Recruitment of Goal Information When Generating Action Predictions. *PLOS ONE*, *9*(5), e98085. https://doi.org/10.1371/journal.pone.0098085

Krupenye, C., & Hare, B. (2018). Bonobos Prefer Individuals that Hinder Others over Those that Help. *Current Biology*, *28*(2), 280-286.e5. https://doi.org/10.1016/j.cub.2017.11.061

Kudrnova, V., Spelke, E., & Thomas, A. J. (2023). *Infants Infer Social Relationships between Individuals who Engage in Imitative Social Interactions*. PsyArXiv. https://doi.org/10.31234/osf.io/zuwpc

Kuhlmeier, V. A., Dunfield, K. A., & O'Neill, A. C. (2014). Selectivity in early prosocial behavior. *Frontiers in Psychology*, *5*, 836.

Kuhlmeier, V. A., Wynn, K., & Bloom, P. (2003). Attribution of dispositional states by 12-month-olds. *Psychological Science*, *14*(5), 402–408. https://doi.org/10.1111/1467-9280.01454

Landy, J. F., Piazza, J., & Goodwin, G. P. (2016). When It's Bad to Be Friendly and Smart: The Desirability of Sociability and Competence Depends on Morality. *Personality and Social Psychology Bulletin*, *42*(9), 1272–1290. https://doi.org/10.1177/0146167216655984

Leahy, B. P., & Carey, S. E. (2020). The Acquisition of Modal Concepts. *Trends in Cognitive Sciences*, *24*(1), 65–78. https://doi.org/10.1016/j.tics.2019.11.004

Lee, W., Kim, E. Y., & Song, H. (2020). Do infants expect others to be helpful? *British Journal of Developmental Psychology*, *38*(3), 478–490. https://doi.org/10.1111/bjdp.12331

Lee, Y., Yun, J. E., Kim, E. Y., & Song, H. (2015). The Development of Infants' Sensitivity to Behavioral Intentions when Inferring Others' Social Preferences. *PLOS ONE*, *10*(9), e0135588. https://doi.org/10.1371/journal.pone.0135588

Li, J., & Tomasello, M. (2018). The development of intention-based sociomoral judgment and distribution behavior from a third-party stance. *Journal of Experimental Child Psychology*, *167*, 78–92. https://doi.org/10.1016/J.JECP.2017.09.021

Liu, D., Gelman, S. A., & Wellman, H. M. (2007). Components of Young Children ' s Trait Understanding: Behavior-to-Trait Predictions Inferences and Trait-to-Behavior. *Child Development*, *78*(5), 1543–1558. https://doi.org/10.1111/j.1467-8624.2007.01082.x

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., Pepe, B., Ganesh Kumar, M., Ullman, T. D., Tenenbaum, J. B., & Spelke, E. S. (2022). Dangerous Ground: One-Year-Old Infants are Sensitive to Peril in Other Agents' Action Plans. *Open Mind*, *6*, 211–231. https://doi.org/10.1162/opmi_a_00063

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

Liu, S., Ullman, T. D., Tenenbaum, J. B., & Spelke, E. S. (2017). Ten-month-old infants infer the value of goals from the costs of actions. *Science*, *358*(6366), 1038–1041. https://doi.org/10.1126/science.aag2132

Livesley, W. J., & Bromley, D. B. (1973). *Person perception in childhood and adolescence* (pp. xi, 320). John Wiley & Sons.

Lockhart, K. L., Nakashima, N., Inagaki, K., & Keil, F. C. (2008). From ugly duckling to swan?: Japanese and American beliefs about the stability and origins of traits. *Cognitive Development*, *23*(1), 155–179. https://doi.org/10.1016/j.cogdev.2007.08.001

Lucas, C. G., Griffiths, T. L., Xu, F., Fawcett, C., Gopnik, A., Kushnir, T., Markson, L., & Hu, J. (2014). The Child as Econometrician: A Rational Model of Preference

Understanding in Children. *PLoS ONE*, *9*(3), e92160. https://doi.org/10.1371/journal.pone.0092160

Lucca, K., Capelier-Mourguy, A., Byers-Heinlein, K., Cirelli, L., Ben, R. D., Frank, M. C., Henderson, A. M. E., Kominsky, J. F., Liberman, Z., Margoni, F., Reschke, P. J., Schlingloff-Nemecz, L., Scott, K., Soderstrom, M., Sommerville, J., Su, Y., Tatone, D., Uzefovsky, F., Wang, Y., … Hamlin, K. (2024). *Infants' Social Evaluation of Helpers and Hinderers: A Large-Scale, Multi-Lab, Coordinated Replication Study*. https://doi.org/10.31234/osf.io/qhxkm

Lucca, K., Pospisil, J., & Sommerville, J. A. (2018). Fairness informs social decision making in infancy. *PLoS ONE*, *13*(2), 1–14. https://doi.org/10.1371/journal.pone.0192848

Luo, Y., & Baillargeon, R. (2005). Can a Self-Propelled Box Have a Goal?: Psychological Reasoning in 5-Month-Old Infants. *Psychological Science*, *16*(8), 601–608. https://doi.org/10.1111/j.1467-9280.2005.015

Luo, Y., & Baillargeon, R. (2007). Do 12.5-month-old infants consider what objects others can see when interpreting their actions? *Cognition*, *105*(3), 489–512. https://doi.org/10.1016/j.cognition.2006.10.007

Luo, Y., & Johnson, S. C. (2009). Recognizing the role of perception in action at 6 months. *Developmental Science*, *12*(1), 142–149. https://doi.org/10.1111/j.1467-7687.2008.00741.x

Macfarlan, S. J., & Lyle, H. F. (2015). Multiple reputation domains and cooperative behaviour in two Latin American communities. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *370*(1683), 20150009. https://doi.org/10.1098/rstb.2015.0009

Malle, B. F. (2022). Attribution Theories: How People Make Sense of Behavior. In D. Chadee (Ed.), *Theories in Social Psychology, Second Edition* (2nd ed., pp. 93–120). Wiley. https://doi.org/10.1002/9781394266616.ch4

Malle, B. F., & Holbrook, J. (2012). Is there a hierarchy of social inferences? The likelihood and speed of inferring intentionality, mind, and personality. *Journal of Personality and Social Psychology*, *102*(4), 661–684. https://doi.org/10.1037/a0026790

Manzi, F., Ishikawa, M., Di Dio, C., Itakura, S., Kanda, T., Ishiguro, H., Massaro, D., & Marchetti, A. (2023). Infants' Prediction of Humanoid Robot's Goal-Directed Action. *International Journal of Social Robotics*, *15*(8), 1387–1397. https://doi.org/10.1007/s12369-022-00941-7

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. (2018). Infants' evaluation of prosocial and antisocial agents: A meta-analysis. *Developmental Psychology*, *54*(8), 1445–1455. https://doi.org/10.1037/dev0000538

Martin, J. W., Leddy, K., Young, L., & McAuliffe, K. (2022). An earlier role for intent in children's partner choice versus punishment. *Journal of Experimental Psychology: General*, *151*(3), 597–612. https://doi.org/10.1037/xge0001093

Martin, J., Young, L., & McAuliffe, K. (2019). *The psychology of partner choice* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/weqhz

Mascaro, O., & Csibra, G. (2012). Representation of stable social dominance relations by human infants. *Proceedings of the National Academy of Sciences*, *109*(18), 6862–6867. https://doi.org/10.1073/pnas.1113194109

Mascaro, O., & Csibra, G. (2022). Infants expect agents to minimize the collective cost of collaborative actions. *Scientific Reports*, *12*(1), 17088. https://doi.org/10.1038/s41598-022-21452-5

Maxwell, L., & Rafetseder, E. (2015). *Do toddlers have an innate moral core? An investigation into whether toddlers reward a good outcome over a good goal*. Poster presented at BCCCD15.

McAuliffe, K., Bogese, M., Chang, L. W., Andrews, C. E., Mayer, T., Faranda, A., Hamlin, J. K., & Santos, L. R. (2019). Do Dogs Prefer Helpers in an Infant-Based Social Evaluation Task? *Frontiers in Psychology*, *10*, 591. https://doi.org/10.3389/fpsyg.2019.00591

McDonough, K. L., Hudson, M., & Bach, P. (2019). Cues to intention bias action perception toward the most efficient trajectory. *Scientific Reports*, *9*(1), 6472. https://doi.org/10.1038/s41598-019-42204-y

McElreath, R. (2020). *Statistical Rethinking: A Bayesian Course with Examples in R and STAN* (2nd ed.). Chapman and Hall/CRC. https://doi.org/10.1201/9780429029608

Meltzoff, A. N. (1995). Understanding the intentions of others: Re-enactment of intended acts by 18-month-old children. *Developmental Psychology*, *31*(5), 838–850. https://doi.org/10.1037/0012-1649.31.5.838

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

Meristo, M., & Surian, L. (2014). Infants Distinguish Antisocial Actions Directed towards Fair and Unfair Agents. *PLoS ONE*, *9*(10), e110553. https://doi.org/10.1371/journal.pone.0110553

Michael, J., & Székely, M. (2019). Goal Slippage: A Mechanism for Spontaneous Instrumental Helping in Infancy? *Topoi*, *38*(1), 173–183. https://doi.org/10.1007/s11245-017-9485-5

Mikhail, J. (2011). *Elements of Moral Cognition: Rawls' Linguistic Analogy and the Cognitive Science of Moral and Legal Judgment*. Cambridge University Press. https://doi.org/10.1017/CBO9780511780578

Miller, C. B. (2016). Empirical Approaches to Moral Character. In *The Stanford Encyclopedia of Philosophy* (Summer 2023 Edition). https://plato.stanford.edu/archives/sum2023/entries/moral-character-empirical/

Miller, J. G. (1984). Culture and development of everyday social explanation. *Journal of Personality and Social Psychology*, *46*(5), 961–978. https://doi.org/10.1037//0022-3514.46.5.961

Mody, S., & Carey, S. (2016). The emergence of reasoning by the disjunctive syllogism in early childhood. *Cognition*, *154*, 40–48. https://doi.org/10.1016/j.cognition.2016.05.012

Morey, R. D., & Rouder, J. N. (2018). *BayesFactor: Computation of Bayes factors for common designs.* (0.9.12-4.2) [Computer software]. https://CRAN.R-project.org/package=BayesFactor

Morris, M. W., & Peng, K. (1994). Culture and Cause: American and Chinese Attributions for Social and Physical Events. *Joumal of Personality and Social Psychology*, *67*(6), 949–971. https://psycnet.apa.org/doi/10.1037/0022-3514.67.6.949

Mou, Y., Province, J. M., & Luo, Y. (2014). Can infants make transitive inferences? *Cognitive Psychology*, *68*, 98–112. https://doi.org/10.1016/j.cogpsych.2013.11.003

Na, J., & Kitayama, S. (2011). Spontaneous Trait Inference Is Culture-Specific: Behavioral and Neural Evidence. *Psychological Science*, *22*(8), 1025–1032. https://doi.org/10.1177/0956797611414727

Netanyahu, A., Shu, T., Katz, B., Barbu, A., & Tenenbaum, J. B. (2021). *PHASE: PHysically-grounded Abstract Social Events for Machine Social Perception* (arXiv:2103.01933). arXiv. https://doi.org/10.48550/arXiv.2103.01933

Nighbor, T., Kohn, C., Normand, M., & Schlinger, H. (2017). Stability of infants' preference for prosocial others: Implications for research based on single-choice paradigms. *PLoS ONE*, *12*(6), e0178818. https://doi.org/10.1371/journal.pone.0178818

Noë, R., & Hammerstein, P. (1994). Biological markets: Supply and demand determine the effect of partner choice in cooperation, mutualism and mating. *Behavioral Ecology and Sociobiology*, *35*(1), 1–11. https://doi.org/10.1007/BF00167053

Nucci, L., Turiel, E., & Roded, A. D. (2017). Continuities and Discontinuities in the Development of Moral Judgments. *Human Development*, *60*(6), 279–341. https://doi.org/10.1159/000484067

Olson, K. R., Banaji, M. R., Dweck, C. S., & Spelke, E. S. (2006). Children's Biased Evaluations of Lucky Versus Unlucky People and Their Social Groups. *Psychological Science*, *17*(10), 845–846. https://doi.org/10.1111/j.1467-9280.2006.01792.x

Olson, K. R., & Spelke, E. S. (2008). Foundations of cooperation in young children. *Cognition*, *108*(1), 222–231. https://doi.org/10.1016/j.cognition.2007.12.003

Paulus, M., & Hauf, P. (2011). Infants' use of material properties to guide their actions with differently weighted objects. *Infant and Child Development*, *20*(4), 423–436. https://doi.org/10.1002/icd.704

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), 976–983. https://doi.org/10.1037/a0023785

Paulus, M., & Moore, C. (2011). Whom to ask for help? Children's developing understanding of other people's action capabilities. *Experimental Brain Research*, *211*(3), 593–600. https://doi.org/10.1007/s00221-011-2676-1

Paulus, M., & Sodian, B. (2015). Which way to take? Infants select an efficient path to their goal. *Journal of Experimental Child Psychology*, *137*, 111–124. https://doi.org/10.1016/J.JECP.2015.03.014

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y

Pepe, B., & Powell, L. (2023). *Infants' expectations for prosociality in imitators*. OSF Preprints. https://doi.org/10.31219/osf.io/cjnkb

Perner, J., Sprung, M., & Steinkogler, B. (2004). Counterfactual conditionals and false belief: A developmental dissociation. *Cognitive Development*, *19*(2), 179–201. https://doi.org/10.1016/j.cogdev.2003.12.001

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

Phillips, J., Morris, A., & Cushman, F. (2019). How We Know What Not To Think. *Trends in Cognitive Sciences*, *23*(12), 1026–1040. https://doi.org/10.1016/j.tics.2019.09.007

Pomiechowska, B., & Csibra, G. (2020). Ten-month-olds infer relative costs of different goal-directed actions. *Proceedings of the 42nd Annual Meeting of the Cognitive Science Society*, *42*, 391–396. https://cognitivesciencesociety.org/cogsci20/papers/0072/0072.pdf

Pomiechowska, B., & Csibra, G. (2022). Infants infer motor competence from differences in agent-specific relative action costs. *Proceedings of the 44th Annual Meeting of the Cognitive Science Society*.

Pomiechowska, B., Tatone, D., Mészégető, D., Revencu, B., & Csibra, G. (2022, May 15). Children, but not adults, prioritize relational over dispositional interpretations of dominance interactions. *Proceedings of the 44th Annual Meeting of the Cognitive Science Society*. https://doi.org/10.31234/osf.io/49ah2

Poulin-Dubois, D. (2020). Chapter 7 - Theory of mind development: State of the science and future directions. In S. Hunnius & M. Meyer (Eds.), *Progress in Brain Research* (Vol. 254, pp. 141–166). Elsevier. https://doi.org/10.1016/bs.pbr.2020.05.021

Powell, L. J. (2022). Adopted Utility Calculus: Origins of a Concept of Social Affiliation. *Perspectives on Psychological Science*, *17*(5), 1215–1233. https://doi.org/10.1177/17456916211048487

Powell, L. J., & Spelke, E. S. (2013). Preverbal infants expect members of social groups to act alike. *Proceedings of the National Academy of Sciences*, *110*(41), E3965–E3972. https://doi.org/10.1073/pnas.1304326110

Powell, L. J., & Spelke, E. S. (2018a). Human infants' understanding of social imitation: Inferences of affiliation from third party observations. *Cognition*, *170*, 31–48. https://doi.org/10.1016/j.cognition.2017.09.007

Powell, L. J., & Spelke, E. S. (2018b). Third-Party Preferences for Imitators in Preverbal Infants. *Open Mind*, *2*(2), 61–71. https://doi.org/10.1162/opmi_a_00018

Premack, D., & Premack, A. J. (1997). Infants attribute value± to the goal-directed actions of self-propelled objects. *Journal of Cognitive Neuroscience*, *9*(6), 848–856. https://doi.org/10.1162/jocn.1997.9.6.848

Prétôt, L., Gonzalez, G., & McAuliffe, K. (2020). Children avoid inefficient but fair partners in a cooperative game. *Scientific Reports*, *10*(1), 10511. https://doi.org/10.1038/s41598-020-65452-9

Pun, A., Birch, S. A. J., & 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

R Core Team. (2023). *R: A Language and Environment for Statistical Computing* (4.3.2) [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/

Raihani, N. J., & Barclay, P. (2016). Exploring the trade-off between quality and fairness in human partner choice. *Royal Society Open Science*, *3*(11), 160510. https://doi.org/10.1098/rsos.160510

Rakoczy, H. (2022). Foundations of theory of mind and its development in early childhood. *Nature Reviews Psychology*, *1*(4), 223–235. https://doi.org/10.1038/s44159-022-00037-z

Rapp, D. J., Engelmann, J. M., Herrmann, E., & Tomasello, M. (2019). Young children's reputational strategies in a peer group context. *Developmental Psychology*, *55*(2), 329–336. https://doi.org/10.1037/dev0000639

Read, S. J., Jones, D. K., & Miller, L. C. (1990). Traits as goal-based categories: The importance of goals in the coherence of dispositional categories. *Journal of Personality and Social Psychology*, *58*(6), 1048–1061. https://doi.org/10.1037/0022-3514.58.6.1048

Redshaw, J., & Ganea, P. A. (2022). Thinking about possibilities: Mechanisms, ontogeny, functions and phylogeny. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *377*(1866), 20210333. https://doi.org/10.1098/rstb.2021.0333

Redshaw, J., & Suddendorf, T. (2016). Children's and Apes' Preparatory Responses to Two Mutually Exclusive Possibilities. *Current Biology*, *26*(13), 1758–1762. https://doi.org/10.1016/j.cub.2016.04.062

Reeder, G. (2009). Mindreading: Judgments About Intentionality and Motives in Dispositional Inference. *Psychological Inquiry*, *20*(1), 1–18. https://doi.org/10.1080/10478400802615744

Repacholi, B. M., & Gopnik, A. (1997). Early reasoning about desires: Evidence from 14- and 18-month-olds. *Developmental Psychology*, *33*(1), 12–21. https://doi.org/10.1037/0012-1649.33.1.12

Revencu, B., & Csibra, G. (2020). *Opening the black box of early depiction interpretation: From whether to how in the Theory-of-Puppets debate. Commentary on Kominsky, Lucca, Thomas, Frank, & Hamlin (submitted)* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/u3em8

Rhodes, M., Hetherington, C., Brink, K., & Wellman, H. M. (2015). Infants' use of social partnerships to predict behavior. *Developmental Science*, *18*(6), 909–916. https://doi.org/10.1111/desc.12267

Rholes, W. S., & Ruble, D. N. (1984). Children's Understanding of Dispositional Characteristics of Others. *Child Development*, *55*(2), 550–560. https://doi.org/10.2307/1129966

Riggs, K. J., Peterson, D. M., Robinson, E. J., & Mitchell, P. (1998). Are errors in false belief tasks symptomatic of a broader difficulty with counterfactuality? *Cognitive Development*, *13*(1), 73–90. https://doi.org/10.1016/S0885-2014(98)90021-1

Roberts, G. (1998). Competitive altruism: From reciprocity to the handicap principle. *Proceedings of the Royal Society B: Biological Sciences*, *265*(1394), 427–431. https://doi.org/10.1098/rspb.1998.0312

Robinson, E. J., & Beck, S. R. (2000). What is difficult about counterfactual reasoning? In *Children's Reasoning and the Mind* (pp. 101–119). Psychology Press.

Robson, S. J., & Kuhlmeier, V. A. (2019). *Reaching for an object or that object: Context and limitations in encoding object features may constrain infant goal attribution*. PsyArXiv. https://doi.org/10.31234/osf.io/hj7f9

Robson, S. J., Lee, V., Kuhlmeier, V. A., & Rutherford, M. D. (2014). Infants use contextual contingency to guide their interpretation of others' goal-directed behavior. *Cognitive Development*, *31*, 69–78. https://doi.org/10.1016/j.cogdev.2014.04.001

Ross, L. (1977). The Intuitive Psychologist And His Shortcomings: Distortions in the Attribution Process1. In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology* (Vol. 10, pp. 173–220). Academic Press. https://doi.org/10.1016/S0065-2601(08)60357-3

Ruffman, T., Aitken, J., Wilson, A., Puri, A., & Taumoepeau, M. (2018). A re-examination of the broccoli task: Implications for children's understanding of subjective desire. *Cognitive Development*, *46*, 79–85. https://doi.org/10.1016/j.cogdev.2017.08.001

Rule, J., Goddu, M., Chu, J., Pinter, V., Reagan, E. R., Bonawitz, E., Gopnik, A., & Ullman, T. (2023). *Fun isn't easy: Children choose more difficult options when "playing for fun" vs. "trying to win."* PsyArXiv. https://doi.org/10.31234/osf.io/q7wh4

Salvadori, E., Blazsekova, T., Volein, A., Karap, Z., Tatone, D., Mascaro, O., & Csibra, G. (2015). Probing the Strength of Infants' Preference for Helpers over Hinderers: Two Replication Attempts of Hamlin and Wynn (2011). *PLoS ONE*, *10*(11), e0140570. https://doi.org/10.1371/journal.pone.0140570

Scarf, D., Imuta, K., Colombo, M., & Hayne, H. (2012). Social Evaluation or Simple Association? Simple Associations May Explain Moral Reasoning in Infants. *PLoS ONE*, *7*(8), e42698. https://doi.org/10.1371/journal.pone.0042698

Schlingloff-Nemecz, L., Stavans, M., & Csibra, G. (2022). *CoCollectors: An iPad game for developmental research on action understanding, partner choice, and cooperative behavior*. PsychArchives. https://doi.org/10.23668/psycharchives.8176

Schmidt, M. F. H., & Sommerville, J. A. (2011). Fairness Expectations and Altruistic Sharing in 15-Month-Old Human Infants. *PLoS ONE*, *6*(10), e23223. https://doi.org/10.1371/journal.pone.0023223

Schuhmacher, N., Köster, M., & Kärtner, J. (2019). Modeling Prosocial Behavior Increases Helping in 16-Month-Olds. *Child Development*, *90*(5), 1789–1801. https://doi.org/10.1111/cdev.13054

Scola, C., Holvoet, C., Arciszewski, T., & Picard, D. (2015). Further Evidence for Infants' Preference for Prosocial Over Antisocial Behaviors. *Infancy*, *20*(6), 684–692. https://doi.org/10.1111/infa.12095

Scott, R. M., & Baillargeon, R. (2013). Do Infants Really Expect Agents to Act Efficiently? A Critical Test of the Rationality Principle. *Psychological Science*, *24*(4), 466–474. https://doi.org/10.1177/0956797612457395

Scott, R. M., & Baillargeon, R. (2017). Early False-Belief Understanding. *Trends in Cognitive Sciences*, *21*(4), 237–249. https://doi.org/10.1016/j.tics.2017.01.012

Sebanz, N., Bekkering, H., & Knoblich, G. (2006). Joint action: Bodies and minds moving together. *Trends in Cognitive Sciences*, *10*(2), 70–76. https://doi.org/10.1016/j.tics.2005.12.009

Shimizu, Y. A., & Johnson, S. C. (2004). Infants' attribution of a goal to a morphologically unfamiliar agent. *Developmental Science*, *7*(4), 425–430. https://doi.org/10.1111/j.1467-7687.2004.00362.x

Shimizu, Y., Lee, H., & Uleman, J. S. (2017). Culture as automatic processes for making meaning: Spontaneous trait inferences. *Journal of Experimental Social Psychology*, *69*, 79–85. https://doi.org/10.1016/j.jesp.2016.08.003

Shimizu, Y., Senzaki, S., & Uleman, J. S. (2018). The Influence of Maternal Socialization on Infants' Social Evaluation in Two Cultures. *Infancy*, *23*(5), 748–766. https://doi.org/10.1111/infa.12240

Shu, T., Kryven, M., Ullman, T. D., & Tenenbaum, J. B. (2020). Adventures in Flatland: Perceiving Social Interactions Under Physical Dynamics. *Proceedings of the 42nd Annual Meeting of the Cognitive Science Society*.

Shum, M., Kleiman-Weiner, M., Littman, M. L., & Tenenbaum, J. B. (2019). *Theory of Minds: Understanding Behavior in Groups Through Inverse Planning* (1901.06085v1). arXiv. https://doi.org/10.48550/arXiv.1901. 06085shu

Sierksma, J., Thijs, J., Verkuyten, M., & Komter, A. (2014). Children's Reasoning About the Refusal to Help: The Role of Need, Costs, and Social Perspective Taking. *Child Development*, *85*(3), 1134–1149. https://doi.org/10.1111/cdev.12195

Singh, L. (2020). Bilingual Infants are More Sensitive to Morally Relevant Social Behavior than Monolingual Infants. *Journal of Cognition and Development*, *21*(5), 631–650. https://doi.org/10.1080/15248372.2020.1807987

Siposova, B., Tomasello, M., & Carpenter, M. (2018). Communicative eye contact signals a commitment to cooperate for young children. *Cognition*, *179*, 192–201. https://doi.org/10.1016/j.cognition.2018.06.010

Skerry, A. E., & Spelke, E. S. (2014). Preverbal infants identify emotional reactions that are incongruent with goal outcomes. *Cognition*, *130*(2), 204–216. https://doi.org/10.1016/j.cognition.2013.11.002

Sloane, S., Baillargeon, R., & Premack, D. (2012). Do Infants Have a Sense of Fairness? *Psychological Science*, *23*(2), 196–204. https://doi.org/10.1177/0956797611422072

Smaldino, P., & Richerson, P. (2012). The Origins of Options. *Frontiers in Neuroscience*, *6*, 50.

Smith, K. M., & Apicella, C. L. (2020a). Hadza Hunter-Gatherers Disagree on Perceptions of Moral Character. *Social Psychological and Personality Science*, *11*(5), 616–625. https://doi.org/10.1177/1948550619865051

Smith, K. M., & Apicella, C. L. (2020b). Partner choice in human evolution: The role of cooperation, foraging ability, and culture in Hadza campmate preferences. *Evolution and Human Behavior*, *41*(5), 354–366. https://doi.org/10.1016/j.evolhumbehav.2020.07.009

Smith-Flores, A. S., Herrera-Guevara, I. A., & Powell, L. J. (2024). Infants expect friends, but not rivals, to be happy for each other when they succeed. *Developmental Science*, *27*(1), e13423. https://doi.org/10.1111/desc.13423

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

Sommerville, J. A., Enright, E. A., Horton, R. O., Lucca, K., Sitch, M. J., & Kirchner-Adelhart, S. (2018). Infants' prosocial behavior is governed by cost-benefit analyses. *Cognition*, *177*, 12–20. https://doi.org/10.1016/j.cognition.2018.03.021

Sommerville, J. A., Schmidt, M. F. H., Yun, J.-E., & Burns, M. (2013). The Development of Fairness Expectations and Prosocial Behavior in the Second Year of Life. *Infancy*, *18*(1), 40–66. https://doi.org/10.1111/j.1532-7078.2012.00129.x

Sommerville, J. A., & Woodward, A. L. (2005). Pulling out the intentional structure of action: The relation between action processing and action production in infancy. *Cognition*, *95*(1), 1–30. https://doi.org/10.1016/j.cognition.2003.12.004

Southgate, V., Johnson, M. H., & Csibra, G. (2008). Infants attribute goals even to biomechanically impossible actions. *Cognition*, *107*(3), 1059–1069. https://doi.org/10.1016/j.cognition.2007.10.002

Spaepen, E., & Spelke, E. (2007). Will any doll do? 12-month-olds' reasoning about goal objects. *Cognitive Psychology*, *54*(2), 133–154. https://doi.org/10.1016/j.cogpsych.2006.06.001

Spelke, E. S. (2016). Core Knowledge and Conceptual Change: A Perspective on Social Cognition. In *Core Knowledge and Conceptual Change* (pp. 279–300). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780190467630.003.0016

Spelke, E. S. (2022). *What Babies Know: Core Knowledge and Composition Volume 1*. Oxford University Press.

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*(3), 482–500. https://doi.org/10.1111/bjdp.12232

Steckler, C. M., Woo, B. M., & Hamlin, J. K. (2017). The limits of early social evaluation: 9-month-olds fail to generate social evaluations of individuals who behave inconsistently. *Cognition*, *167*, 255–265. https://doi.org/10.1016/J.COGNITION.2017.03.018

Surian, L., & Franchin, L. (2017). Toddlers Selectively Help Fair Agents. *Frontiers in Psychology*, *8*, 944.

Surian, L., Ueno, M., Itakura, S., & Meristo, M. (2018). Do Infants Attribute Moral Traits? Fourteen-Month-Olds' Expectations of Fairness Are Affected by Agents' Antisocial Actions. *Frontiers in Psychology*, *9*. https://www.frontiersin.org/articles/10.3389/fpsyg.2018.01649

Sylwester, K., & Roberts, G. (2010). Cooperators benefit through reputation-based partner choice in economic games. *Biology Letters*, *6*(5), 659–662. https://doi.org/10.1098/rsbl.2010.0209

Tasimi, A., & Wynn, K. (2016). Costly rejection of wrongdoers by infants and children. *Cognition*, *151*, 76–79. https://doi.org/10.1016/j.cognition.2016.03.004

Tatone, D., Geraci, A., & Csibra, G. (2015). Giving and taking: Representational building blocks of active resource-transfer events in human infants. *Cognition*, *137*, 47–62. https://doi.org/10.1016/j.cognition.2014.12.007

Tatone, D., Hernik, M., & Csibra, G. (2019). Minimal Cues of Possession Transfer Compel Infants to Ascribe the Goal of Giving. *Open Mind*, *3*, 31–40. https://doi.org/10.1162/opmi_a_00024

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

Tauzin, T., & Gergely, G. (2018). Communicative mind-reading in preverbal infants. *Scientific Reports*, *8*(1), 9534. https://doi.org/10.1038/s41598-018-27804-4

Téglás, E., Vul, E., Girotto, V., Gonzalez, M., Tenenbaum, J. B., & Bonatti, L. L. (2011). Pure reasoning in 12-month-old infants as probabilistic inference. *Science*, *332*(6033), 1054–1059. https://doi.org/10.1126/science.1196404

Thiele, M., Hepach, R., Michel, C., Gredebäck, G., & Haun, D. B. M. (2021). Social interaction targets enhance 13-month-old infants' associative learning. *Infancy*, *26*(3), 409–422. https://doi.org/10.1111/infa.12393

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*(9), 662–669. https://doi.org/10.1038/s41562-018-0415-3

Todorov, E. (2004). Optimality principles in sensorimotor control. *Nature Neuroscience*, *7*(9), 907–915. https://doi.org/10.1038/nn1309

Tooby, J., & Cosmides, L. (2008). The Evolutionary Psychology of the Emotions and Their Relationship to Internal Regulatory Variables. In *Handbook of emotions* (3rd edition, pp. 114–137). The Guilford Press.

Török, G., Pomiechowska, B., Csibra, G., & Sebanz, N. (2019). Rationality in Joint Action: Maximizing Coefficiency in Coordination. *Psychological Science*, *30*(6), 930–941. https://doi.org/10.1177/0956797619842550

Török, G., Stanciu, O., Sebanz, N., & Csibra, G. (2021). Computing Joint Action Costs: Co-Actors Minimize the Aggregate Individual Costs in an Action Sequence. *Open Mind*, *5*, 100–112. https://doi.org/10.1162/opmi_a_00045

Trivers, R. L. (1971). The Evolution of Reciprocal Altruism. *The Quarterly Review of Biology*, *46*(1), 35–57. https://doi.org/10.1086/406755

Trope, Y., & Gaunt, R. (2007). Attribution and Person Perception. In *The SAGE Handbook of Social Psychology: Concise Student Edition* (pp. 176–194). SAGE Publications Ltd. https://doi.org/10.4135/9781848608221

Ullman, T. D., Baker, C. L., Macindoe, O., Evans, O., Goodman, N. D., & Tenenbaum, J. B. (2009). Help or Hinder: Bayesian Models of Social Goal Inference. *Advances in Neural Information Processing Systems 22 (NIPS 2009)*.

Ullman, T. D., & Tenenbaum, J. B. (2020). Bayesian Models of Conceptual Development: Learning as Building Models of the World. *Annual Review of Developmental Psychology*, *2*, 533–558. https://doi.org/10.1146/annurev-devpsych-121318-084833

Uzefovsky, F., Paz, Y., & Davidov, M. (2020). Young infants are pro-victims, but it depends on the context. *British Journal of Psychology*, *111*(2), 322–334. https://doi.org/10.1111/bjop.12402

Vaish, A., Carpenter, M., & Tomasello, M. (2010). Young Children Selectively Avoid Helping People With Harmful Intentions: Selectivity in Helping. *Child Development*, *81*(6), 1661–1669. https://doi.org/10.1111/j.1467-8624.2010.01500.x

Van de Vondervoort, J. W., & Hamlin, J. K. (2016). Evidence for Intuitive Morality: Preverbal Infants Make Sociomoral Evaluations. *Child Development Perspectives*, *10*(3), 143–148. https://doi.org/10.1111/cdep.12175

Van De Vondervoort, J. W., & Hamlin, J. K. (2017). Preschoolers' social and moral judgments of third-party helpers and hinderers align with infants' social evaluations. *Journal of Experimental Child Psychology*, *164*, 136–151. https://doi.org/10.1016/j.jecp.2017.07.004

Van de Vondervoort, J. W., & Hamlin, J. K. (2018). Preschoolers Focus on Others' Intentions When Forming Sociomoral Judgments. *Frontiers in Psychology*, *9*, 1851. https://doi.org/10.3389/fpsyg.2018.01851

Vaporova, E., & Zmyj, N. (2020). Social evaluation and imitation of prosocial and antisocial agents in infants, children, and adults. *PLoS ONE*, *15*(9), e0235595. https://doi.org/10.1371/journal.pone.0235595

Verschoor, S., & Biro, S. (2012). Primacy of Information About Means Selection Over Outcome Selection in Goal Attribution by Infants. *Cognitive Science*, *36*(4), 714–725. https://doi.org/10.1111/j.1551-6709.2011.01215.x

Vesper, C., & Sevdalis, V. (2020). Informing, Coordinating, and Performing: A Perspective on Functions of Sensorimotor Communication. *Frontiers in Human Neuroscience*, *14*, 168.

von Hofsten, C. (1980). Predictive reaching for moving objects by human infants. *Journal of Experimental Child Psychology*, *30*(3), 369–382. https://doi.org/10.1016/0022-0965(80)90043-0

Wallace, R. J. (2003). Practical Reason. In *The Stanford Encyclopedia of Philosophy* (Spring 2020 Edition). https://plato.stanford.edu/archives/spr2020/entries/practical-reason/

Wang, R. E., Wu, S. A., Evans, J. A., Tenenbaum, J. B., Parkes, D. C., & Kleiman-Weiner, M. (2020). Too many cooks: Bayesian inference for coordinating multi-agent collaboration. *arXiv:2003.11778 [Cs]*. http://arxiv.org/abs/2003.11778

Wang, Y., & Henderson, A. M. E. (2018). We cooperated so… now what? Infants expect cooperative partners to share resources. *Infant Behavior and Development*, *52*, 9–13. https://doi.org/10.1016/j.infbeh.2018.04.005

Warneken, F. (2018). How Children Solve the Two Challenges of Cooperation. *Annual Review of Psychology*, *69*(1), 205–229. https://doi.org/10.1146/annurev-psych-122216-011813

Warneken, F., & Tomasello, M. (2007). Helping and Cooperation at 14 Months of Age. *Infancy*, *11*(3), 271–294. https://doi.org/10.1111/j.1532-7078.2007.tb00227.x

Weller, D., & Hansen Lagattuta, K. (2013). Helping the In-Group Feels Better: Children's Judgments and Emotion Attributions in Response to Prosocial Dilemmas. *Child Development*, *84*(1), 253–268. https://doi.org/10.1111/j.1467-8624.2012.01837.x

Westra, E. (2018). Character and theory of mind: An integrative approach. *Philosophical Studies*, *175*(5), 1217–1241. https://doi.org/10.1007/s11098-017-0908-3

Wolpert, D. M., & Landy, M. S. (2012). Motor control is decision-making. *Current Opinion in Neurobiology*, *22*(6), 996–1003. https://doi.org/10.1016/j.conb.2012.05.003

Wong, A., Cordes, S., Harris, P. L., & Chernyak, N. (2023). Being nice by choice: The effect of counterfactual reasoning on children's social evaluations. *Developmental Science*, *26*(6), e13394. https://doi.org/10.1111/desc.13394

Woo, B. M., Liu, S., Gweon, H., & Spelke, E. (2024). *Toddlers prefer agents who help those facing harder tasks*. OSF. https://doi.org/10.31234/osf.io/pb7mz

Woo, B. M., & Spelke, E. (2020a). *Infants' social evaluations depend on the intentions of agents who act on false beliefs* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/eczgp

Woo, B. M., & Spelke, E. S. (2020b). How to Help Best: Infants' Changing Understanding of Multistep Actions Informs their Evaluations of Helping. *Proceedings of the 42nd Annual Meeting of the Cognitive Science Society*.

Woo, B. M., & Spelke, E. S. (2023). Toddlers' social evaluations of agents who act on false beliefs. *Developmental Science*, *26*(2), 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*(1), 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), 1–34. https://doi.org/10.1016/S0010-0277(98)00058-4

Woodward, A. L. (1999). Infants' ability to distinguish between purposeful and non-purposeful behaviors. *Infant Behavior and Development*, *22*(2), 145–160. https://doi.org/10.1016/S0163-6383(99)00007-7

Woodward, A. L., & Sommerville, J. A. (2000). Twelve-Month-Old Infants Interpret Action in Context. *Psychological Science*, *11*(1), 73–77. https://doi.org/10.1111/1467-9280.00218

Woodward, A. M., Horen, L. A., Knoll, S. J., & Beier, J. S. (2022). Children's developing views of social excluders: A dissociation between social evaluation and partner preference. *British Journal of Developmental Psychology*, *40*(3), 410–421. https://doi.org/10.1111/bjdp.12413

Wyman, E., Rakoczy, H., & Tomasello, M. (2013). Non-verbal communication enables children's coordination in a "Stag Hunt" game. *European Journal of Developmental Psychology*, *10*(5), 597–610. https://doi.org/10.1080/17405629.2012.726469

Wynn, K. (2009). Constraints on natural altruism. *British Journal of Psychology*, *100*(3), 481–485. https://doi.org/10.1348/000712609X441312

Yin, J., Csibra, G., & Tatone, D. (2022). Structural asymmetries in the representation of giving and taking events. *Cognition*, *229*, 105248. https://doi.org/10.1016/j.cognition.2022.105248

Ziv, T., & Sommerville, J. A. (2017). Developmental differences in infants' fairness expectations from 6 to 15 months of age. *Child Development*, *88*(6), 1930–1951. https://doi.org/10.1111/cdev.12674

Appendix A: Comparing alternative means vs. outcomes

This study was meant to be a follow up to Experiment 1.2, where we found that infants did not look longer when an agent chose the one of two identical goal objects located behind a high wall compared to when he chose another behind a low wall. To investigate potential reasons for this failure, and get a more fine-grained understanding of the scope and limitations of 10-month-olds' utility-based reasoning about other agents' goal-directed actions, we designed the present experiment.

We intended to run two of three experiments, the second (B or C) dependent on the outcome of the first (A). First, in Experiment A, we wanted to test whether infants would expect an agent to choose between two means for approaching a goal object the one that was least costly. Infants here were to be familiarized to an agent repeatedly detouring an obstacle consisting of two walls of equal length (long and short walls in alternation) to approach a goal object. At test, one of the walls was short and the other one long. In one test event the agent detoured the long wall, in the other, the short wall. If infants interpreted the agent's behavior during familiarization as a cost-efficient way to achieve a goal, they should look longer at test if the agent chooses the path around the long wall.

If Experiment A yielded the predicted result, we planned to move on to Experiment B, where we aimed to test whether infants would expect an agent to choose between two outcomes the one that can be reached at lower costs. The structure of this experiment was similar to that of Experiment A, except that now were are two identical goal objects, each located behind one side of the wall, respectively. During familiarization, the agent sequentially approached both goal objects, again, by detouring long and short walls in alternation. At test, one wall was short and one long, and there was a goal object behind each of them. In one test event, the agent detoured the long wall, in the other, the short wall. If infants assigned equal value to the goal objects and interpreted the agent's behavior during familiarization as a cost-efficient way to achieve a goal, they should look longer if the agent chooses the object behind the long wall.

In contrast, if Experiment A did not yield the predicted result, we instead planned to run Experiment C, to test whether infants would at least be sensitive to departures from efficiency that do not require comparing two spatially distinct means or outcomes. The structure was also similar to Experiment A: The agent approached a single goal object located behind an obstacle with two walls of equal length (here, the walls were always long); the agent detoured the wall on the left and on the right side in alternation. At test, one wall was short and the other long. In one test event, the agent detoured the long wall in an efficient manner (i.e. by moving close to the wall, thus only detouring as much as necessary); in the other test event, the agent detoured the short wall in an inefficient manner (i.e., with the same motion trajectory as the one around the long wall, thus detouring further away from the wall than necessary). If infants expected the agent to minimize action costs, they should look longer if the agent detours the short wall inefficiently (see Liu & Spelke, 2017).

Depending on the outcome of these experiments, we reasoned that different conclusions could be drawn about infants' understanding of utility maximization in goal-directed actions. The following explanations represent accounts that differ in the richness of the action concept they ascribe to infants: If we obtained the predicted results in both Experiments (A) and (B), this would constitute evidence that infants expect agents to maximize utility, both by choosing means and outcomes that minimize action costs. If we obtained the predicted results in Experiment (A), but not in Experiment (B), this would suggest that infants at this age do not compare actions with different outcomes – either because they fail to compute and compare the relative utility of these outcomes, or because they cannot represent counterfactual outcomes. The success of Experiment A would show, however, that they can generate and compare counterfactual representations of different actions directed towards the same end state. If we did not obtain the predicted results in Experiment (A), but we did in Experiment (C), this might indicate that infants do not compare the relative efficiency of different actions aimed at the same goal but modeled on spatially distinct constraints, but are sensitive to the inefficiency of actions when improperly modeled to the same physical constraints.

We started running Experiment A, but quit before completing a full sample, as results were not looking promising. In particular, many infants were failing to meet our prespecified attentiveness criteria and had to be excluded from the sample.

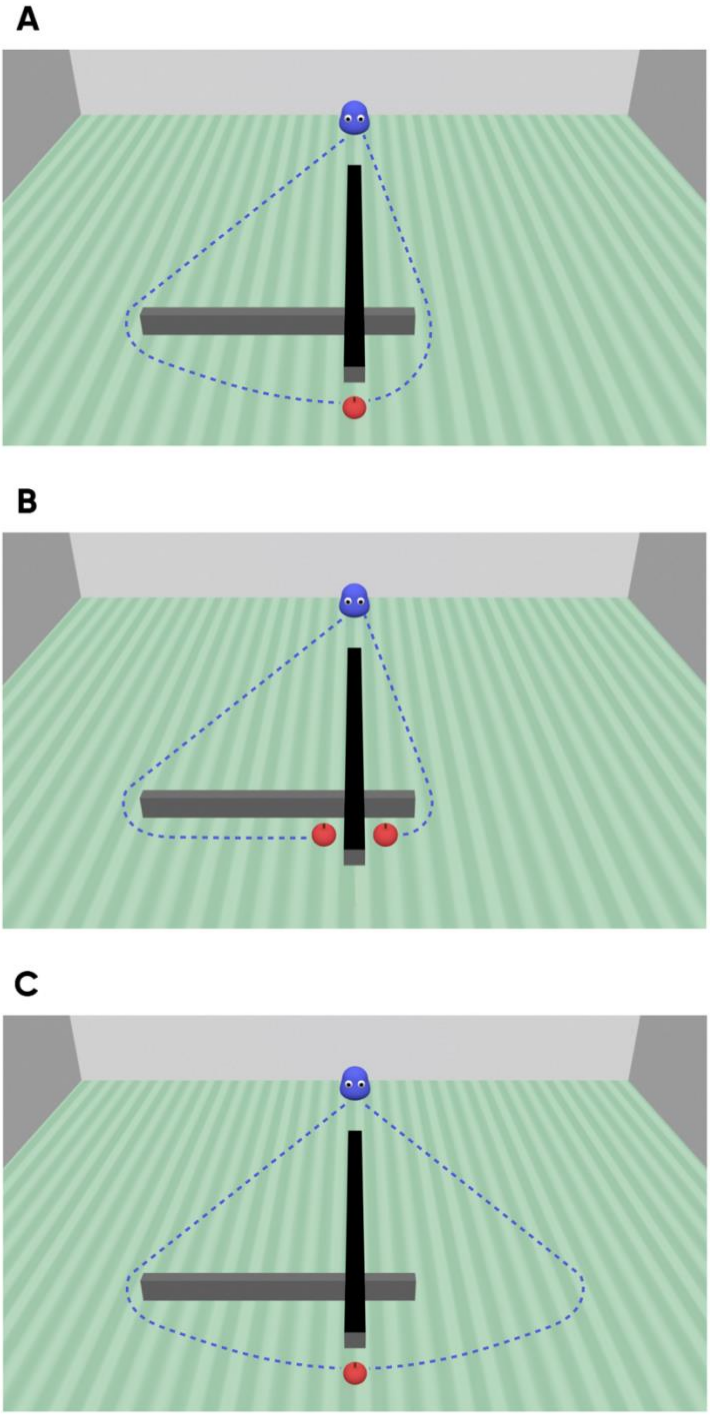

Figure A.1. Test event stimuli of the project reported in Appendix A. Dotted lines indicate the motion trajectories of the agent in the consistent and inconsistent test events, respectively. In Experiment A (for which we collected some data, reported here), the agent at test takes a short path (right, consistent event) or long path (left, inconsistent event) to reach the goal. In Experiment B, there are two goals, and the agent again takes a short path (right, consistent event) or long path (left, inconsistent event) to reach one of them. In Experiment C, the agent either takes a long path around the long wall that is efficiently adjusted to this constraint (left, consistent event) or an identical long path around the short wall, which is inefficient in this context (right, inconsistent event).

Methods

This experiment was preregistered at the OSF (https://osf.io/wz4ek).

Participants

Fourteen infants participated in this experiment (4 male, mean age: 303 days). An additional 21 infants were tested but were excluded from the analysis for failing to meet the preregistered attentiveness criteria (11), experimenter error (6), parental interference (2), technical failure (1), and fussiness (1).

Apparatus

The apparatus was the same as in Experiment 1.2.

Procedure and stimuli

Caregivers were instructed to hold the infants by their hips without impeding their ability to attend or disengage from the screen. Caregivers' eyes were covered with opaque sunglasses. Before each trial, an attention-getting clip was shown.

Familiarization. In the familiarization events, infants watched an agent detour an obstacle to reach a goal object (an apple). The obstacle was shaped like an upside-down cross (with long walls or short walls protruding left and right). At the beginning of each video, the agent was located at the center top of the screen, facing forward (2 s). Then, an apple fell from above and landed at the center of the screen, in front of the bottom end of the obstacle (1 s). The agent then hopped once, before moving towards the apple by detouring around the obstacle, either past the right or left side (5 s). Upon reaching the apple, he hopped once more and came to a standstill while touching the apple (3 s).

Infants viewed 6 familiarization trials, each containing 4 events. Within a trial, the 4 events show the agent detouring around the same obstacle (i.e., it had long walls or short walls on both sides), twice around one side and twice around the other (e.g., long-right, long-right, long left, long-left). The duration of the videos (detour around long vs. short wall) was identical. Trials began with different events in alternation (right first, left first; short first, long first). The videos within a trial were separated by short displays of a black screen (1 s). Each trial was preceded by a short attention-getter clip. A trial ended either after all its 4 videos finished playing, or after the infant looked away from the screen for a minimum of 2 seconds.

Test. The test phase consisted of 2 trials, one showing a consistent test event ("efficient" goal approach), the other an inconsistent test event ("inefficient" goal approach). The layout of the scene at test differed from that used at familiarization: It showed the long and short wall simultaneously (one on each side). Otherwise, the videos were identical to those used in familiarization. In each of the two test trials, the test event was looped and interspersed with the brief display of a black screen (1 s). Each test trial was preceded by a longer attention-getter clip (15 s), which was different from the one used at familiarization. A test trial ended either when the infant looked away from the screen for a minimum of 2 seconds, or when 60 seconds had elapsed.

We counterbalanced the order of test events (inconsistent first vs. consistent first), the order of familiarization events (short path first vs. long path first), and the side of the short wall at test (left vs. right).

Stimuli can be accessed at https://osf.io/k83xv/.

Coding and analyses

Data coding and (Bayesian) analysis were the same as in Experiment 1.2. We preregistered the following exclusion criteria: During familiarization, participants had to attend to at least 6 goal approach motions for >50% of their duration (from the beginning of the approach motion until contact with the goal object); and they had to meet this criterion at least once for each of the familiarization events (left-long, right-long, left-short, right-short). Additionally, subjects had to attend to a minimum of 50% of the crucial goal approach action at test. Participants who attended to both test events for their entire duration (i.e., who did not terminate at least one test trial by looking away from the screen for a minimum of 2 seconds) were also excluded. Further exclusion criteria were fussiness, caregiver intervention (e.g., talking to the infant, readjusting the position at test to direct the infant's gaze to the screen), experimenter error (e.g., overexposure at test), external distractions, and technical failure.

Results

Infants did not look longer when the agent took a long path towards the goal ($M_{\text{long path}}$ = 15.79 s, $SD_{\text{long path}}$ = 7.64 s) compared to a short path ($M_{\text{short path}}$ = 22.27 s, $SD_{\text{short path}}$ = 15.33 s). The Bayesian analysis was inconclusive: We obtained a BF of 2.489. The average looking times were going in the other direction than we predicted.

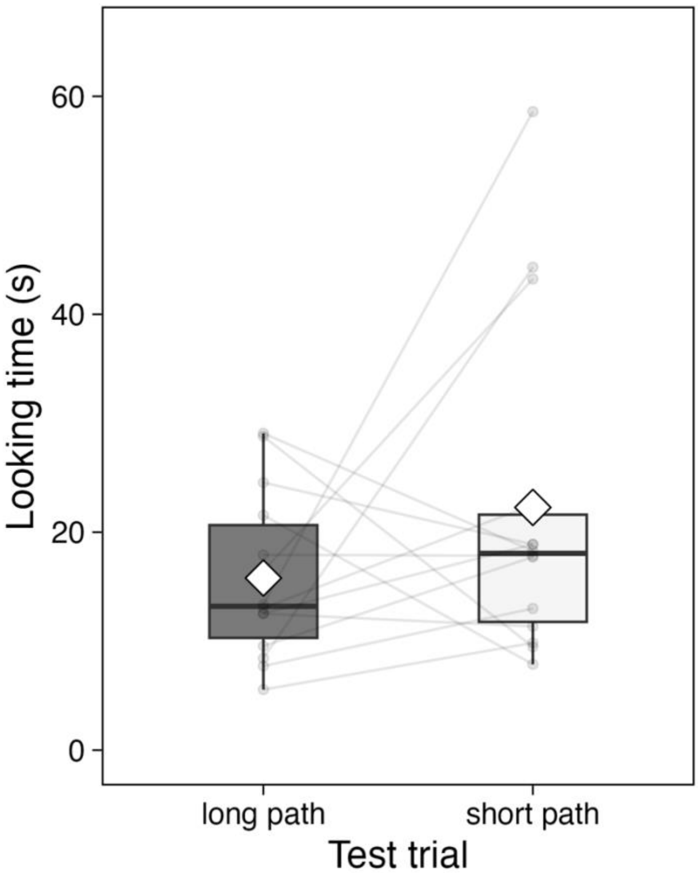

Figure A.2. Boxplot of average looking times (in seconds) toward the test events in the experiment reported in Appendix A. Light grey lines connect the looking times of individual participants, white diamonds indicate means, horizontal lines indicate medians, boxes indicate middle quartiles, and whiskers indicate points within 1.5 times the interquartile range from the upper and lower edges of the middle quartiles.

Appendix B: Interpreting subgoals as means to reduce the overall cost of an action sequence

After obtaining null results in Experiments 2.2.1 and 2.2.2 (which investigated whether infants ascribe a second-order goal to an agent who moved an obstacle so that another agent could take a relatively shorter path), we speculated that this failure may be due to infants' difficulty to make sense of the efficiency of instrumental means. The 12-month-olds we tested may have struggled to understand the hierarchical structure of the action sequence we presented to them, i.e., that by moving the block, the Helper targeted a subgoal contributing to the Helpee's ultimate goal of reaching the target object. Some prior research has found that infants at this age can comprehend the logic of means-ends sequences (Sommerville & Woodward, 2005; A. L. Woodward & Sommerville, 2000).

However, the task posed in these studies differs in two potentially important ways from ours: First, the stimuli feature actions that are likely familiar to infants (opening a box, pulling a cloth) and which are in infants' own motor repertoire, while our stimuli depict animations. It is not known whether infants can make sense of the hierarchical structure of more abstract and possibly unfamiliar actions. Second, in prior studies studying infants' understanding of means-ends relations, the means or subgoal was causally related to the targeted outcome (i.e., reaching the goal object) such that it was necessary to perform the action to bring about the goal. In our design, during familiarization, the Helper performed an action which allowed the Helpee to reach his goal at a lower cost, but it was not an enabling action: The Helpee always had the possibility to reach the goal on his own with more effort. It may be that infants do not comprehend that pursuing subgoals can serve the purpose of simply making it easier to reach one's ultimate goal. Think about a scenario where someone has to carry 20 soccer balls to the other end of the field: She might go back and forth, carrying 2 balls at a time, or she might first grab a large bag and put the balls in there before making her way across. Putting the balls in the bag is not necessary to accomplish the goal, but it saves the person quite a bit of effort.

In this study, we aimed to directly test whether infants can understand that steps in an action sequence which result in a cost reduction for the agent are not goals in themselves, but rather means. We familiarized infants to an agent who approached a goal object by passing through a gap in a wall; to do so, he pushed aside an obstacle blocking the gap, although a longer path around the far side of the wall was available. At test, the agent always pushed aside an obstacle. In one test event, this again served the purpose of reaching the goal object and was thus consistent with the agent's behavior during familiarization. In the other, the agent's most direct path to the object was already free. Therefore, pushing the obstacle was superfluous and the action inconsistent with the goal the agent previously pursued. If infants represented the means-ends structure of the familiarization events as we intended, they should look longer at the latter. We abandoned data collection after testing 10 infants as results were looking inconclusive.

Methods

An earlier version of this experiment was preregistered at the OSF (https://osf.io/zvwtf). However, we modified the design after preregistration (as many infants were failing to meet our attentiveness inclusion criteria), and restarted data collection. Here we report the testing procedure and preliminary results from the modified version.

Participants

Ten infants participated in this experiment. All ten met the inclusion criteria.

Apparatus

The apparatus was the same as in Experiment 2.2.1-3.

Procedure and stimuli

Caregivers were instructed to hold the infants by their hips without impeding their ability to attend or disengage from the screen. Caregivers' eyes were covered with opaque sunglasses. Before each trial, an attention-getting clip was shown.

Familiarization. In the familiarization events, infants watched an agent move to reach a goal object (a green ball). The agent was initially located at the top of the screen, left of center, facing forward. There was a wall extending horizontally from one border of the screen (left or right) for around ¾ of the screen width. The wall was interrupted by a gap located directly under the agent, which was obstructed by a blue rectangular block.

In the video, the agent moved until he was standing in front of the block (2 s), then moved around it to its side and pushed it away from the opening with two shoves (4 s). Finally, the agent passed through the gap, approached the object and stood next to it (4 s).

Infants watched 6 familiarization trials. After the end of each video, a still image of the final frame was displayed for a maximum of 10 seconds. A trial ended either after 10 seconds had elapsed, or when the infant looked away from the screen for a minimum of 2 seconds. Each trial was preceded by a short attention-getter clip.

Test. The test phase consisted of 2 trials, one showing a consistent, the other an inconsistent test event. The layout of the scene at test differed from that used at familiarization: In the consistent trial, the path to the goal was completely blocked by the wall and obstacle, such that the agent could only reach the obstacle by pushing the obstacle aside. In the inconsistent trial, there was a large opening in the wall such that the agent could approach the goal on a straight path. The agent's actions (pushing the block aside and moving towards the goal), their movement trajectories and timing were identical in the two test videos. After each test trial, a still image of the final frame was displayed for a maximum of 60 seconds. A test trial ended either when the infant looked away from the screen for a minimum of 2 seconds, or when 60 seconds had elapsed.

We counterbalanced the order of test events (inconsistent first vs. consistent first).

Stimuli can be accessed at https://osf.io/rwn7e/.

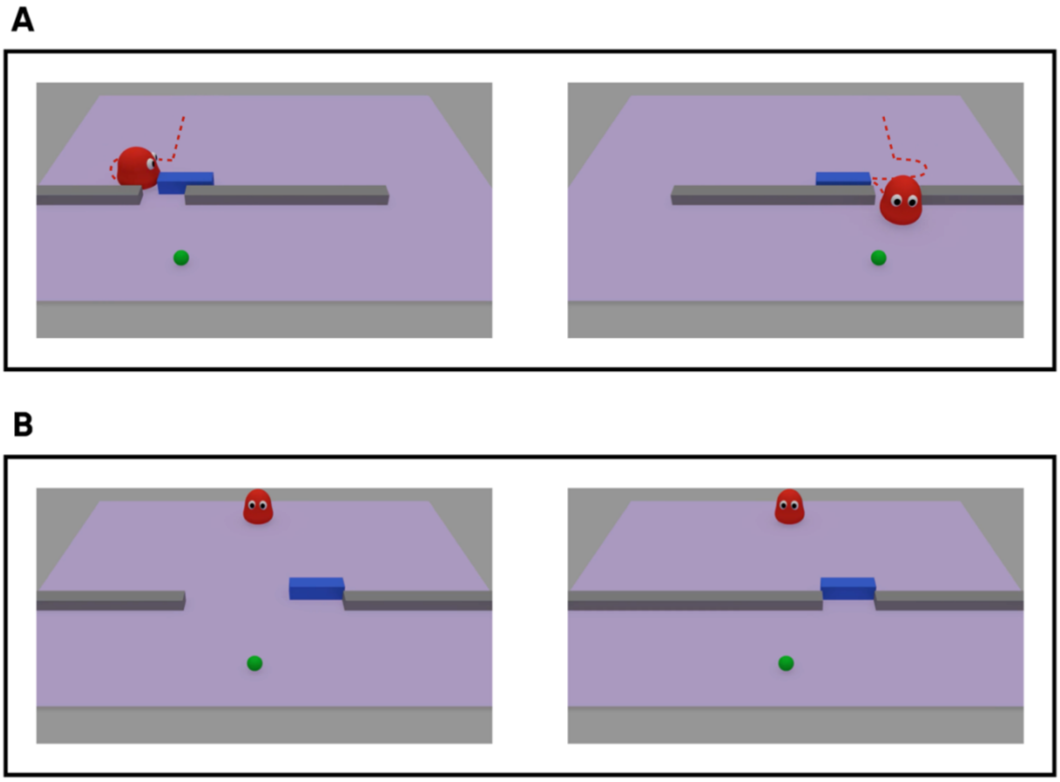

Figure B.1. Stimuli of the experiment reported in Appendix B. In the familiarization trials (A), the agent pushes aside an obstacle that is blocking the direct path to the goal, then approaches the goal. In the test trials (B), the agent does the same, but in one event this is instrumental, as it allows the agent to reach the goal (consistent event, right), while in the other it is superfluous, as the shortest path is already available (inconsistent event, left). The images in B depict the layout of the scene at the onset of the trial.

Coding and analyses

Data coding and (Bayesian) analysis were the same as in Experiment 2.2.1-3. We preregistered the following exclusion criteria: During familiarization, participants had to attend to at least 4 familiarization trials for >50% of their duration (this time window had to include half of the period in which the agent pushed the block). Additionally, subjects had to attend to the entire crucial block-pushing action in both test events, and look for at least 2 seconds after test trials ended. Participants who attended to both test events for their entire duration (i.e., who did not terminate at least one test trial by looking away from the screen for a minimum of 2 seconds) were also excluded. Further exclusion criteria were fussiness, caregiver intervention (e.g., talking to the infant, readjusting the position at test to direct the infant's gaze to the screen), experimenter error (e.g., overexposure at test), external distractions, and technical failure.

Results

Infants did not look longer when the agent's block-pushing action at test was instrumental ($M_{\text{instrumental}}$ = 10.61 s, $SD_{\text{instrumental}}$ = 11.08 s) compared to when it was superfluous ($M_{\text{superfluous}}$ = 9.38 s, $SD_{\text{superfluous}}$ = 6.4 s). The Bayesian analysis yielded a BF of 0.438, providing neither support for our hypothesis nor for the null hypothesis. However, the mean looking time to the inconsistent trial was strongly affected by a single outlier (40.1 s, all other looking times to this trial were 10 seconds and less; see Figure B.2).

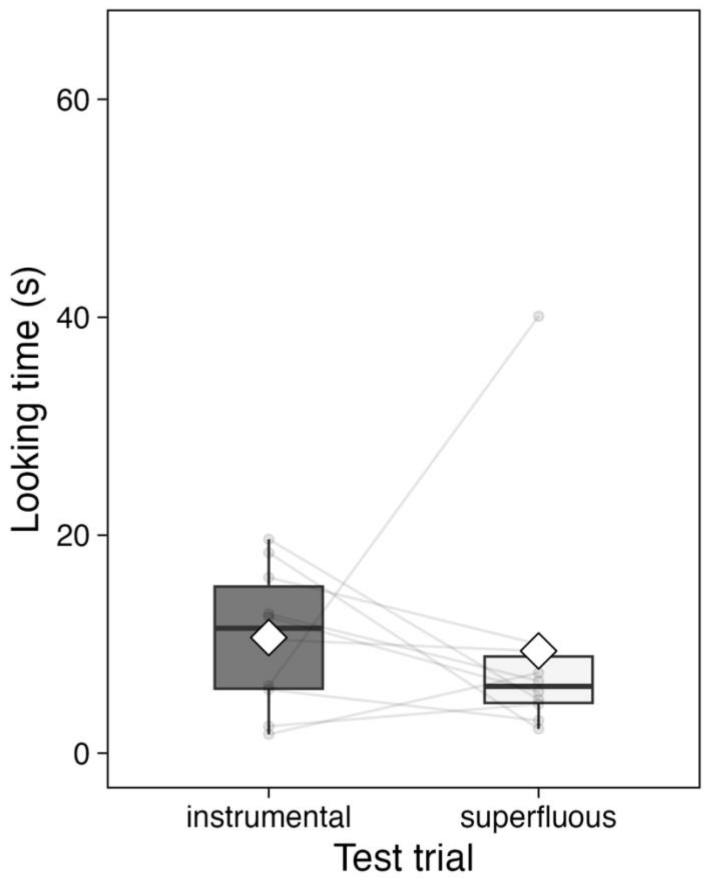

Figure B.2. Boxplot of average looking times (in seconds) toward the test events in the experiment reported in Appendix B. Light grey lines connect the looking times of individual participants, white diamonds indicate means, horizontal lines indicate medians, boxes indicate middle quartiles, and whiskers indicate points within 1.5 times the interquartile range from the upper and lower edges of the middle quartiles.

Appendix C: Replication of Experiment 2.2.2 with 18-month-old toddlers

In Experiment 2.2.2, we did not find evidence that 12-month-olds interpreted the goal of a Helper to be increasing the Helpee's action cost. We then attempted to replicate this experiment with a group of older participants (18-month-olds toddlers), to see if they would show the predicted pattern of looking times.

Methods

Participants

Thirteen infants participated in this experiment (8 male, mean age: 544 days). An additional 11 infants participated, but were excluded from the analysis for failure to meet our prespecified attentiveness criteria (n = 5), fussiness (n = 3), experimenter error (n = 2), and external distraction (n = 1).

Apparatus

The apparatus was the same as in Experiment 2.2.2.

Procedure and stimuli

Procedure and stimuli were the same as in Experiment 2.2.2.

Coding and analyses

Data coding, participant exclusion criteria, and (Bayesian) analysis were the same as in Experiment 2.2.2.

Results

Toddlers did not look longer at the event where the Helper pushed aside a block that was not in the way of the Helpee ($M_{inconsistent}$ = 16.62 s, $SD_{inconsistent}$ = 14.02 s) compared to when this action allowed the Helpee to take a shorter path ($M_{consistent}$ = 14.92 s, $SD_{consistent}$ = 8.98 s). With this sample size, the Bayesian analysis yielded a BF of 0.135, providing some evidence for the null hypothesis of no effect.

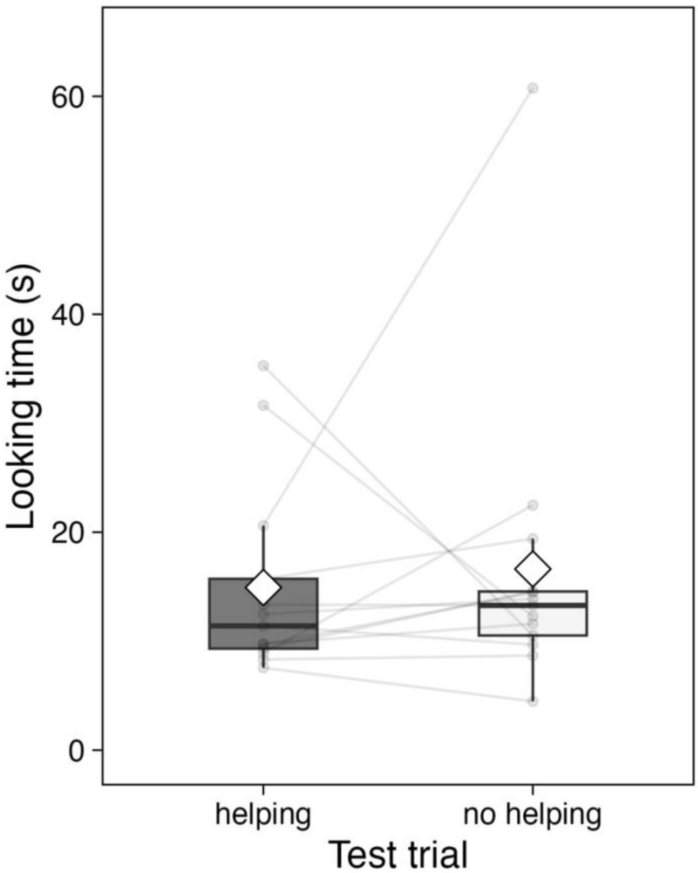

Figure C.1. Boxplot of average looking times (in seconds) toward the test events in the experiment reported in Appendix C. Light grey lines connect the looking times of individual participants, white diamonds indicate means, horizontal lines indicate medians, boxes indicate middle quartiles, and whiskers indicate points within 1.5 times the interquartile range from the upper and lower edges of the middle quartiles.

Appendix D: Supplementary materials to Section 3

D.1: Detailed description of configuration structure

A Configuration is a data structure that determines the parameters of a Co-Collectors game in JSON format. Within a Configuration, an ordered list of Rounds specifies the progress of the game. The specification of a round refers to an element in the coconutsSets array to set the environment of the round, and to one or two kobos from the agents array to identify the players of the round. Each kobo in the agents array is specified by linking a profile from the prosocialities array and a profile from the competences array to an avatarType. Table D.1 describes the Configuration structure and the other structures embedded in it. Variables within a structure can be specified in any order. Among the arrays of structures, only in rounds the order of the elements within the array is relevant. “Int” requires input values tobe specified as an integer, “Float” as a floating point number.

Table D.1. Game specification by JSON structures	
Structure	Explanation

Configuration[1] {
“description”: String,
“coconutSets”: [CoconutSet],
“agents”: [Agent],
“prosocialities”: [Prosociality],
“competences”: [Competence],
“rounds”: [Round]
}
	
	The title of the configuration
	An array of environments
	An array of agents
	An array of prosociality profiles
	An array of competence profiles
	An array of rounds

Round[2] {
“id”: Int,
“description”: String,
“type”: *RoundType*,
“seed”: Int,
“coconutSetId”: Int,
“leftAgentId”: Int,
“rightAgentId”: Int
}
	
	A unique identifier
	Name of the round [optional]
	From *RoundType* categories (see Table D.2)
	Seed of the random generator[2] [optional]
	The ID of the coconutSet in this round
	The agent ID for the left kobo[4]
	The agent ID for the right kobo[5]

Agent { “id”: Int, “description”: String, “prosocialityId”: Int, “competenceId”: Int,	 A unique identifier Name of the kobo [optional] ID of the prosociality profile of the kobo ID of the competence profile of the kobo

"avatarType": *AgentAvatarType* }	Selected from a specified list[6]
Prosociality { "id": Int, "description": String, "singleProCont": [Float], "singleProContOwn": [Float], "singleProContNeutral": [Float], "singleProContOther": [Float] }	A unique identifier Name of the profile [optional] Probability of approaching a new coconut – when no coconut is being cracked[7] Probability of approaching a new coconut – when an own coconut is being cracked[7] Probability of approaching a new coconut – when a brown coconut is being cracked[7] Probability of approaching a new coconut – when a partner's coconut is being cracked[7]
Competence { "id": Int, "description": String, "hitTime": Float, "reactionTime": Float }	A unique identifier Name of the profile [optional] Duration of a hit in seconds Delay of reacting to a new coconut in seconds
CoconutSet { "id": Int, "description": String, "number": Int, "agentProportion": [Float], "leftBerries": Int, "followGap": [Float], "hardness": [Int], "doomTime": [Float], }	A unique identifier Name of the set [optional] Number of coconuts Probability of the three types of coconuts[8] Berries for the left agent inside brown coconuts[9] Min and max time between two coconuts Min and max hits to crack coconuts open Min and max lifetime of coconuts in seconds

1) The Configuration structure also includes further parameters that are set not by the user but by the server. These include an "id" (a unique integer identifier), a "creationDate" and a "modificationDate" (in yyyymmdd format), and an "identifierString" (an identifier of the user).

2) The Round structure also includes an internal parameter ("progress") set by the server. The user does not have to specify this.

3) If specified, random numbers are generated in the round starting from this seed. This ensures that rounds without players (Demo, Observation) will be played exactly the same way each time. If not specified, a real random number is used as a seed.

4) This must be specified in partnerSelection, Demo, and Observation rounds.

5) This must be specified in partnerSelection, and Observation rounds.

6) Can be "red", "orange", "teal", "lightblue", "darkblue", "purple", "pink", "lightpurple", or "lightpink".

7) Each of these arrays consists of three numbers between 0 and 1, specifying the

probabilities of approaching own, brown, or partner's coconuts, respectively.

8) Three numbers between 0 and 1, specifying the probability of a coconut belonging the left kobo, to the right kobo, or being brown, respectively. The three numbers must add up to 1. If the third number is not specified, it becomes 1 minus the sum of the two numbers in the array.

9) This number specifies how many of the 4 berries contained in brown coconuts should belong to the left kobo (must be between 0 and 4).

Table D.2. Round specification by RoundType values

RoundType	Explanation
avatarSelection	In this round, the player is offered to select her own avatar from 5 available options (see Figure 1B). Such a round should precede practice and cooprounds. Here seed, coconutSetId, leftAgentId, and rightAgentIdparameters are ignored and can be left out.
partnerSelection	The player's avatar appears in between two kobos, specified by leftAgentId, and rightAgentId, and the player can select the preferred partner by tapping it. Here seed, and coconutSetId parameters are ignored.
demo	The kobo identified by leftAgentId cracks coconuts and collects berries alone. Parameter rightAgentId is ignored.
practice	The player can collect berries alone. Parameters leftAgentId and rightAgentId are ignored.
observation	Two kobos, specified by leftAgentId and rightAgentId crack nuts and collect berries.
coop	If such a round is preceded by both an avatarSelection and a partnerSelection round, the player and the previously selected partner play together; the most recent ones among these are taken into account in specifying the kobos. In this case leftAgentId and rightAgentId are ignored. If one of these parameters are given a value and the round is not preceded by a partnerSelection round, the player plays together with the specified partner.

An example configuration can be seen below. Note that only the text in bold constitutes the JSON specification; the text in italic serves as explanation.

{
"configuration": {
"description": "Example", *a title given to the configuration*
"creationDate": "20200710", *the date in which the configuration was created, added by the server*
"modificationDate": "20200806", *the date in which the configuration was modified, added by the server*
"identifierString": "f07822e8b92ae968", *the unique identifier of the user, set by the server*

"prosocialities": [*the array of all prosociality profiles used in the game*
{
"id": 1, *a unique identifier, referred by agent specifications*
"description": “Committed", *name of the profile (optional) – an agent with this profile never leaves an uncracked nut*
"singleProCont": [*when no other coconut is being cracked, the probability of approaching a new coconut*
1, *if it is OWN (this agent always approaches a new OWN coconut)*
1, *if it is BROWN (this agent always approaches a BROWN coconut)*
1 *if it is PARTNER’s (this agent always approaches a PARTNER’s coconut)*
],
"doubleProContOwn": [*when OWN coconut is being cracked, the probability of approaching new coconut*
0, *if it is another OWN (this agent never approaches a new OWN coconut)*
0, *if it is BROWN (this agent never approaches a BROWN coconut)*
0 *if it is PARTNER’s (this agent never approaches a PARTNER’s coconut)*
],
"doubleProContNeutral": [*when BROWN coconut is being cracked, the probability of approaching new coconut*
0, *if it is OWN (this agent never approaches a new OWN coconut)*
0, *if it is another BROWN (this agent never approaches a BROWN coconut)*
0 *if it is PARTNER’s (this agent never approaches a PARTNER’s coconut*
],
"doubleProContOther": [*when PARTNER coconut is being cracked, the probability of approaching new coconut*
0, *if it is OWN (this agent never approaches a new OWN coconut)*
0, *if it is BROWN (this agent never approaches a BROWN coconut)*
0 *if it is another PARTNER’s (this agent never approaches a PARTNER’s coconut)*
]
},
{
"id": 2,
"description": "Selfish",
"singleProCont": [1, 1, 0], *no approach toward partner’s coconut*
"doubleProContOwn": [0, 0, 0], *doesn’t leave while cracking own coconuts*
"doubleProContNeutral": [1, 0, 0], *leaves a brown coconut for a new own coconut*
"doubleProContOther": [1, 1, 0] *leaves partner’s coconuts for own or brown ones (redundant)*
},
{
"id": 3,
"description": "Moderately Prosocial",
"singleProCont": [1, 1, 1],
"doubleProContOwn": [0, 0.3, 0.2],
"doubleProContNeutral": [0.6, 0, 0.4],
"doubleProContOther": [0.7, 0.4, 0]
},
{
"id": 4,
"description": "Highly Prosocial",
"singleProCont": [1, 1, 1],
"doubleProContOwn": [0, 0.5, 0.8],
"doubleProContNeutral": [0.2, 0, 0.2],
"doubleProContOther": [0.2, 0.2, 0]
}
],
"competences": [*the array of all competence profiles used in the game*
{
"id": 1, *a unique identifier*
"description": “Example”, *name of the profile (optional)*
"hitTime": 0.5, *Duration of a hit in seconds (an agent with this profile delivers 2 hits per second)*
"reactionTime": 0.5 *Delay of reacting to a new coconut in seconds*
},
{
"id": 2,
"description": "Moderate",
"hitTime": 0.5,
"reactionTime": 0.8

```
    },
    {
      "id": 3,
      "description": "Quick",
      "hitTime": 0.35,
      "reactionTime": 0.3
    },
    {
      "id": 4,
      "description": "Slow",
      "hitTime": 0.9,
      "reactionTime": 0.8
    }
  ],
  "coconutSets": [ the array of all coconut sets used during the game
    {
      "id": 1, a unique identifier
      "number": 3, there are 3 coconuts in this set
      "leftBerries": 2, inside brown coconuts there are 2 berries for the left agent
      "followGap": [ the time between two coconuts spawning ranges from 4 to 6 seconds
        4, minimum duration
        6 maximum duration
      ],
      “hardness”: [ the number of hits required to crack the coconut open ranges from 2 to 6 hits
        2, minimum number of hits
        6 maximum number of hits
      ],
      "doomTime": [ the lifetime in seconds of coconuts ranges from 3.5 to 7 seconds
        3.5, minimum lifetime duration
        7 maximum lifetime duration
      ],
      "agentProportion": [ the probability of the three types of coconuts spawning
        0.5, a new coconut will belong to the left agent with p = 0.5
        0 no coconut will have the color of the right agent
      ] only 2 values are specified, so the probability of brown nuts spawning = 1-the sum of probabilities
above, here 0.5
    },
    {
      "id": 2,
      “description”: “Skewed”,
      "number": 12,
      "leftBerries": 2,
      "followGap": [3, 6]
      "hardness": [3, 12],
      "doomTime": [4, 6],
      "agentProportion": [0.5, 0.25 ]
    },
    {
      "id": 3,
      "description": "Medium length”,
      "number": 16,
      "leftBerries": 2,
      "followGap": [2, 5],
      "hardness": [8, 12],
      "doomTime": [3, 6.5],
      "agentProportion": [0.33, 0.33]
    },
    {
      "id": 4,
      "description": "Short (for playing)”,
      "number": 8,
      "leftBerries": 2,
      "followGap": [2, 5],
      "hardness": [8, 12],
      "doomTime": [3, 6.5],
      "agentProportion": [0.33, 0.33]
    }
```

],
"agents": [
{
"id": 1, a unique identifier
"avatarType": "orange", *a particular agent appearance from a defined set. This one is the 'orange' one*
"prosocialityId": 3, *the agent has the prosociality profile ID 3. Refer to the prosociality array for the details*
"competenceId": 4, *the agent has the competence profile ID 4. Refer to the competence array for the details*
"description": "Low skill" *name of the agent*
},
{
"id": 2,
"avatarType": "lightpurple",
"prosocialityId": 3,
"competenceId": 3,
"description": "High skill"
},
{
"id": 3,
"avatarType": "teal",
"prosocialityId": 2,
"competenceId": 2,
"description": "Selfish"
},
{
"id": 4,
"avatarType": "orange",
"prosocialityId": 4,
"competenceId": 2,
"description": "Highly Prosoc"
},
{
"id": 5,
"avatarType": "darkblue",
"prosocialityId": 2,
"competenceId": 3,
"description": "Selfish_highSkill"
},
{
"id": 6,
"avatarType": "pink",
"prosocialityId": 4,
"competenceId": 4,
"description": "Highly Prosoc_lowSkill"
},
{
"id": 7,
"avatarType": "lightpink",
"prosocialityId": 1,
"competenceId": 1,
"description": "solo kobo"
}
],
"rounds": [*the array of rounds in the game in the order of presentation*
{
"id": 1, *a unique identifier*
"type": "demo", *this is a demo round in which the kobo identified by leftAgentId cracks coconuts alone*
"seed": 6, *Seed of the random generator. Because it is specified, this round will be identical every time it is played*
"coconutSetId": 1, *this round uses coconut set ID 1. Refer to the coconet sets array for more details*
"leftAgentId": 7, *agent ID 7 appears on the left*
"description": "shows lightpink agent cracking 3 nuts successfully" *a description of the round*
},
{
"id": 2,

"type": "avatarSelection",
"description": "child chooses its avatar"
},
{
"id": 3,
"type": "practice",
"seed": 121,
"coconutSetId": 2,
"description": "child practices" *with 12 coconuts of varying hardness*
},
{
"id": 4,
"type": "observation",
"seed": 3861,
"coconutSetId": 3,
"leftAgentId": 1,
"rightAgentId": 2,
"description": "Low skill vs. High skill"
},
{
"id": 5,
"type": "partnerSelection",
"leftAgentId": 1,
"rightAgentId": 2,
"description": "Low skill vs. High skill"
},
{
"id": 6,
"type": "coop",
"seed": 13,
"coconutSetId": 4,
"description": "Cooperation" *Player collects berries with the partner selected in Round* 5
},
{
"id": 7,
"type": "observation",
"seed": 1000,
"coconutSetId": 3,
"leftAgentId": 3,
"rightAgentId": 4,
"description": "Selfish vs. Highly Prosoc"
},
{
"id": 8,
"type": "partnerSelection",
"leftAgentId": 3,
"rightAgentId": 4,
"description": "Selfish vs. Highly Prosoc"
},
{
"id": 9,
"type": "coop",
"seed": 12,
"coconutSetId": 4
},
{
"id": 10,
"type": "observation",
"seed": 371,
"coconutSetId": 3,
"leftAgentId": 5,
"rightAgentId": 6,
"description": "Selfish_High skill vs. Highly Prosoc_Low skill"
},
{
"id": 11,
"type": "partnerSelection",

```
      "leftAgentId": 5,
      "rightAgentId": 6,
      "description": "Selfish_High skill vs. Highly Prosoc_Low skill"
    },
    {
      "id": 12,
      "type": "coop",
      "seed": 14,
      "coconutSetId": 4,
    }
  ]
 }
}
```

D.2: Detailed description of log files

Log files are text files generated by each game played with an active configuration. They can be uploaded to, and then accessed from, the server.

The log files have three sections: Header, Events, and Configuration, which are separated from each other by a blank line. The Header contains 5 lines:

1. Version of the app
2. Configuration ID and name
3. Researcher
4. Participant
5. Date

The Events section lists all events occurring in the game in separate lines. The first element in each line is the timestamp, showing the time elapsed from the start of the game in milliseconds. The next element identifies the type of event, which is then followed by values of variables describing the event. The table below describes these events and their parameters.

Event	Description	Parameters
GameStart	Game has started	real time hh:mm:ss
GameFinish	Game has finished	real time hh:mm:ss
Pause	Game is paused	
Resume	Game is resumed after pause	
RestartRound	Round is restarted after a pause	
Quit	Game is quit after pause	
Next	Game is proceeding to the next round after pause	
Previous	Game is returning to the previous round after pause	
RoundStart	Round has started	Round number Round ID Round seed Round type Round description leftAgentId (-1: player) rightAgentId (-1: player)
RoundFinish	Round has finished	Round number

		leftAgentGain (berries collected) rightAgentGain (berries collected)
AvatarSelected	An avatar is selected	avatar ID (“playerx” x = 1…5)
PartnerSelected	A partner is selected	Agent ID Agent description
Spawn	A coconut has spawned	nut number within round owner (agent ID or -1 if player or 0 if brown coconut) hardness doomTime
Land	A coconut has landed	nut number within round
Hit	A coconut has been hit	nut number within round agent ID (-1 if player)
Crack	A coconut is cracked open	nut number within round berries gained by left agent berries gained by right agent
Doom	A coconut is spoiled	nut number within round

The Configuration section simply repeats the configuration structure that generated the game.

D.3: Additional figures

<u>**Experiment 3.1**</u>

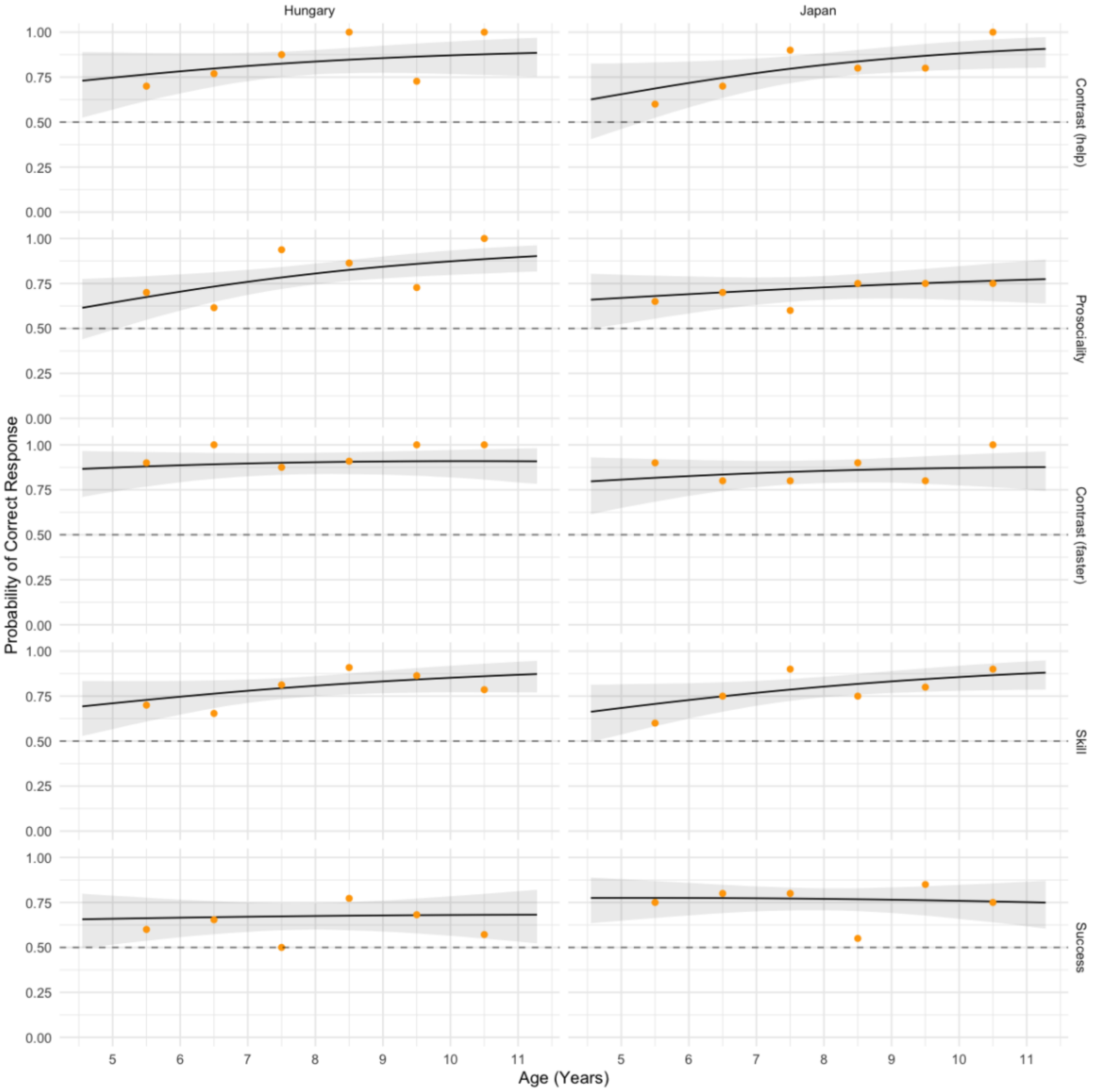

Figure D.1. Age differences in children's response accuracy in Experiment 3.1. The proportion of correct responses is indicated by orange dots. The black line indicates the mean of the parameter estimates; the shaded area represents the 89% CI. The left column shows the results for Hungarian (3.1a), the right for Japanese children (3.1b). The rows show results for the different trial types, with Contrast trials split by whether children were asked who helped more and who was faster (from the top: Contrast—who helped?; Prosociality; Contrast—who was faster?; Skill; Success).

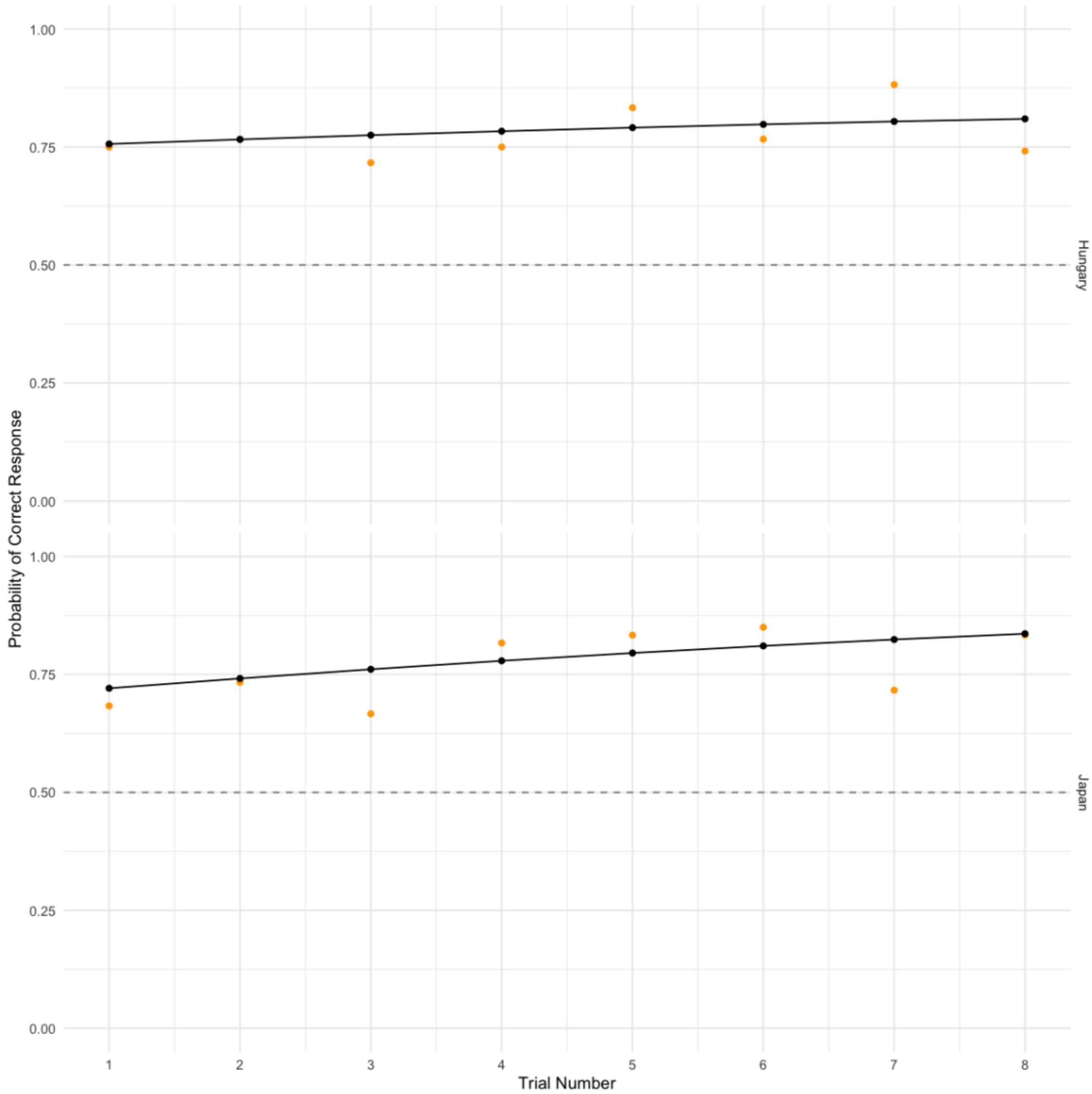

Figure D.2. Children's average response accuracy as a function of trial position in Experiment 3.1. The proportion of correct responses is indicated by orange dots. The black dots and line indicate the mean of the parameter estimates. The top row shows the results for Hungarian, the bottom for Japanese children.

Experiment 3.2

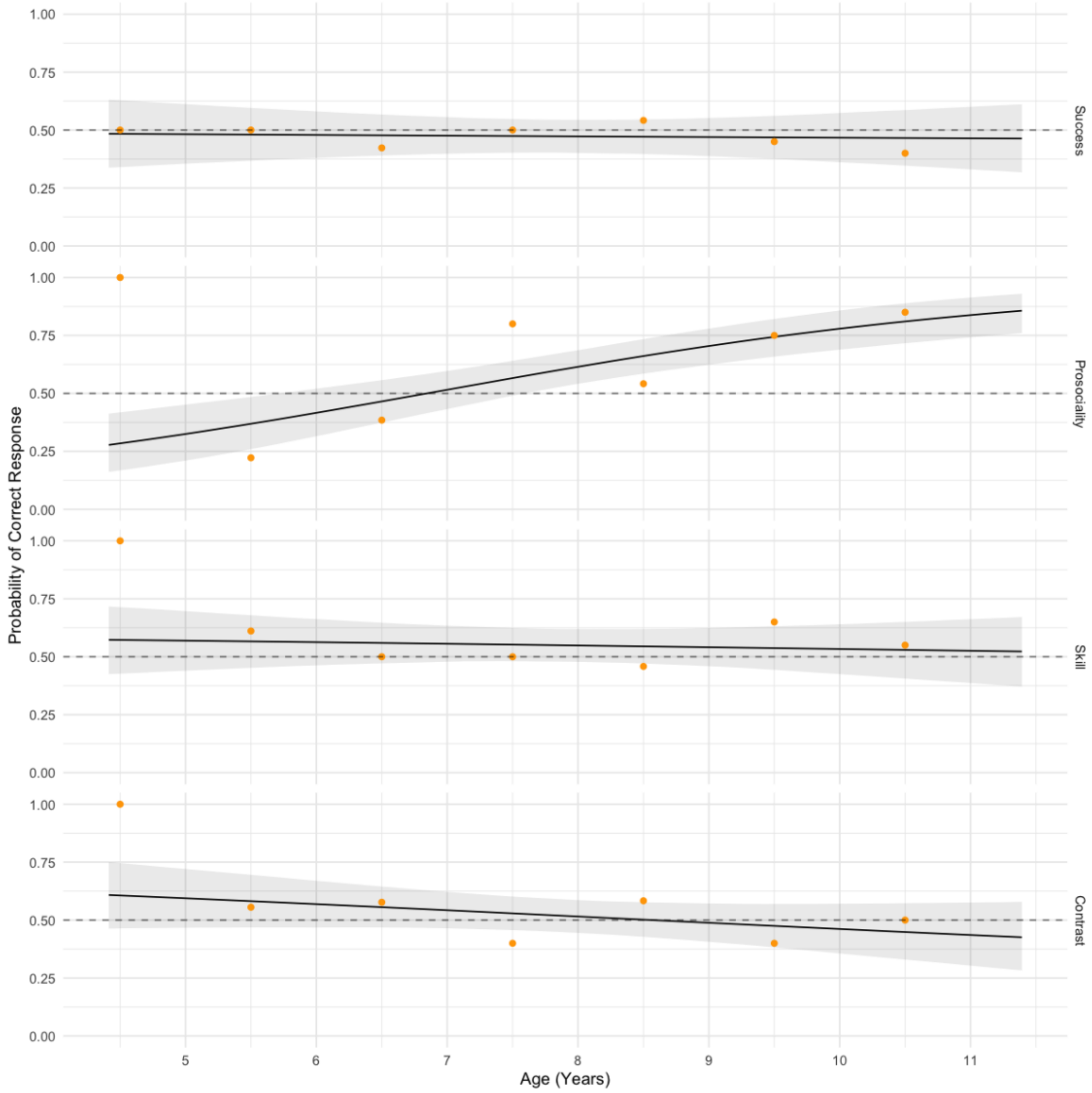

Figure D.3. Age differences in children's response accuracy in Experiment 3.2a. The proportion of correct responses is indicated by orange dots. The black line indicates the mean of the parameter estimates; the shaded area represents the 89% CI. The rows show results for the different trial types (from the top: Success, Prosociality, Skill, Contrast).

Experiment 3.3

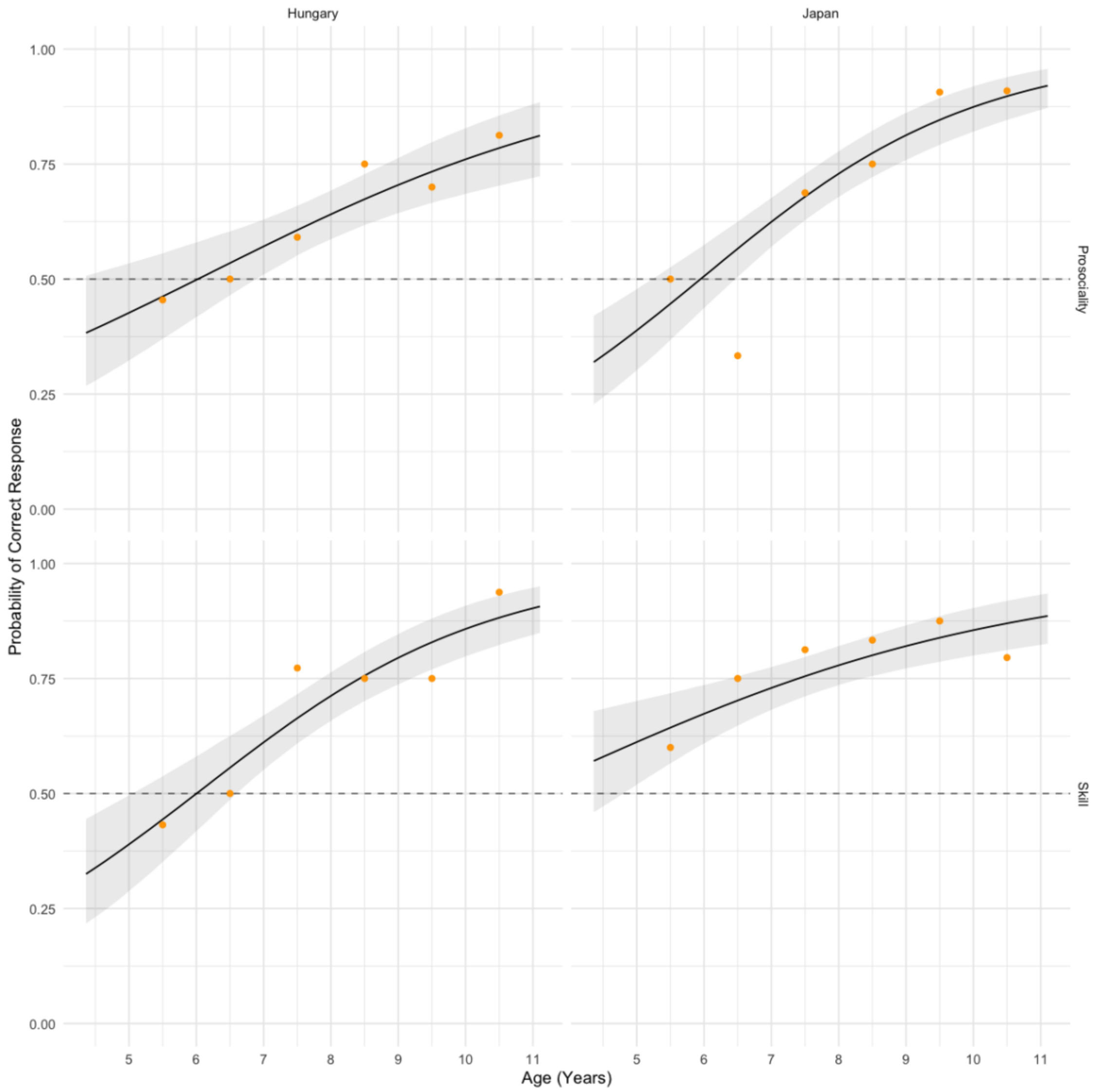

Figure D.4. Age differences in children's response accuracy in Experiment 3.3. The proportion of correct responses is indicated by orange dots. The black line indicates the mean of the parameter estimates; the shaded area represents the 89% CI. The left column shows the results for Hungarian (3.3a), the right for Japanese children (3.3b). The rows show results for the different trial types (from the top: Prosociality, Skill).

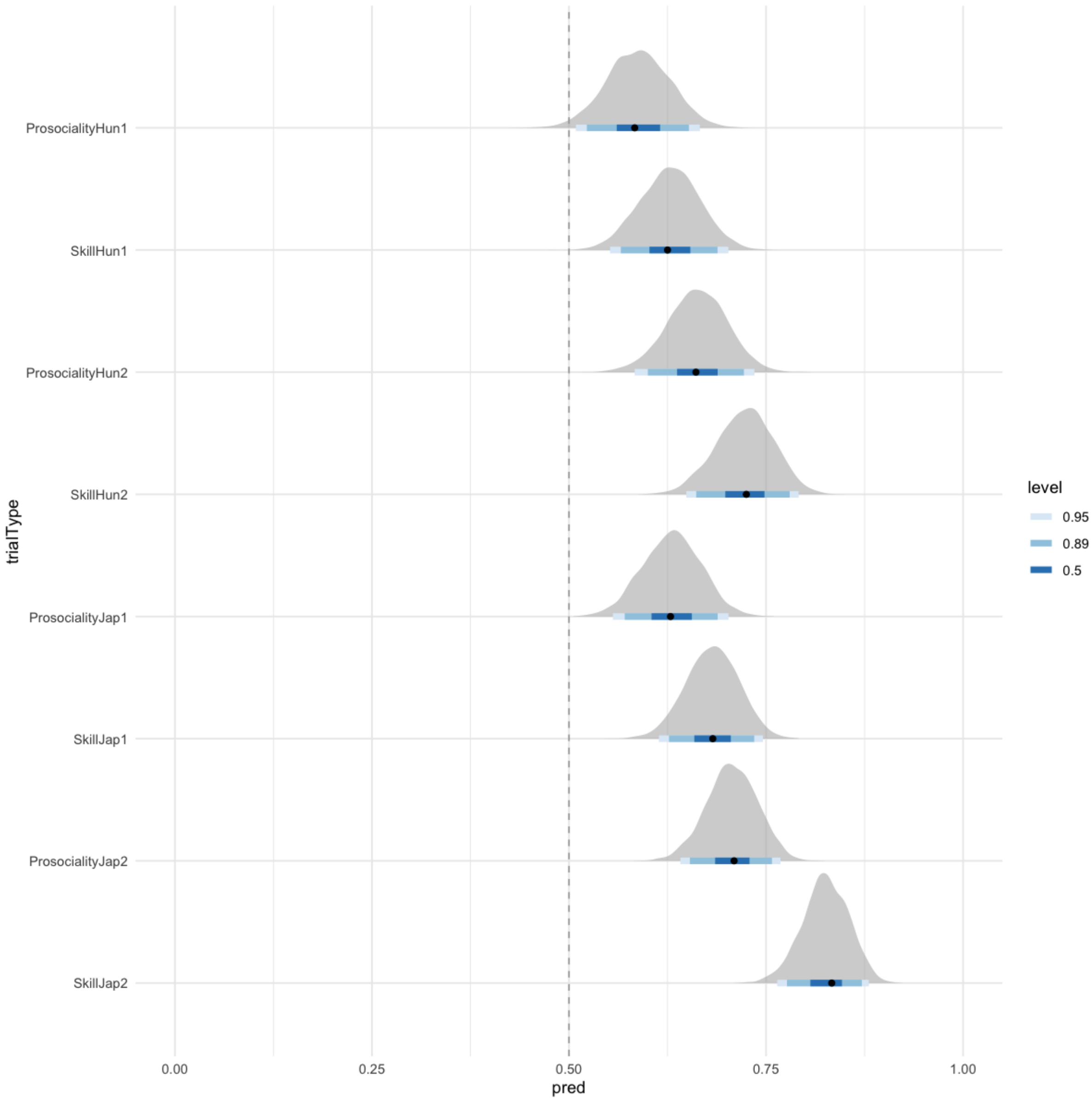

Figure D.5. Children's response accuracy as a function of Block (i.e., trials were presented in the first or second block) in Experiment 3.3. The proportion of correct responses is indicated by black dots. The light blue line indicates the 89% CI of parameter estimates, the shaded areas represent the posterior distributions. The rows show results for the different trial types (Prosociality, Skill), split by block (1 or 2) and country.

Experiment 3.4

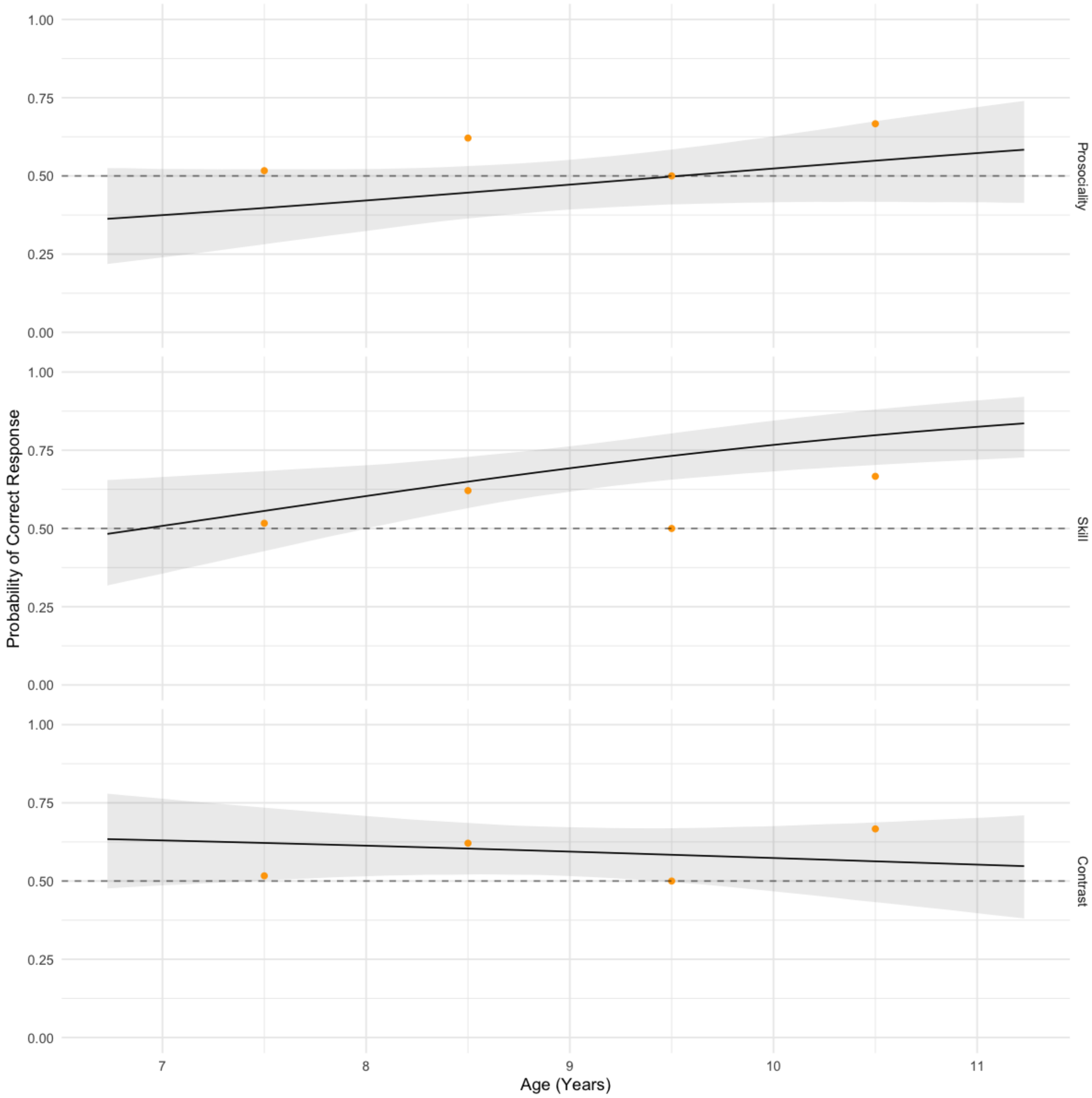

Figure D.6. Age differences in children's response accuracy in Experiment 3.4. The proportion of correct responses is indicated by orange dots. The black line indicates the mean of the parameter estimates; the shaded area represents the 89% CI. The rows show results for the different trial types (from the top: Prosociality, Skill, Contrast).

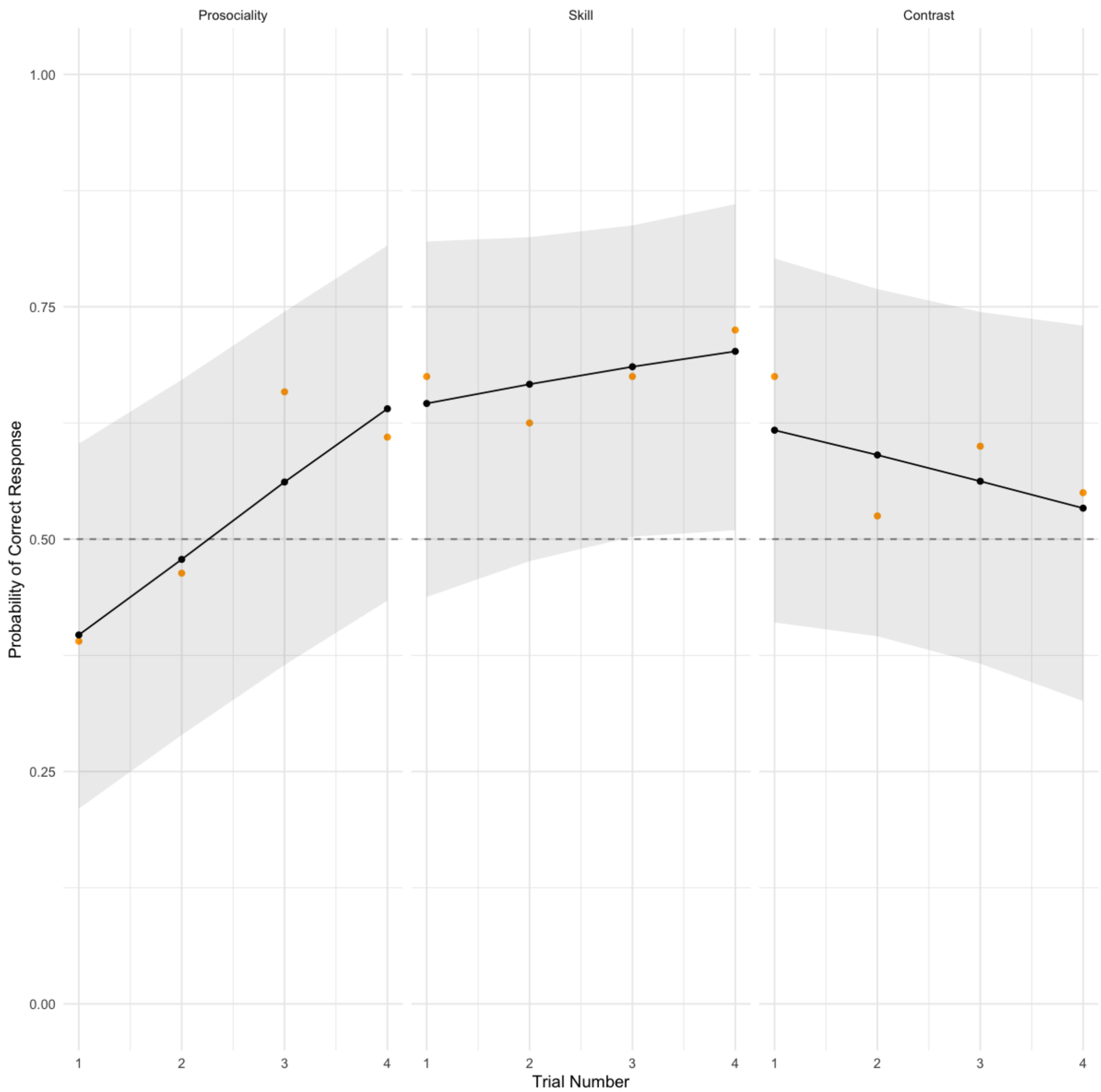

Figure D.7. Children's response accuracy as a function of Trial position in Experiment 3.4. The proportion of correct responses is indicated by orange dots. The shaded areas represent the 89% CI. The columns show results for the different conditions/trial types (Prosociality, Skill, Contrast).