Curiosity And Social Interpretative Schemas: Their Role In Cognitive Development

Kíváncsiság és társas értelmezési keretek: szerepük a megismerés fejlődésében

Akadémiai Doktori Értekezés

Ildikó Király

2018

Table of Contents

Acknowledgments

This dissertation represents a collection of ideas that has emerged as an outcome of joint thinking of a group of people, and also stands as a bridge between generations of researchers with the same interest and shared inspiration. So there are a lot of people without whom it could not have been realized.

First of all, I want to thank my mentors, Anikó Kónya, György Gergely, Gergely Csibra, and Csaba Pléh, for their trust, patience and time spent on sharing thoughts with me.

To an even degree, I owe special thanks to Fruzsina Elekes, Kata Oláh, Dorka Kampis, Erna Halász, Marci Nagy and Gábor Bródy, my former PhD and MA students for all their excellent work they did and especially for being tolerant and helping me in staying motivated and in discovering new areas of research.

I am also very grateful to my co-authors, friends and colleagues, Eszter Somogyi, Kata Egyed, Ágnes Lukács, Anett Ragó, Ágnes Kovács, Ernő Téglás, Fanni Tolmár, Mikolaj Hernik, Dezső Németh, Gerda Szalai, Bálint Forgács, who encouraged me constantly during the writing of the dissertation.

I have to thank Réka Pető, Krisztina Peres, Ágoston Galambos, Judit Kárpáti, Nazli Altinok, Anna Berei, Iulia Savos, Júlia Baross, Krisztina Andrási, Renáta Szűcs, Lívia Elek, Andrea Hegedűs, Zsuzsanna Üllei and Zsófia Kalina (current PhD students, students and research assistants) for their reliable work in designing, conducting and organizing the experiments.

The research presented in the dissertation would not have been possible without the support and encouragement of my institution, ELTE PPK Psychology Institute and without the financial support over the years from the Hungarian Scientific Research Fund (OTKA 109352; OTKA 116779) and from the János Bolyai Research Scholarship of the Hungarian Academy of Sciences. I owe special thanks to the Institute for Advanced Studies and the Department of Cognitive Science at CEU for inviting me to get acquainted with an inspiring academic atmosphere. I would like to thank the Momentum Program of the Hungarian Academy of Sciences for lending me wonderful opportunities to continue.

I am especially thankful to the child participants and their families for their indispensable contribution.

Certainly, I am very grateful to my family for being always there.

Curiosity and social interpretative schemas: their role in cognitive development

Exposition

The importance of the developmental perspective

The cognitive machinery of humans has been the target of investigation from the very beginning of scientific thought. At the same time, the ontogenetic emergence of the cognitive system, the development of cognition became a research topic only with the foundation of psychology as a discipline. The description of development in terms of its normal course together with potential variability and exceptable delineations can in itself contribute to the understanding of human behavior. Yet, the psychological approach brought an interest in the developmental trajectory of human cognition itself. This approach can aid the understanding and identification of the building blocks and developmental milestones of human intelligence and also might open a window on the understanding and identification of human-specific capacities. The promise is a better understanding of behavior organization as a whole. (Piaget, 1962; Karmiloff-Smith; 1995; Pléh, 2010)

One of the most persistent enigmas of development was, what kind of capacities and contents are part of our biological inheritance and what kind of capacities and contents emerge as a consequence of the influence of the environment, and being active part of it. In other words, how rich and prewired is our biological inheritance? This ever-dispute has been reformulated as a consequence of extensive research: since both sources of information seem to be necessary for knowledge accumulation, these days no-one argues for one or the other extreme, for pure innatism (claiming that the cognitive machinery is already born with ideas, knowledge and beliefs, no knowledge can be derived exclusively from one's experiences) or for pure empiricism (stating that knowledge is mainly derived from sensory experiences). Briefly, even if there is an emphasis on learning, in the sense that experiences have key importance in knowledge acquisition, knowledge could only be derived if there are some presumptions on what should be learnt and selected. In light of this, there is still a debate both on the division of labour between the two sources, and also on how to think of innate principles.

Piaget, founder of the scientific investigation of development *per se* was amongst the first ones who understood that both of the above two sources of information are necessary for knowledge acquisition, and introduced a theory proposing a dynamic interplay between innate principles and learning in the environment: his model postulated pre-wired, general cognitive capacities that drive the process of adaptation to the environment through active participation (Piaget, 1985). His theory is still very influential and recent theoretical models, while criticizing some of his concepts, still build on the basic claims of them.

We will briefly introduce those distinguished explanatory models that attempt to solve the developmental enigma. The dominant frameworks in the field take it seriously to revisit Piaget. The discussion he initiated considers the claim that (i) cognition and specifically its emergence is guided by innate, universal and general principles; (ii) while the unfolding process of the cognitive system is driven by the constructive role of the active and developing subject.

Introduction

Current models of cognitive development in a nutshell

Modularism. Jerry Fodor (1983) proposed that the human mind is composed of innate neural structures, *modules,* which have distinct filogenetically developed functions. Modules are equipped with domain specific processes that are inferential, however these processes are encapsulated, meaning that they operate mandatorily on a certain kind of input, while they do not communicate with other modules, or psychological systems during their operation (so modules in this sense are similar to reflexes). He supposes a central processing part as well that employs various logical manipulations on contents different in origin, and takes care of the relations between the specified domains. Having a modular system yields the benefit that processing is fast, and the output of such modules is simple. The original ideas of Fodor received support from the work of Elisabeth Spelke (2000) and Susan Carey (2011) among others from the field of developmental psychology, who argue that the astonishing competencies of infants, who are able to reason about things like numerosity, goal-directed behavior, or the physical properties of objects in the first months of life cannot be entirely explained by domain-general learning processes. These abilities appear to be too sophisticated to have been learnt via associative learning, given the short time and also the constraints of infant's perceptual, attentional, and motor competencies. As an example, Spelke (2000) argues for specific knowledge domains with unique cognitive signatures that are designed to overcome persistent problems in the environment. She postulates five innate core knowledge organizations: the system of objects, actions, numbers, space and social relations. These independent modules provide preconditions for further development.

Carey (2011) in her persuasive work on the *Origins of Concepts,* - similarly to Spelke -, suggests systems of core cognition, as innate starting point for conceptual development. She advocates that the innate primitives should not be limited to perceptual or sensory-motor representations, rather she postulates conceptual core representations. Indeed, the representations in core cognition are the output of *innate input analyzers*, domain specific information filtering processors. These inbuilt capacities make it possible to create conceptual representations out of perceptual and spatiotemporal primitives. Conceptual development starts from three domains of core cognition: the *domain of objects*, including representations of causal and spatial relations among them, the *domain of agents*, including their goals, communicative interactions, attentional states, and causal potential, and the *domain of numbers*.

In Carey's view, there could be other innate conceptual representations as well, including innate central (non-domain specific) processes that are responsible for computing causal relations from observed patterns (like statistical dependence among events). She hypothesizes, as an alternative account, that there may be specific aspects of causality that are part of distinct core cognition systems (she refers to physical causality in the domain of objects, or intentional causality in the domain of agents).

Carey believes that conceptual development consists of episodes of qualitative change: new representational systems emerge that have more expressive power than core cognition and are also incommensurate with core cognition and other earlier representational systems. The learning mechanism that could contribute to the emergence of novel representational systems is the gradual understanding of concepts with the help of bootstrapping. Language acquisition represents a special example, as it is supposed to become possible by domain specific learning mechanisms, comprising a language acquisition device (LAD). However, language acquisition plays a special role in conceptual development as words could be applied as semantic placeholders for novel concepts in the process of bootstrapping. Therefore language acquisition and conceptual development are closely related. While the representations in core cognition support language learning, providing some of the meanings that language expresses, language creates representations whose format is novel, and non-iconic any more (like it was in core cognition). This representational change makes possible the integration of the concepts from core cognition with the rest of language. This theoretical angle emphasizes that language learning necessarily shapes thought, as the learning of language opens new representational resources that might comprise concepts that were previously unrepresentable.

Neuroconstructivism. Annette Karmiloff-Smith (1995), a student of Piaget, challenged the view of domain specific modularism, and emphasized the availability of innate domain relevant biases, instead of encapsulated contents and processes. In her view, our cognitive system has innate predispositions and tendencies that are prepared for an expected environment. So, these biases are innate, but they only become fully operational through interaction with the environment. The emergence of the cognitive system in development is not the mere unfolding of an existing system, rather an interplay of the initial biases, the active thinker and its environment.

This theoretical account emphasizes the context-dependence of development, as the construction of representations depends on the exploration of the environment performed by the individual. Considering the continuous pro-activity and its modulatory effect on mental representations, a progressive specialization can be posited in relation to the past and current learning environment.

In a developmental perspective, this progressive specialization is a result of the process of *representational re-descriptions*. Representations of the world are becoming themselves object of cognition gradually, which leads to a much more formalized conceptual pattern. In this framework, information becomes *available as knowledge* for processes other than those in which it was originally embedded, because the organization repeatedly re-represents the information stored previously for special purposes *in* the system.

Modul-like structures arise as a result of experiential learning through innate biases. Alterations, improvements in the existing (neural) representations occur as small adjustments that the environment requires. This process is dynamic and always present: the brain's plasticity grants a nest of ever-changing representations by virtue of interacting pro-actively with the environment. The output of each re-description is stored and may influence behavior differently. The most important implication of this theory is that any current representations are the optimal outcome for a specific environment.

The dominant theories of cognitive development thus share the assumption of innate principles, however consider the nature of them slightly differently. According to modularism, the innate principles are ready-made and enclosed in separate, domain specific core modules. Innate input analyzers translate perceptual primitives into conceptual ones. Likely, neuroconstructivism suggest more general, yet domain related innate principles, biases or predispositions that shape information processing.

These models provide subtle descriptions on the gradual enrichment of knowledge base. Carey introduces two innate mechanisms that can account for information organization from the very beginning: domain specific content-mapping is subserved by innate input analyzers, while innate central causal representations could help the combination, linking of domains. The question remains though, whether there are other innate principles beyond those responsible for content mapping or central manipulators. Are there innate principles that could help information selection and inference based learning as well?

Also, the current models of development advocate constructivism, in the form of representational enrichment. Modularism sees the mind as a set of modules that make it able to understand causal relations with external objects, and form mental states with contents that are about things in the world. The central processing part is responsible for developing logical relations between domains, thus context independent contents: as a result, through interacting with the environment, the mind develops into an independent and rather context-free information processor. The discontinuity hypothesis of Carey (2011) underlines that the initial capacities, and representational systems of core cognition are qualitatively different from, and hence discontinuous, with explicit, verbally represented, intuitive theories that arise as a result of qualitative change. Still, neuroconstructivism proposes the process of gradual specialization as a result of the continuous interaction between the environment and the pro-active mind. In fact, by always building on preexisting representations, these representations become increasingly context bound and specialized (modularized).

Similar in these models is that they handle the environment in a general sense, and do not accredit special relevance to the social environment. All of the current models of cognitive development introduced share the viewpoint that they interpret and explain the development and knowledge acquisition of the *individual mind*. Nonetheless, as discussed above, these theoretical angles share the presumption as well that some natural preconditions for development are innate as a consequence of filogenetic factors. Thus, it is striking that the special role of social context is undervalued despite the fact that living and acting in social groups represented an overwhelming benefit for survival in human evolution (Caporael, 2007). In order to consider the potential role of social context in cognitive development, we turn to the work of Bruner.

Social constructivism in cognitive development. According to the theory of Bruner (1996) the development of cognition follows a gradual change and decontextualisation in the dominant format of representations: from indexical (action-based), through iconic (image-based), into symbolic (linguistic) representations. However, Bruner proposes that representational systems emerge from social-communicative interactions, emphasizing the role of everyday teaching or demonstration situations. His work rests on the assumptions that active, motivated participation in the social life of a group, as well as meaningful use of language involve an interpersonal, intersubjective, collaborative process of creating shared meaning. Bruner, as an outstanding representative of social constructivism highlighted the need to consider the role of knowledgeable, social partners in scaffolding the cognitive machinery.

The above introduced models of development recognize the importance of the social context, yet handle it as a specific domain, like the core domain of social relations (Spelke, 2000), or the core knowledge domain of agents (Carey, 2011). Neuroconstructivism takes a step further, emphasizing that the specific environment in which the individual develops has constraining effects on the possible neural representations, as a result of possible limitations of the experience, (this process is called ensocialment, see Westermann et al., 2007). Yet alterations in social context are managed in a similar vein as alterations in the physical world. The approach of Bruner (1996) is exceptional in the sense that he recognizes the fact that most of the knowledge of a human individual is learnt through interactions, thus through indirect sources, and not necessarily from direct experience. We follow in his footsteps and examine the possibility that humans have special innate capacities that make them able to exploit the routes of indirect learning, *learning from others*.

The overall aim of the dissertation is to refine and integrate the assumptions of the above introduced, current models of the *development of cognition* with the help of empirical research, in order to investigate the early availability of innate *inferential* principles. In addition, we address the need to elaborate these models by adding a special emphasis on the ways how *being social* might play a role in cognitive development.

Social Curiosity: the role of socially induced generative models in cognitive development

The dissertation builds on empirical research that tries to provide evidence on the availability of core interpretative schemas or *generative models* that enhance information selection and inference based learning from very early on. While previous theories proposed innate domain specific modules (Fodor, 1983), core knowledge (Spelke, 2000), domain related dispositions, (Karmiloff-Smith, 1995) or even innate central causal representations (Carey, 2011), we presume prewired or early available interpretation frameworks or generative models, which rather function as "innate algorithms". First the concept of generative models will be introduced; then the constructive role of social partners in knowledge acquisition will be briefly outlined, with a special emphasis on the advancement even in the instrumental and conceptual domain.

The common features of such innate algorithms are that they provide a structure both with (1) default expectations and (2) inherent well-formedness conditions, so they are *inferential* in nature. The default expectation slots prescribe the set of features the model is applicable for: when a preset cue occurs, the model is triggered to operate. In other words, generative models also prescribe cues that trigger their operation. In addition, with the help of inherent inferential principles, information that is directly not available in the situations can be revealed and understood. Inferential principles guide information selection, and facilitate the interpretation of an observed event or behavior in a format that will allow predictions for upcoming similar instances. Thus, generative models are different from innate input analyzers (see Carey, 2011), as the latter only compute representations of one kind of entity in the world in a dedicated manner, and this process does not involve abstraction and inference based prediction.

Teleological stance has been introduced by Gergely and colleagues as an innate algorithm for helping young infants understand goal directed actions (see Gergely &

Csibra, 2003; Gergely et al., 1995). Their theory introduced the conceptual structure of early available generative models. (Indeed, in the Thesis, I present my work as an empirical contribution in support of the availability and extended functional role of this generative model in development).

The complex nature of understanding goal directed actions and teleological stance as an operator in this process can be illustrated by one of the violation-of expectation studies of Gergely et al (1995). In that study, twelve-month-olds were habituated to a computer-animated event with a distinct end-state, in which a small circle approached and contacted a large circle ('goal') by jumping over ('means act') an obstacle separating them ('situational constraint'). During the test phase the situational constraints were changed by removing the obstacle. Infants then saw two test displays: the same jumping goal-approach as before, or a perceptually novel straight-line goal-approach. They looked longer (indicating violation-of-expectation) at the old jumping action, but showed no dishabituation to the novel straight-line goal-approach. This pattern of looking time allows the presumption that infants dishabituated to the perceptually similar, old movement because it seemed to them an inefficient 'means' to the given 'goal' in the novel situation, as there was no obstacle to jump over. However, in the alternative event infants did not dishabituate despite the perceptual novelty of the straight movement, possibly because this action appeared to them the most efficient means to the given goal in this new situation.

Gergely and colleagues claim that the above results show that infants use an inference based model, the Teleological Stance for the sake of understanding and predicting actions around them. They interpret their own results as an indication that by 12 months infants can (1) interpret others' actions as goal- directed, (2) evaluate which one of the alternative actions available within the constraints of the situation is the most efficient means to the goal, and (3) expect the agent to perform the most efficient means available. The expectation to act efficiently is served by the *rationality principle* itself. The rationality principle arises from the normative assumption that intentional actions are essentially functional in nature. In this sense, the rationality principle provides a criterion for 'well-formedness' and can be applied as an 'inferential principle' that can direct the construction of action interpretations, at the same time. Overall, *Teleological Stance* allows to form two important presumptions even by young infants: (1) actions function to bring about future goal states, and (2) goal states are realized by the most efficient means available in a given situation.

A recent extension of teleological stance is the *Naïve Utility Calculus* (Jara-Ettinger, Gweon, Schulz & Tenenbaum, in press). This calculus originates from - and is compatible with the earlier model of goal attribution, that is based on the principle of action efficiency and rational agency (introduced above, Gergely & Csibra, 2003), but this naïve utility calculus (NUC) extends and goes beyond accounting for inference based goal attribution and choice of means action in a number of significant respects.

The central claim is that humans, from early infancy, interpret the intentional actions of others in terms of a ‘naïve utility calculus’, applying the core assumption that others make decisions and choose actions by maximizing their (expected) utilities (i.e., the rewards they expect to obtain relative to the costs they expect to incur). Mainly, NUC is a scientific account of people's intuitive theory of how people act, while NUC does not require that agents actually compute and maximize fine-grained expected utilities.

In essence, NUC is a generative model that specifies how costs and rewards determine behavior. This model suggests the following way of action selection. When agents decide whether to pursue a goal or which goal to pursue, they estimate the expected utility of each goal. This is calculated by (1) estimating the rewards the agent would obtain if she completed the given goal, and then (2) subtracting the estimated cost she would need to incur to complete it. Through this process, agents build a Utility Function (UF) that maps possible plans onto expected utilities. Agents then select and pursue the plan with the highest positive utility. So, the core inferential assumption of NUC is that agents are utility maximizers.

Indeed, NUC is supplemented by the core knowledge structures about basic domains and concepts of the physical and psychological world (e.g., knowledge about objects, forces, action, perception, goals, desires, and beliefs). In this angle, this calculus can integrate the output of different core cognition domains. Relatedly, in the domain of *social relations* (see Spelke, 2000) or *agents (*see Carey, 2011), NUC supports a wide range of core social-cognitive inferences, and is already present in infancy and persists stably through adulthood. So, the advantage of NUC is that - beyond physical actions - it accounts for a wide range of social situations as well, such as sampling-sensitive and preference judgements, communication, pedagogy and social and moral evaluation (Jara-Ettinger et al., 2015a; b)

However, as the model builders also emphasize, the real world has several difficulties that the idealized NUC presented above cannot handle. The formal description provided for the NUC model is highly idealised and assumes fully deterministic scenarios where agents have perfect information about costs and rewards. Though it is supposed that the NUC likely applies to abstract and more complicated situations as well, in the present format it cannot account for how more complex social factors influence behavior selection, (e.g., cases where cultural norms are in play).

All in all, NUC is a generative model that offers an account for an intuitive theory that specifies how costs and rewards determine how people act. This model allows both children and adults to *infer* complex mental states as causes after observing the behavior of others: to infer their beliefs and desires, their long-term knowledge and preferences, and even their character, who is knowledgeable or competent (Jara-Ettinger et al., in press). This model, however, builds on personal experience as the source of information, observers apply their own assumption of expected utilities, despite the fact that these estimates may be inexact. In other words, this is a model of how individuals learn about others. Therefore, it cannot explain, as we emphasized above, the role of social partners in knowledge accumulation and organisation, yet make possible to identify agents with specific purposes and knowledge.

The objective of the dissertation is to expand on the role of social partners in learning in a more general sense. Since humans always experience being *together* with conspecifics, knowledge *about the partners* is essential, we agree. This knowledge guides behavior and also makes it possible to successfully interact with each other. However, we underline that social context also represents a group of individuals who are knowledgeable experts with full-blown cognitive machinery. This raises the opportunity that observation of partners, the active behavioral exchange, including different forms of interactions could carry information in two ways for the novices: information about the environment and information about the partners themselves (as shown by the already introduced generative models). The question of the dissertation is whether there are generative models dependent on the availability of partners and their active contribution in learning or not.

Our model presumes a basic motivation, *information seeking or curiosity* (Silvia, 2012). While we accept the existence and relevance of core knowledge domains, we propose that there are generative models that exploit the presence of knowledgeable partners in order to fulfill information seeking motivation: novices are ready to identify knowledgeable expert partners around them. With the help of potential prewired inferential frames, they filter and follow the behavior of others to maximize their epistemic benefit. The hypothesized generative models, in this sense, are bootstrapping mechanisms to get ‘novices’ started on the long path leading to an eventual expert understanding of intentional minds and actions together with the complex environment. Meanwhile, knowledgeable partners act as observable models and interactional partners and through these roles actively modulate and gradually refine the training of the novices.

In our view, the *teleological stance* (note that this formula is compatible with the recent model of *naive utility calculus*) makes it possible to *identify agents* and learn the rules of their basic behavior. We will take this generative model as the main, initial model in our investigations. The active modulatory role of knowledgeable partners might also provide an account how knowledge from different domains could be integrated. The interplay of the suggested learning modes, *learning from* and *learning about* others could contribute to the organization of information in a hierarchical manner: acquire information in order to share with others, or value information based on its source (being shared or not). We thus posit that the most important catalyst of development - fueled by curiosity - is to become experts in social contexts, namely sharing the knowledge of partners.

Hence, the dissertation considers the proposal that generative models bootstrap the development of cognitive competencies *by the active contribution of expert others*. In this process, experts represent cues both for triggering these models and also for boosting their refinement and integration. In order to explore this general supposition, we have developed four lines of empirical investigations. First, we introduce briefly the state-of-the-art debates in each subfield of cognitive development, and then, we outline the specific thesis points that answer the introduced problems in light of the general proposal of the dissertation.

I. Social Learning.

Humans have evolved specialized cognitive mechanisms adapted for the acquisition of organized knowledge, culture (Boyd and Richerson, 1985). The first line of research focuses on the description of such basic mechanisms, the capacities children employ to learn from social partners. The minimal definition assumes that any type of information gain in the individual that occurs as a result of observing the behavior of a social partner is *social learning* (Want and Harris, 2002).

The formats of the social learning capacity are different in complexity: from simple *enhancement* where the presence of the social partner only boosts the encoding of a location or a stimulus, to complex *imitation* where the learner follows the observed behavior with high fidelity. Even in the case of imitation the mechanisms that subserve the process is disputed (Want and Harris, 2002; Call, Carpenter és Tomasello, 2005). Note that while imitation is the process where following an observed behavior results in attainment of a similar outcome, this could happen without being aware of the outcome as the goal of the behavior– so here we are talking about blind imitation. However, imitation could happen as a consequence of planning ahead to reach the outcome, the goal, when we are talking about insightful imitation. In the latter case, the agent forms knowledge of the means-end relation of the observed behavior: a specific movement will result in a certain endstate, and as such, re-enactment the observed action sequence is initiated by a goal concept in mind. The challenge is (in accordance with the general question of the dissertation) to clarify whether observational learning is rooted in associative processes or is supported by generative, interpretative models in human infants.

Piaget (1962) illustrated the first qualitative shift (the emergence of symbols) in the development of cognitive processes by the description of delayed imitation: imitation after a delay can only emerge as a consequence of having a symbolic representation formed at the time of the perception of a motor behavior, and consequently, this symbol is used to build a novel motor program, in a novel spatial and temporal context. Call, Carpenter és Tomasello (2005) enriched this conceptual approach by claiming that the symbolic representation of an observed behavior contains an element on the means action and also an element on the goal or outcome separately, although these two elements of behavior are activated together in the course of re-enactment. There is an advantage of encoding these two elements apart: their segmentation permits flexible use of information: facing or even imagining the outcome of an action induces a search to select the means adequate to attain it.

In opposition to this account, Meltzoff (1990) describes imitation as a blind copying mechanism, subserved by the process of automatic intermodal transfer. During this transfer, the perception of a behavior automatically activates the matching motor program and the behavior is executed. Thus, in this view, a single associative representational unit is formed. The prerequisite of successful intermodal transfer - and so imitation per se - is the similarity of the perceived actor and the perceiver herself/himself. This theoretical angle gave rise to several alternative approaches that assume a *common code* for percieved and performed actions, and share the assumption that this common code is associatively formed as a consequence of repeated co-occurrences of certain end-states and actions. Imitation is induced by any cues on end-states that are already associated with a given motor program, without any inference at play (Prinz, 1997; Paulus et al, 2011a, b; Heyes, 2012).

The challenge for any account –including the above approaches - is to provide explanation for the acquisition of new behavioral skills on the one hand, and for their flexible (but functionally adequately constrained) generalization and selective reproduction in appropriate novel contexts on the other. According to Call, Carpenter, Tomasello (2005) ‘blind’ mimicry (i. e., resonance-based automatic motor copying) cannot account for the infant’s capacity of reproducing the adult’s actual behavior in their appropriate functional contexts, since reproduction remains dependent on learning through repetition and matching contextual cues. While flexible, functional use of an imitatively learnt behavior implies an understanding of the intentional state underlying the behavior, namely understanding of the other’s mental intentions and reasons behind his or her action choice (Tomasello & Carpenter, 2007).

We provide a model that can answer the above challenges without postulating the full-blown capacity of understanding intentions: we propose that imitative re-enactment is an output of an analytic learning process. In order to learn novel actions infants need to be able to identify and filter the relevant target actions. We suggest that during this selection process, the child as observer monitors the efficiency of the performed action in relation to its goal. The observed behavior induces the encoding of the novel behavior as a function of the evaluation of a given action sequence as *efficient* for goal attainment. Specifically, when the situational constraints justify the novel method chosen for the action, the observer will turn to use the most efficient means available in the situation, and won’t imitate the modeled behavior with high fidelity (Gergely, Bekkering, Király, 2002). In our view, this selection process is governed by the *teleological stance,* a competence that enable children to interpret actions as goal directed (Gergely & Csibra, 2003; Csibra et al, 2003). Moreover, for this reason, we argue, infants need to rely on the active inferential guidance provided by the social partner, that is served by the generative model of *Natural Pedagogy*. Natural Pedagogy

Theory argues that human infants are prepared to recognize a single act of demonstration *as* communication, and have the expectation that the content of the communication represents knowledge that is generalizable along some relevant dimension to other objects and other situations (Csibra & Gergely, 2006, 2011).

This developmental model is introduced in details in the first line of investigations (I. Social Learning, Theses 1-3), where we present the role of *teleological stance* and *natural pedagogy* as generative models in inference based selective social learning. We tested the predictions of this model in contrast to that of alternative theories, and we used delayed deferred imitation paradigms.

II. Memory development.

Childhood amnesia refers to the phenomenon that humans in general are unable to recall specific, personal events from their first years of life. The so called 'very first memories' appear to be recalled from an age when the recaller was around 3 and a half years old (or even older), irrespective of the recaller being 8 or 70 years old at the time of retrieval (Eacott & Crawley, 1998). This phenomenon received special attention in the field, and obtained several explanations. It has been proposed that childhood amnesia occurs as a consequence of the late development of the hippocampus and related cortical areas (Nadel & Zola-Morgan, 1984). This neuromaturational account has received significant support and is still a dominant view in the field of infant memory (Bauer & Leventon,2013), however, there is controversy how neural maturation actually influences the functional characteristics of memory competence, since cumulative evidence support that very young infants perform ordered recall in different tasks, and this competence is incompatible with the immature neural substrate view (see Mullaly and McGuire, 2013). Indeed, there are several alternative proposals on the level of psychological interpretation: the mismatch of retrieval strategies applied before and after skilled language use is developed could be responsible for the unavailability of early person specific memories (Simcock & Hayne, 2003); the development of cognitive self concept might also contribute to the qualitative change in memory organisation (Howe & Courage, 1997), and the narrative socialization of joint remembering could also provide a plausible account for becoming able to recall distinctive, specific memories (Fivush & Nelson 2004). Despite the richness of explanatory frameworks, all of the above-mentioned approaches build on the observation that there is a shift in the functions of memory that is best described as the emergence of personal memories that provide the bases of autobiographical memory.

At the same time, these approaches disagree whether children are able to form contextually rich memories or not *before* the emergence of autobiographical memory. With introducing the term of *episodic memory,* Tulving (1972) emphasized that the main function of episodic memory capacity is retaining information together with its spatial-temporal context. According to this theoretical angle, when a memory contains components that can answer the What? Where? and When? questions, this memory should be episodic in nature. In this sense, episodic memory could be available already for young children and even for other species as well (Clayton, Bussey & Dickinson, 2003; Russel & Hanna, 2012). However, Tulving (2005) reconsidered his early suggestion on what the essential criteria of episodic memory were. He proposed that recollection is a process that elicits the retrieval of contextual information pertaining to a specific event or experience that has occurred. Furthermore, a key property that makes 'recollection' possible is autonoetic consciousness. That is a special kind of consciousness, which enables an individual to be aware of the self in a subjective time during the act of remembering. Tulving emphasizes that *episodic memory* (in this angle) is late developing, qualitatively different from other types of memory and human specific. Indeed, with this definition he introduced a serious problem for research. Namely, investigation of recollection is troublesome in the absence of refined linguistic skills (Clayton, Bussey, & Dickinson, 2003). Tulving (2005) suggested a test that requires recollection in his view, so, this protocol could be the litmus test of nonverbal episodic memory. He called this test the spoon test and is derived from an Estonian children's story. In this tale, a young girl dreams about a party where her favourite pudding is being served. Unfortunately, she is unable to eat any of the pudding because, unknown to her, guests were requested to bring their own spoons. The next night after this event, before going to bed, the little girl hides a spoon under her pillow, possibly with the idea on mind that she returns to the party in her dreams. According to Tulving (2005), the act of placing the spoon under the pillow provides behavioural evidence that the little girl remembers the dream (episodic memory) and has prepared for the same event to occur again in the future (episodic foresight). (Scarf, Gross, Colombo & Hayne, 2011).

The debate on the early availability of episodic memory – as well as on the explanation of childhood amnesia - is ongoing. The followers of the 'what-where-when' conceptual approach argue for the early emergence of episodic memory, and emphasize that the development of episodic memory involves only quantitative changes, mainly in capacity constraints (Bauer et al, 2000; Hayne, 2004). According to a minimalist version, episodic memory is indeed a form of re-experience as it takes over two things from the original experience: its spatiotemporal context and its 'synthetic unity'. However, this experience lacks self-awareness of the original experience, and therefore associative (and not conceptual) in nature (Russel & Hanna, 2012). In contrast, the advocates of the conceptually rich approach to episodic memory claim that there is a qualitative shift in memory development: the emergence of self-reflection gives rise to episodic memory as children start to understand time as a causal factor (McCormack & Hoerl, 2001; Povinelli et al, 1999).

We present a specific theoretical angle with respect to the above debate, claiming that memory competencies measured in the first years of life can be best described as dominantly semantic in nature – lacking any characteristics that would point to being specific in content,. The semantic/noetic memory bias – we posit - is a byproduct of the need of novices to search for generics, information with predictive values. This claim evidently presumes that memory formation is dependent on online behavior selection. Young children not only interpret ongoing behavior using their model of teleological stance, but this online behavior interpretation allows them to filter information and possibly encode only the selected elements. This process could supply fast mapping of relevant information, and so, relatedly, could be a dedicated mechanism that subserves the formation of generic, essentially semantic memories as well, even after the first encounter with an event. This bias has special advantages for those with limited knowledge base. Collecting generic information provide valuable benefit for adjustment in upcoming situations. Consequently, it could be a byproduct of this information filtering bias that young children seem to be unable to encode the distinctive, specific features of an event.

In the second line of studies we investigate the above proposal, namely whether teleological stance plays a role in the formation of enduring semantic memories by encoding only the selection of goal relevant information. (II. Memory: Theses 4-6.). We apply the method of delayed imitation in our experiments.

A related important question remains: what explains the emergence of episodic memory then? The dominant view for the qualitative shift in memory organization and relatedly for the emergence of episodic memory argues that joint remembering with expert partners implant novel representational strategies for encoding and organizing memories (Fivush & Nelson, 2004). Discourse on joint experience highlights for children that different observers can have different perspectives on the exact same event. The emphasis on such differences in recalled details sensitizes children to detect the distinctive properties of memories. Additionally, sharing memories in interactions inherit a special representational structure for embedding such distinctive features, and children could be taught how to narrate their memories.

In our approach, however, episodic memory might function to support flexible integration of information. Supposedly, episodic memories might enable the reconsideration and updating of the inferential consequences of a past event in light of some newly acquired information (see Klein et al, 2009). As an example, it is much easier to change our views, beliefs if we are able to recall the context of acquiring them, especially, if we realize that our view arose from a misunderstanding, or was acquired from an unreliable source. Yet, this flexible revision process is only possible if the original experience is likely available.

Let us turn again to the suggestion of Tulving, the spoon test as litmus test for episodic memory. There are several attempts to implement the 'spoon test' (Atance és Sommerville, 2013; Scarf, Gross, Colombo, & Hayne, 2013; Suddendorf, Nielsen, & von Gehlen, 2011). For example, in the study of Scarf and colleagues (Scarf, Gross, Colombo, & Hayne, 2013), 3- and 4-year-old children could find a hidden treasure case in a sandbox. When they explored the treasure case, children realized that it was locked. Later, children returned to the lab and were asked to select one of three props (including a key to unlock the treasure box and two distractor objects) to take with them to the sandbox scene. Three-year-olds performed above chance if the delay between the events was 15 minutes or less: they tended to take the key, while 4-year-olds performed above chance even with a 24-hour delay. This test, (similarly to any spoon test, seemingly) demonstrates that children can identify and select information as relevant for an upcoming or reoccurring event based on a memory of a past event. Note that this achievement can be based on the encoding of some semantic information derived from that past event ("sandboxes have locked treasure cases") without retrieving a specific past event. This information can then be recalled when some related novel information (key for unlocking) is obtained in a context that promises revisiting the original scene. In our view, the so-called spoon tests show evidence on selective encoding of predictive information, rather than episodic retrieval. We propose an alternative way to test whether young children can apply episodic memory, especially for information updating. (II. Memory Thesis 5. and III. Thesis 8).

III. Navigating in the social world: Naïve Psychology

The main objective of the third line of investigations is to further analyze the epistemic role of interactional partners. The availability of an expert partner in communication offers the possibility to learn information filtered by the knowledgeable partner (see also I. Social learning). In addition, social partners represent a special category to understand: interacting with a social partner opens an opportunity to learn about 'others' as well. In essence, observation of the behavior of the social partner provides information both about the world in general and about the partner specifically. We introduce the possibility that young children are equipped to endorse information from expert social partners in two different ways – in an object-centered mode and in a person-centered mode. These two modes of information selection are intertwined in the course of communication in social contexts. Our studies were aimed to disentangle them. Furthermore, in accordance with the general goal of the dissertation, we suggest that the above two cognitive goals are subserved by different generative models that we try to grasp with the help of experiments.

The first line of investigations (I. Social Learning) introduces the role of the interactional partners in guiding the acquisition of the world (learning FROM others). In the third line of investigations, we focus on the ways children learn ABOUT others, their person specific features and special capacities (we will call this Naïve Psychology).

Mindreading is the ability that allows humans to predict and interpret others' behavior based on their mental states, thus, this is the capability that makes humans successful in most of their social interactions. Consequently, it has been studied extensively both in the domain of developmental science and adult social cognition. However, the underlying cognitive architecture and the exact mechanisms that enable humans to use such powerful abilities are still unclear. Mindreading sometimes occurs online, spontaneously, and implicitly without any deliberation. On other occasions, it involves planned, explicit, verbally expressed and often offline reasoning about mental states. While there is a wide consensus that human adults can perform complex belief inferences and use sophisticated mental representations in an explicit manner, we know much less about the processes that are spontaneously invited by implicit mindreading.

Hence, one of the biggest puzzles in ToM research is to understand what underlying processes make it possible to successfully track beliefs online and later explicitly update these beliefs that are attributed to other people.

In support of the role of Naïve Psychology in interactions, it is argued that if novices can successfully track the belief-representations of other social agents and the potential differences between their own knowledge and that of the partner, this will allow them to use such representations to successfully communicate and learn about relevant knowledge that is socially distributed among different agents (Keil et al., 2008). Despite this presumption, and in light of empirical evidence, there is a controversy that dominates the field: while adults use mental state reasoning in their everyday lives with great ease (Friedman & Leslie, 2004; but see Apperly, Riggs, Simpson, Chiavarino & Samson, 2006, who claim that fast, efficient use of Naïve Psychology is inflexible and limited in adults as well), children, in contrast, seem to have difficulties in explicit reasoning about complex mental states before the age of 4 (Wellman, Cross & Watson, 2001). In the recent years, nevertheless, a new line of studies revealed evidence that infants already in their second year of life seem to possess mindreading abilities (Kovács, Téglás, & Endress, 2010; Surian, Caldi & Sperber, 2007; Onishi & Baillargeon, 2005); for instance, two-year-olds are able to anticipate where an agent will look for a target object based on her false beliefs (Southgate, Senju & Csibra, 2007). These findings provide evidence in favor of a recently articulated account proposing that humans from very early on track other's belief (Kovács, 2015), like adults, (Kovács et al., 2010).

In our view, online, real time belief computation enables adequate, fast behavioral adjustments in social situations. Therefore, this system should be in place already in infants, and necessarily, should be used by adults as well. The challenge is, in our view, to grasp whether children are able to monitor simultaneously other partners as sources for two types of information: to learn about the world FROM others and also to learn ABOUT others per se. We posit that there should be a dynamically applied switch between the above two modes of learning approaches. While we have identified teleological stance as the main model that subserves learning from others (I. Social Learning), in this line of investigations we are in search for the features of the basic generative model, *Naï*ve *Psychology*. We investigate whether there is a spontaneous monitoring of the partner's attentional focus that could be the central mechanism of the inferential capacity to track others' beliefs. This supposition of ours emerged from the opinion that the main function of naïve psychology could be to help interactional partners to set common ground and detect partitioned knowledge flexibly. Relatedly, we explore the context and cues that possibly trigger switching between the two modes of learning: we test whether natural pedagogy indeed biases the context to learn rather the generics about the world and not the specifics about the social partner.

We have developed novel looking time and behavioral tests for our empirical purposes (III. Navigating in the social world: Naïve Psychology; Theses 10-12.).

IV: Cultural learning and Naïve Sociology

The milieu of ubiquitous interactions is claimed to be a key factor in the emergence of the unique complexity of human culture. In Bruner's words, there is no mind without culture, and no culture, without mind (Bruner, 2008). The question that emerges from this interdependence is how to account for social group members' capacity to achieve and maintain coordinated knowledge of their environment. In other words, what are the basic social and psychological factors that allow for the emergence of cultural or shared knowledge (Shteynberg, 2010)? The fourth line of research investigates the function and emergence of the competence used for learning about and from groups of people, usually studied as categorizing social partners, and called Naïve Sociology.

It is a delicate problem of psychology: what kind of benefit Naïve Sociology could represent for humans? The central problem we investigate is whether Naïve Sociology simply arises as a result of cumulative perceptual differentiation of input information, or it reflects a systematic semantic information selection, fueled by generative models.

With respect to Naïve Sociology, it is well documented in the social psychology literature that humans are excessively sensitive to social categories (Tajfel, Billig, Bundy & Flament, 1971). Research in developmental psychology has provided evidence that this social category formation is present in young infants as well; for example, language (and accent) has been identified as a reliable indicator of social similarities for infants (e.g. Kinzler, Dupoux & Spelke, 2007; Kinzler, Corriveau & Harris, 2011). According to some proposals, the human mind has evolved a special domain or module to form and represent social categories (Spelke & Kinzler, 2007; Sperber & Hirschfeld, 2004). Nevertheless, what advantages this capacity may provide is under debate. For example, it has been suggested that in the ancient environment a module dedicated to processing social kinds may have served to help tracking coalitions and potential coalitional partners (Cosmides, Tooby & Kurzban, 2003) or that it helps people to make sense of the complex structures of human societies (Hirschfeld, 1996). We propose a novel explanation, namely, that the function of Naïve Sociology is the identification of representatives of shared knowledge.

The ultimate challenge is how social groups transmit their rules and norms, the cultural phenomena, to those who are novices? Recent approaches have identified that both the causal opacity, and the uncertainty regarding generality and sharedness of knowledge are common problems with which any learner is necessarily confronted (Csibra & Gergely, 2011). Natural Pedagogy Theory describes a solution for this challenge (Csibra & Gergely, 2006, and see also *I. Social learning* in this dissertation). Firstly, novices should be prepared to recognize instances of information exchange, and show readiness to engage in interactions as recipients. In line with this, recent research support that very early on infants pay special attention to communicative signals and treat them as referential (Senju & Csibra, 2008). Moreover, novices have to have a default expectation that the content of the communicative episode, the demonstration itself, represents shared knowledge. It is also argued that these expectations are universally available (Hewlett & Roulette, 2016), and applied by everyone around, utilized by adults as well (Csibra & Gergely, 2011).

However, we suggest that while this preparedness for shared knowledge in communication is very important in order to set the basic framework of generic knowledge (we will discuss the relevance of this theory in research line I. and III.), it cannot explain in itself the development of understanding cultural variances. In our view, the 'sociality competence' should include components that both help novices to get access to the shared representational space, as well as help expert members to maintain access, and allow flexible contribution to it. In order to develop into a competent member of society, novices need to acquire culturally relevant knowledge from culturally competent individuals. For this, the novices need to select culturally knowledgeable and reliable sources of information.

On the one hand, this requires the tracking of the interactional *partner's access to potential knowledge*, and on the other, special sensitivity to *cues* that point to *the shared knowledge base*.

In our fourth line of investigation we explore the proposal that the seemingly more complex function of Naïve Sociology, namely the systematic information selection for organization and categorization of social partners has important epistemic advantages for humans, most prominently at the beginning of their life. We investigate whether children are equipped with capabilities to identify reliable sources of information for the sake of fast cultural knowledge acquisition.

THESES

I. SOCIAL LEARNING

Social learning, imitation especially, is the main tool for acquiring instrumental and social instrumental knowledge. Imitation -as an information transmission mechanism - is served by two interpretative schemas or generic models that grant inference based learning – teleological stance and natural pedagogy. Teleological stance enables the interpretation of goal directed, efficient actions. Natural pedagogy induces the expectation that the partner intends to teach relevant and novel information in the communicative situation. Teleological stance and natural pedagogy — while being two separate cognitive adaptations to interpret instrumental versus communicative actions— work in tandem for learning socially created instrumental knowledge in humans.

1. Imitation as a tool for social learning enables infants to enrich their individual learning strategies through the observation of their partners.

Imitation as a cognitive apparatus for information transmission is not a unitary and blind mechanism. In this process, the observer monitors the efficiency of the performed action in relation to its goal. An observed behavior triggers selective social learning as a function of the evaluation of a given action sequence as efficient for goal attainment. When the situational constraints justify the method chosen for the action, however there is a simpler alternative at the time of reproduction, the observer will use the most efficient means available in the situation, and won't imitate the modeled behavior with high fidelity. This model is called the theory of rational imitation. Our results provide evidence that already 14-month-old infants are able to infer the most efficient means in the situation, and selectively use it to attain the goal. This pattern of results highlight that social learning is inference based and selective.

As an extension to this theory, it has been shown that when the modelled behavior involves tool use, and the situation highlights the benefit of using it, children not only learn the tool use but also inhibit the prepotent means, the direct manipulation with hands. Moreover, it has been confirmed that the function of imitation and social learning is *epistemic*, despite the fact that the source of information is a communicative partner.

2. thesis. Children with atypical developmental trajectory also exploit the teleological stance: they are able to interpret goal directed actions, but not intentions.

The ability to apply the interpretative schema of rational, goal directed action is not only available for normally developing children. We investigated the ability to understand goals and attribute intentions in the context of two imitation studies in low-functioning, nonverbal children with autism. Down syndrome children and typically developing children were recruited to form matched comparison groups. In the two sets of simple action demonstrations only contextual indicators of the model's intentions were manipulated. The results suggest that nonverbal children with autism attributed goals to the observed model, but did not show an understanding of the model's prior intentions even in simplified, nonverbal contexts.

3. Thesis. Imitation (in terms of underlying mechanisms) is not restricted to the competence that children are able to interpret actions as goal directed. In addition to the application of the analysis of teleological stance subserving the selection of their own action, children also profit from natural pedagogical stance. They interpret the model's behavior as teaching communication that induces the search for novel and relevant information in them.

The model of rational imitation received significant criticism, challenging the view that high fidelity imitation occurs and is explained by higher order interpretative processes. Rather, from an alternative theoretical angle, it has been suggested that imitation is a result of simple associative learning: imitation occurs as a result of motor resonance induced by the observed behavioral pattern (Paulus et al., 2011a, b). In order to answer this challenge, we proposed a novel elaborated version of the theory, *the model of relevance based selective emulation*. In this model, we highlight that the main problem with the initial model of rational imitation was that it essentially made the idea of appealing to the rationality principle unfalsifiable; when infants did not reproduce the demonstrated action, it was treated as evidence of the application of the rationality principle, and when they did reproduce it, it was also interpreted as evidence for the operation of the same inference. Overall, the original explanation of selective head touch imitation purely in terms of the application of the rationality principle fails to capture all of the relevant aspects of the selective imitation phenomenon.

The new approach to the selective imitation phenomenon attempts to explain why infants reproduce actions when those are obviously subefficient. We found that 14-month-olds reenacted a novel arbitrary means action, the head touch, only following a communicative demonstration, and not in an incidental observation context. Our results revealed that infants' tend to reproduce communicatively manifested novel actions.

Recent results of ours further confirm that the sensitivity to teaching intention allows the observer to segment the demonstrated behavior into subgoals. This segmentation means that the observer is able to discriminate the teaching context and the related behavioral elements from the concrete content, the relevant behavioral steps that constitute a novel subgoal-goal relation in the demonstrated behavior. As such, the new account of selective, imitative learning of novel actions shows how the teleological stance and natural pedagogy—two separate cognitive adaptations to interpret instrumental versus communicative actions—are integrated as a system for learning socially constituted instrumental knowledge in humans

II. MEMORY

The main interpretative schemas of social learning not only help the interpretation of ongoing, real time behavior; additionally, they also shape the selection of valid information for longer periods of time.

4. thesis. The interpretative schema for understanding goal directed actions plays a significant role in forming predictive, general memories.

Early memory competences can be characterized by a bias towards collecting general memories: the main purpose of our experiments was to show that even in cases when young participants encounter an event only once, the formed memory can be best described as a filtered sequence of events for successful goal attainment. In other words, despite the fact that the source of encoding is a certain, so specific event, the representational format of memories resembles to general memory representations, and not episodic (distinct, specific) ones.

In fact, our studies highlight the concern that the method of deferred imitation is not a conclusive method for grasping early episodic memory competences. When we observe that a child re-enacts an event that she has seen only once before, this performance not necessarily entails that this behavior is rooted in a memory that is *episodic in nature*, meaning that is distinguished as pointing to a *specific* event. Indeed, this is the central argument used to claim that the development of declarative memory is quantitative and linear, the main formats of memory representations are available from the very beginning (Bauer et al, 2000; Hayne, 2004).

Our results reveal that 2 and 3 year old children recall solely those event steps that were necessary to attain the demonstrated goal in the given context and omit those elements that were not. This pattern of results suggests that children rely on the teleological stance as an available action interpretation schema in order to interpret the ongoing event and additionally, to encode those components that will help them to re-identify a similar event with a similar goal. This can be seen as a bias towards collecting information as the generics of events in order to become able to predict behavior (even after only one single encounter).

5. thesis. The inflexibility of early memory competences can be considered as the byproduct of searching for generics, and consequently being unable to encode the specifics of an event. Albeit, the emergence of episodic memory may serve the function to reorganize and update knowledge based on mnemonic retrieval.

There is a general concern with respect to imitation, a method used as a nonverbal format of free recall. Free recall implies that the person retrieves and shares all the information that is accessible to them, has been encoded and is thus part of the memory representation. While it is inevitably true for imitation that it builds on a representation that is formed following an observed motor behavior (Meltzoff, 1990), a serious interpretation constraint emerges from the fact that the recall process itself – the reenactment – can be guided by the same selection pressure as the learning phase or the immediate retrieval phase. In other words, during the phase of retrieval, the person can apply the interpretative schema of goal directed action again, and thus will perform the most efficient means, regardless of having a detailed memory (including any non-relevant components) of the original event. Given this possibility, it cannot be confirmed, and at the same time it cannot be ruled out that there is an episodic memory as a source for re-enactment.

The challenge for imitation as a method for recall can be formed like this: we cannot tell during which memory process information selection takes place or plays a significant role. In the framework suggested by our theory of relevance based selective imitation (see theses 1 and 3), the prediction in this case is biased - selection processes should be guided online by the help of generative models, so during the phase of acquisition guided by the expert teacher.

We investigated whether the memories and memory processes of children are flexible enough to adapt to a situation with changed contextual parameters. That is, what is the nature of selection observed in imitative behavior – does it happen online, during the encoding phase or is it a flexible process and selection occurs at retrieval?

In our experiment we presented 2-year-old children events with changing contexts, where the different contexts either verified or disproved the use of a novel tool as opposed to a familiar action to reach a certain goal. First, we tested what the typical copying mechanism during immediate re-enactment is: do children selectively imitate the tool use in the condition where the context requires it, and leave it out when it is doesn't? We also tested whether in deferred imitation, if the context changes children adapt their retrieval processes flexibly: whether they integrate an action element into their actions when it becomes necessary, and relevant (despite being irrelevant at the initial learning phase).

Results suggest that at immediate test children imitate the use of the tool selectively if the model previously directed their attention to the situational constraints determining the relevance of the tool use. This pattern of finding confirms our model of relevance guided selective imitation. Interestingly, at the second time when they faced a different context in which a previously irrelevant step turned to be relevant, they did not change their strategy: whatever they did the first time, they repeated the second time.

This pattern of data implies that young children's memory processes seem to be inflexible: it is not just that they „forget" the element in question and therefore cannot use it in the changed situation (in the condition where the successful goal attainment requires the recall of that specific element), rather we argue that they have not encoded the original event in a specific format and that is why they cannot recollect the original event including specific details.

6. thesis. The access to shared semantic (cultural) knowledge can be optimized

through the memory organization of the individual. Consequently, different memory organization strategies are hypothesized that are dependent on the actual capacities and capacity constraints of the individual.

The main role of memory is to allow access to a broader knowledge base in support of the ideal accomplishment of actual goals. Undoubtedly, this broader knowledge base is getting wider with development and includes cultural information as well.

With investigating the memory competences of young children we wanted to prove that their mnemonic strategies indicate the motivation to maximize availability of shared knowledge with the help of a given strategy.

The dominant view of children's memory is that it is slow to develop and inferior to that of adults. In a study we contrasted four-year-old children and adults in a test of verbatim recall of verbal material. Parents read a novel rhyming verse (and an integrated word list) as their child's bedtime story on ten consecutive days. A group of young adults listened to the verse, matching the exposure of children. All participants subsequently performed a verbatim free-recall of the verse. Four-year-olds significantly outperformed both their parents and the young adults. There were no significant differences in the ability to recall the gist of the verse, nor the integrated word list, ruling out concerns about differences in engagement or motivation. In our view, verbatim recall of verse is a skill that is practiced by children, as they rely on oral transmission in their culture .

The acquisition of reading and writing empower children to have access to the so called extended memories of culture (Donald, 2010). When the accessibility of wide knowledge and alternative knowledge organization provided by extended memories open up for them, text memory, literal remembering for what has been heard becomes unnecessary. The essential message of this thesis, in other words, is that the mnemonic strategies of an individual emerge from and are related to the motivation for having access to shared accumulated knowledge (of a group) and maximize the array of accessible information.

III. NAVIGATING IN THE SOCIAL WORLD: NAÏVE PSYCHOLOGY

Sociality is rooted in the capability that humans can *learn from and about a partner* as well in the course of communication. In the process of learning, the communicative partner can play two separate roles: could be the main

source of information, and also the target of observation and learning. We would like to highlight that from very early on, children are able to exploit both roles of the partner in order to enrich their knowledge on different fields of the environment in an integrated format.

7. thesis. In their early months, children primarily learn FROM others and encode the content learnt as predictive, generic information.

The benefit of social learning can be characterized by the description that young children learn about objects in the world through the observation of others' object directed behavior. This form of learning allows the transmission of information filtered by a knowledgeable partner. Thus, social learning exploits the mental achievements of an expert partner and supports fast acquisition of complex information.

The most prominent example of this learning situation is social referencing. When an infant first encounters a novel, unfamiliar, and even surprising object, she first checks the behavior and emotional expression of a close partner and tries to read her/his emotional, referential expression whether the target object is approachable or avoidable (Walden és Ogan, 1988). We have reinterpreted this phenomenon arguing that the social partner's emotional expression not simply modulates the infant's behavior through emotion regulation, but also conveys information about the object itself. Supposedly, valence information is transmitted that is already known by the expert individual and unknown for the infant novice partner: in this situation, the young observer learns generic, predictive information about the object quickly from the partner. So, it can be supposed that in social referencing situations infants assume that the other's object-directed emotion manifestations conveys universally shared information about the referent that is available to *all* individuals.

In a study with 14-month-olds we have provided evidence that infants learn and remember the object valence information, and not the personal, consistent preferences of partners observed. Based on this, we argue that infants rely on their so-called '*object-centred*' interpretations to form generalized expectations that all others will perform the same kind of object-directed actions that are appropriate given the objective valence quality of the referent that the infant's newly formed object representation contains.

Indeed, object-directed emotion expressions provide two types of information: they can be understood as communicative signals that convey culturally shared knowledge about referents that can be generalized to other individuals (as we have shown above as object-centered interpretation), or they can convey the expressers' *person-specific*, subjective disposition toward objects. In a further study, by presenting object-directed emotion expressions in communicative versus non-communicative contexts, we demonstrated that 18-month-olds could flexibly assign either a person-centered interpretation or an object-centered interpretation to referential emotion displays. The findings indicate that infants are prepared to learn shared knowledge from nonverbal communicative demonstrations addressed to them at an early age. While they are also capable to learn about the partner: they attribute person specific dispositions and preferences.

8. thesis. The possibility that children are able to apply both the object centered and the person centered interpretative schemas when interpreting an other person's behavior, invites the reconsideration of findings on mindreading or naïve psychology competencies in young infants: we propose the primacy of the object-centered approach. In addition, we postulate that the emergence of flexible use of interpretative models might contribute to improvement in performance.

There is an ongoing debate on the availability, and nature of early mindreading (Naïve Psychology) competences (Perner and Ruffman, 2005; Rakoczy, 2012). The so called implicit minreading tests rely on robust behavioral measures, like looking time or anticipatory looks applied in simple object choice context. In most of the cases, these simple object choice scenarios can be interpreted by both the object-centered and by the person centered interpretative frames as well, and these possibilities cannot be disentangled in the classic, existing approaches. However, in the framework of mindreading, there is a potential hierarchy between the above two interpretative schemas: the person specific interpretation could be described as an object-centered content bound to a specific person, distinctively.

Revisiting the findings in the domain of implicit mindreading, recent findings claimed to suggest that infants understand others' preferential choice and can use the perspectives and beliefs of others to interpret their actions. The standard interpretation in the field is that infants understand preferential choice as a dispositional state of the agent, so learn about that specific agent. In our view, it is possible, however, that these social situations trigger the acquisition of more general, object centered knowledge. We propose that early mindreading processes lack the binding of belief content to the belief holder. However, such limitation may in fact serve an important function, allowing infants to acquire information through the perspectives of others in the form of universal access to general information.

In our view, binding mental states to specific persons, as a consequence of the simultaneous availability of object centered and person centered interpretative schemas, is a developmental milestone with respect to the emergence of mindreading capacities, as it makes possible the individuation of mental states approximately at the age of 18 months .

The integrated, flexible use of interpretative models might contribute to a more and more elaborate and complex performance, even inviting different domains. As an example, we show that a further developmental achievement in mindreading is brought about by the involvement of emerging episodic memory competences.

A study of ours aimed to investigate the contribution of episodic memory to mindreading, by proposing two different processes of belief attribution: prospective (online) belief tracking and retrospective inference based belief attribution. The experiment explored whether 18- and 36-month-old children could flexibly use episodic memory to attribute or update a protagonist's belief. After a displacement event that an experimenter witnessed wearing sunglasses, 36-month-olds correctly attributed to her a false belief, when finding out after that event that the sunglasses were opaque. They successfully ascribed a false belief based on this new information and behaved accordingly. In contrast, 18-month-olds behaved as if the sunglasses were transparent. This suggests that 18-month-olds cannot use their memories to (re)compute a belief retrospectively, although they performed well when they could track false beliefs prospectively. This dissociation reflects that 18-month-olds rely primarily on prospective (online) belief tracking, while 36-month-olds can also flexibly compute beliefs retrospectively, based on episodic memories, well before they pass explicit tasks.

9. thesis. The primary function of mindreading (naïve psychology) is to enable the observer to monitor in the real time, say spontaneously and prospectively, the knowledge state of the partner. This capability facilitates both learning about the partner in the here and know and also evaluating the possibility whether learning from the partner would contribute to a valuable shared representational space for the long term or not.

Both children and adults are able to compute and monitor the perspective of their interactional partner, fostering the rich net of social interactions. However, due to the mindreading system's limited capacity perspective taking was argued to occur spontaneously in a restricted manner, only for all or nothing type, so level-1 perspective taking, but not for the more sophisticated level-2 perspectives. We proposed that level-2 perspectives (containing aspectual information) could also be computed spontaneously if participants have reason to assume that the partner is indeed aware of the objects' aspectual properties. In a reaction-time study where adults, and school aged-children answered a simple number verification task, the partner's inconsistent perspective was found to interfere with reaction times providing evidence for spontaneous level-2 perspective taking.

Overall, we propose that this capability is a necessary prerequisite for the establishment of *common ground*, both for the ongoing interaction, and also for the enrichment of a long term, knowledge base shared with a broader group.

IV. CULTURAL LEARNING AND NAÏVE SOCIOLOGY

The establishment of long-term knowledge base intertwins with the acquisition of cultural knowledge. The acquisition of shared, cultural knowledge demands that a) already young children should be able to recognize what is shared knowledge; b) already in the phase of acquisition children should be able to select information with respect to its potential relevance in relation to shared, cultural knowledge.

10. thesis. The emotional and affiliative motives behind social categorization are preceded by a cognitive, epistemic function of identifying culturally knowledgeable individuals both for (1) acquiring knowledge, and (2) obtaining access to and maintaining a shared representational space in the service of successful interactions.

Researchers in various fields of cognitive science have suggested that the human mind has evolved a special module to form and represent social categories (Spelke and Kinzler, 2007; Sperber and Hirschfeld, 2004). The open question is what the main function of such mechanism can be that is sensitive to any, even arbitrary grouping cues, yet is used immediately and dynamically at the same time.

To become a competent member of a culture one must be able to conform to the behavioural norms of that particular culture and possess culturally shared knowledge. Acquiring this knowledge is essential for successful social interactions. There is no information inherent to skin colour that would provide guidance in how to adjust our behaviour in interactions, rather such adjustments are related to cultural background. Identifying members of one's own cultural group is of special relevance during cultural transmission. Thus, naïve sociology, social categorization may play a role in cultural transmission, as it helps infants to select culturally knowledgeable and reliable sources of information (see also Kinzler et al, 2012).

We argue that any behavioural cue that indicates that a person shares the knowledge space of the target group/culture will lead infants to categorize that person as "in-group" which, in turn, will necessarily induce an epistemic trust (accept that person as a valuable information source and consequently suspend other individual learning strategies) towards any information that person may manifest later on, even if that is not yet part of the perceivers knowledge base.

11. thesis. Cultural knowledge represents an organized system. Supposedly, children understand that an attributed knowledge base might cover and unite different domains. Based on this, we hypothesize that children are able to form unified expectations induced by different cues on the background knowledge of a partner.

We propose that young children are sensitive to the cues - like tool use or language use - that reliably reflect the borders of shared knowledge and are able to build up common semantic categories of social groups with the help of integrating information induced by different 'knowledge' cues. The expectation of such integrated shared knowledge base should influence children's behavioral expectations as well.

An eye-tracking paradigm was designed to test whether two-year-old children differentially associate conventional versus non-conventional tool use with language-use, reflecting an organization of information that is induced by cues of shared knowledge. The results of the study suggest that children take the conventionality of behavior into account in forming representations about a person, and they generalize to other qualities of the person based on this information.

12. thesis. Naïve sociology contributes to the flexibility of social learning: when

receiving a novel piece of information from a carrier of shared knowledge (e.g. cultural group), that piece of information is treated as part of the culturally shared representational space that the child intends to acquire.

Our aims also included the exploration of how young children reason about social categories, and how this reasoning is reflected in their willingness to learn from agents of certain social groups. Three studies of ours suggest that young children are sensitive to the borders of culturally shared knowledge and confirm that social categorization in early childhood serves epistemic purposes, children selectively endorse information received form cultural in-group members.

The first study investigated whether toddlers would selectively imitate the actions of a demonstrator who exhibits familiarity with cultural practices over a demonstrator who consistently deviates from familiar tool-use practices. We propose that the familiarity of a tool using action as a cue serves to point out those potential informants that belong to the same social group as the novice, therefore relying on these signals in selecting teachers ensures that the knowledge obtained is valid within a particular social environment. Results show that 3-year-old children are more willing to copy the actions of the conventionally behaving model. This suggests that by the age of 3, children are adapt at determining whether someone's knowledge is appropriate within their own social context and thus worth acquiring.

The second study investigated 3-year-old children's learning processes about object functions. We built on children's tendency to commit scale errors with tools to explore whether they would selectively endorse object functions from a linguistic in-group over an out-group model. Participants were presented with different object sets, and a model speaking either in their native or a foreign language demonstrated how to use the presented tools. In the test phase, children received the object sets with two modifications: the original tool was replaced by one that was too big to achieve the goal but was otherwise identical, and another tool was added to the set that looked different but was appropriately scaled for goal attainment. Children in the Native language condition were significantly more likely to commit scale errors – that is, choose the over-sized tool – than children in the foreign language condition. We propose that these results show that children are more likely to generalize object functions to a category of artifacts following a demonstration from a cultural in-group member.

The third study investigated whether 4-year-olds used language as a cue to social group membership to infer whether the tool-use behavior of a model had to be encoded as indicative of the tool's function. We built on children's tendency to treat functions as mutually exclusive, i.e. their propensity to refrain from using the same tool for more than one function. We hypothesized that children would form mutually exclusive tool-function mappings only if the source of the function information is a cultural in-group person, as opposed to an out-group person. Participants were presented with tool-function pairs by a model who previously spoke either in their native or in a foreign language. During the test phase, children encountered new purposes, for what they could either use the demonstrated tools' color variant or another equally suitable, thus far unseen, alternative tool. In line with our predictions, children preferred to use the alternative tool for the new function only in the cultural in-group condition. The findings suggest that children restrict learning artifact functions from cultural in-group models.

Thesis points and empirical papers supporting them

I. SOCIAL LEARNING

Social learning, imitation especially, is the main tool for acquiring instrumental and social instrumental knowledge. Imitation -as an information transmission mechanism - is served by two interpretative schemas or generic models that grant inference based learning – teleological stance and natural pedagogy. Teleological stance enables the interpretation of goal directed, efficient actions. Natural pedagogy induces the expectation that the partner intends to teach relevant and novel information in the communicative situation. Teleological stance and natural pedagogy — while being two separate cognitive adaptations to interpret instrumental versus communicative actions— work in tandem for learning socially constituted instrumental knowledge in humans.

1.1 Thesis 1. Imitation as a tool for social learning enables infants to enrich their individual learning strategies through the observation of their partners.

Király I. (2009): The effect of the model's presence and of negative evidence on infants' selective imitation, *Journal of Experimental Child Psychology 102*, 14-25.

1.2 . Thesis 2. Children with atypical developmental trajectory also exploit the teleological stance: they are able to interpret goal directed actions, but not intentions

Somogyi E, **Király I,** Gergely Gy, Nadel J (2013): Understanding goals and intentions in low-functioning autism, *Research in Developmental Disabilities 34:* (11) pp. 3822-3832.

1.3 Thesis 3. Imitation (in terms of underlying mechanisms) is not restricted to the competence that children are able to interpret actions as goal directed. In addition to the application of the analysis of teleological stance subserving the selection of their own action, children also profit from natural pedagogical stance.

Király, I., Csibra, G., Gergely, Gy. (2013). Beyond rational imitation: Learning arbitrary means actions from communicative demonstrations, *Journal of Experimental Child Psychology 116.*, 471–486., DOI: 10.1016/j.jecp.2012.12.003.

Király, I., Egyed, K., Gergely, Gy. (in preparation). Relevance or Resonance: Inference based selective imitation in communicative context

dc_1353_16

Journal of Experimental Child Psychology 102 (2009) 14–25

Contents lists available at ScienceDirect

Journal of Experimental Child Psychology

journal homepage: www.elsevier.com/locate/jecp

The effect of the model's presence and of negative evidence on infants' selective imitation

Ildikó Király *

Department of Developmental Psychology, Institute for Psychology, Hungarian Academy of Sciences, Budapest 1132, Hungary

ARTICLE INFO

Article history:
Received 31 July 2007
Revised 15 June 2008
Available online 19 August 2008

Keywords:
Selective imitation
Ostensive communicative cues
Cognitive transparency
Failed attempt
Social context of imitation
Imitation and memory

ABSTRACT

This study demonstrated selective "rational" imitation in infants in two testing conditions: in the presence or absence of the model during the response phase. In the study, 14-month-olds were more likely to imitate a tool-use behavior when a prior failed attempt emphasized the logical reason and relevance of introducing this novel means, making it cognitively transparent for the infants. Infants also learned imitatively from the cognitively opaque (yet socially communicated) modeling situation, but to a lesser degree. Furthermore, the presence of the model as a social partner during testing influenced the performance of infants in that they were more likely to imitate the novel means when the model was present during testing. These results highlight the important interaction of interpretive schemas (e.g., causality, teleological stance) and social communicative cues in action interpretation guiding imitative learning.

Introduction

Children can copy what others do, and they are ready to learn from others through imitation as early as around the end of their first year (Barr, Dowden, & Hayne, 1996; Killen & Uzgiris, 1981). Based on recent comparative analyses of social learning, it has been argued that the term *imitation* should be applied only to cases in which infants understand the goal of the model's actions, copy the specific actions used by the model, and reproduce the modeled result (Tomasello, 1999; Want & Harris, 2001). Imitation can be contrasted with emulation, where children understand the goal of the model's actions and reproduce the modeled result but do not copy the specific means used by the model. It had been argued previously that most of the time infants *imitate* others' actions (rather than performing

* Fax: +36 13200496.
E-mail address: kiralyi@mtapi.hu

0022-0965/$ - see front matter
doi:10.1016/j.jecp.2008.06.003

emulation) even if a more efficient means is available (Horner & Whiten, 2005; Nagell, Olguin, & Tomasello, 1993; Want & Harris, 2002).

The growing body of evidence in the domain of social learning has also underlined that imitation is an integrated form of cognition that has at least two functions in that it serves (a) as a basic channel for communication (social function) and (b) as a fundamental tool of learning (cognitive function). According to Uzgiris (1981), the main motivation of young infants to copy others' behavior is primarily an epistemic cognitive motivation to support learning about events in the world, whereas older infants are more inclined to imitate others so as to satisfy social motivations, that is, to fulfill a social function of encouraging shared experience with others. From this perspective, "blind imitation," or persistence in copying the specific actions of others in infants, functions (in addition to its epistemic role) as the initiation and maintenance of interaction with the model.

Recently, however, a number of studies have shown that infants do not blindly imitate everything they observe; in some situations, they demonstrate the selective nature of imitation. A classic study of Meltzoff (1995) showed that 18-month-olds did not imitate what an experimenter actually did when she *failed* to achieve a goal; rather, they imitated what she had intended to do. Actually, infants copied the model's means *and* finished the action successfully. An exploratory study of Huang, Heyes, and Charman (2002) examined whether the reenactment of an unsuccessful attempt was due to reading the model's underlying intentions or to learning emulatively from object affordances or spatial contiguity. Interestingly, when all target acts produced within the 20-s response period were counted, 17- and 19-month-olds in the emulative learning and spatial contiguity conditions produced as many target acts as in the full demonstration and failed attempt conditions, revealing the potential effects of nonimitative learning processes. However, when only the infants' first actions were counted, those who observed the full demonstration produced more target acts. Further studies have confirmed that 18-, 24-, and 30-month-olds copy the means that the model uses unsuccessfully to achieve a goal (Bellagamba & Tomasello, 1999; Call, Carpenter, & Tomasello, 2005; Nielsen, 2006). Moreover, it was shown that prior exposure to a failed attempt followed by the full modeling of the target action sequence results in superior performance in the imitation of event components, as compared with receiving only a full modeling of the event (Carpenter, Call, & Tomasello, 2002; Want & Harris, 2001).

On the contrary, after exposure to a failed attempt, 12-month-olds do not use the modeled unsuccessful means but tend to emulate instead (Bellagamba & Tomasello, 1999; Nielsen, 2006). According to Nielsen (2006), in a situation where a prior failed attempt clarified the logic of a new tool-use behavior, 12-month-olds followed the specific actions of the model only when they were given a logical reason to do so; otherwise they focused on reproducing the outcome of the demonstrated actions (i.e., emulation). In contrast, 18-month-olds focused on copying actions and outcomes irrespective of the apparent logic of the model's behavior. Furthermore, 18-month-olds were more likely to copy the actions of the model when she acted socially than when she acted aloof, and 24-month-olds copied the actions of the model irrespective of the model's behavior but were more likely to produce the end result of these actions when the model acted socially.

In a similar vein, Brugger, Lariviere, Mumme, and Bushnell (2007) showed that infants' choice of what to imitate depended largely on their knowledge of the causal relationship between means and outcomes. They found that 14- to 16-month-olds were more likely to imitate the first action of a two-step event sequence when it was physically necessary to generate the effect. In another setup, infants were also more likely to imitate the action when it was socially cued. According to these authors' major conclusion, infants' knowledge of causality and their sensitivity to others' social signals both contribute to their tendency to imitate an action.

Another interesting and significant attempt to demonstrate and interpret selective imitation shows that infants imitate actions in terms of what they believe to be the demonstrator's goal (Bekkering, Wohlschlager, & Gattis, 2000; Carpenter, Call, & Tomasello, 2005). The authors' main result was that infants copied actions in terms of goals. When there was a clearly visible outcome or end state, infants interpreted the outcome as the goal and reenacted it to obtain the same outcome without copying the modeled means; however, when there was no visible end state as a potential goal, infants interpreted the particular action as the goal itself and followed the model's behavior accurately. Hence, according to the authors, infants interpret the action to be reproduced at the hierarchically highest level of action organization in terms of the overall goal.

Gergely, Bekkering, and Király (2002) demonstrated that 14-month-olds engage in "rational imitation." Using Meltzoff's (1988) head-on-box event, Gergely and colleagues tested 14-month-olds' imitative performance in two groups, varying the situational constraints of the model. In the *hands-occupied* condition, the model's hands were visibly occupied while performing the "head action"; in the *hands-free* condition, the model placed her visibly free hands onto the table before demonstrating the head action. When the model's hands were occupied, 14-month-olds were much less likely to imitate the head action (21%); instead, they illuminated the box by touching it with their hand, performing the most simple, easy-to-perform emulative response available to them. In contrast, when the model's hands were free but she still used her head to illuminate the box, 69% of 14-month-olds imitated her head action.

Rational imitation suggests that imitative learning is not an automatic "copying" process invoked by identification with the human actor. Rather, imitative learning is a selective interpretive process that involves the evaluation of the rationality of the means in relation to the situational constraints of the actor. Schwier, van Maanen, Carpenter, and Tomasello (2006) replicated the result of the above study with 12-month-olds using a different but analogous task. In these authors' interpretation of the results, infants' imitative behavior is guided by an understanding of others' intentions as rational choices between available means, and they can use this understanding in cultural learning contexts. In yet another variation of the head-on-box study (Gergely et al., 2002), where infants observed the head touch demonstration in an incidental learning context (with the model not performing any communicative cues toward the infants), infants were less likely to reproduce the head touch action themselves. Most important, the selectivity between the hands-free and hands-occupied conditions disappeared in the absence of ostensive communication (Király, Csibra, & Gergely, 2004).

Researchers in this field agree that the selective interpretive process of imitation is based on an understanding of the situation, whereas it also depends on social communication. In their recent theory, Gergely and Csibra (2005) made a convincing proposal to solve the problem concerning the fast learning of cultural knowledge and, at the same time, introduced imitative learning as its principal means. The model of natural pedagogy assumes that imitative learning is elicited by ostensive communicative cues accompanying the model's manifestations of cultural procedures and knowledge. The interpretive selectivity guiding what aspect of the modeled behavior will be learned through imitation is directed and constrained by the implicit assumptions of the infant's "pedagogical stance" that the other's ostensive cues activate. The main claim of natural pedagogy assumes that in order to interpret and learn from events, infants rely mostly on their own judgment when they comprehend the relevance of an action (with the help of their modes of construal or interpretive schemas), but they weigh the model's cues more heavily when they themselves are unable to figure out the reason for an action. In the former case, the relevance of an action is conveyed by its cognitive transparency, whereas in the latter case (i.e., cognitive opacity), relevance is inferred from the communication of a knowledgeable other. A challenging overall problem and a possible assumption addressed in this study is that the demonstration of a knowledgeable other not only conveys relevance through ostensive communication but also can enhance cognitive transparency and understanding at the same time.

The way imitation is influenced by social communication is still a compelling question. As we can see, ostensive communicative context and social cues facilitate imitation (see also Brugger et al., 2007; Nielsen, 2006). Furthermore, imitation's selective nature disappears in contexts lacking such communicative cues (see Király et al., 2004). In all of the above-mentioned studies, the original form of presentation was varied; during the modeling phase (when infants could code the information), the model either acted socially or was aloof, whereas the social context of the reenactment phase was not controlled. These studies, as a consequence, could not uncover entirely how social cues influenced imitation and whether imitation in these tasks functioned cognitively or communicatively.

Thus, the current study aims to address two important issues in relation to the factors guiding selective imitation in infants. First, it is still an open question whether young infants (12- and 14-month-olds) learn from failed attempts and are able to incorporate the information conveyed regarding the intended goal of the action. The second question relates to how social communicative contexts influence the retrieval phase of imitation, that is, how the presence of a communicative partner influence reenactment in young infants.

To be able to answer the first question, a new design was created where, with the help of communicative cues and through the presentation of negative evidence, the relevance of new information became cognitively transparent; thus, it could be shown that in this case (in contrast to a similar but cognitively opaque situation) a higher level of imitation occurs. In this new design, identical situational constraints were established for the model and for the infant in both conditions. A new variable was introduced to convey overt information on the efficiency (rationality) of the target act; the rationality of the novel action was indicated explicitly by the demonstration of the failure of the alternative prepotent "hand action." It is supposed that this prior failed attempt could help the comprehension of the efficiency of the means used in terms of physical constraints and, simultaneously, could emphasize the availability of the model's prior intention (see Carpenter et al., 2002). In consequence, a prior failed attempt could promote cognitive transparency.

With the objective of responding to the second question, in this study new testing conditions were introduced; whereas during the modeling phase the model always acted socially, during the test phase the model was either present or absent. With the help of this design, the primary function of imitation could be investigated; the model could be conceived either as the source of information or as a communicative partner. If the model is conceived "only" as the source of information, the amount of imitation cannot be expected to decrease in the model-absent testing condition, underlining that imitation serves primarily epistemic purposes.

Groups of 14-month-olds were tested while varying the situational constraints of the model with a new target object, "a ball as a lifter" for a hidden box, in two testing conditions.

Method

Design

The effects of two independent variables on infants' tendency to imitate a novel action were investigated. The first independent variable was whether the model was present or absent during the test phase (*model present* vs. *model absent*), and the second one represented the mode of presentation (*tool successful only, hands unsuccessful–tool successful,* and *hand manipulation control*). The above two factors were crossed in a factorial design, creating six different groups of participants. A *baseline* condition served to assess the spontaneous performance of infants on the new apparatus.

Participants

A total of 94 14-month-olds were recruited through advertisements in local newspapers. Of these, 9 were excluded from the final sample because of technical error ($n = 2$), parental interference ($n = 2$), or failure to come back for the test phase of the study ($n = 5$). Of the 85 participating infants, 62 were assigned to one of four conditions (*model present/tool successful only* [$n = 13$], *model present/hands unsuccessful–tool successful* [$n = 14$], *model absent/tool successful only* [$n = 18$], and *model absent/hands unsuccessful–tool successful* [$n = 17$]) and a further 23 were assigned to hand manipulation control conditions (*model present/hand manipulation control* [$n = 12$] and *model absent/hand manipulation control* [$n = 11$]). The mean age of the infants was 13 months 28 days (range = 13 months 17 days to 14 months 14 days), and their sex distribution was 41 girls and 44 boys.

An additional 14 infants were tested in the baseline condition (mean age = 14 months 5 days, range = 13 months 20 days to 14 months 15 days, 8 girls and 6 boys).

Apparatus

We introduced a small red box ($9 \times 9 \times 9$ cm) with a slightly smaller box hidden inside it ($9 \times 9 \times 5.5$ cm) and a new tool object—"a ball as a lifter." This new tool consisted of a magnet and a ball, with the magnet fixed to the ball (10 cm) on a short string. The hidden box was placed on a small table between the model and the infants during the modeling phase and was presented to the infants at the same location during the test phase. The sessions were monitored and videotaped from behind a one-way mirror.

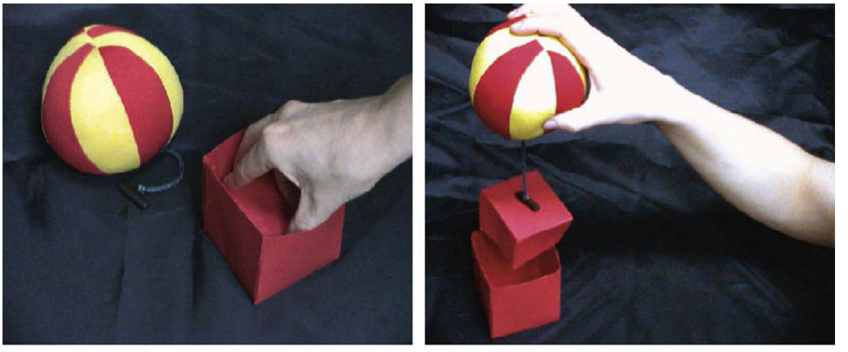


Fig. 1. A ball as a lifter.

Procedure

The infants were brought to the laboratory twice with a 1-week delay between visits. The first session consisted of the modeling phase, and the second session (1 week later) was the test phase of the study. This delay was introduced following the method of Meltzoff (1988), with the overall aim of testing the permanent effect of possible selective imitative behavior.

Modeling phase

The infant was seated on his or her parent's lap in front of the table with the apparatus covered with a cloth. The apparatus was placed approximately 1 m away from the participant to prevent him or her from touching it. The mother was instructed not to interact with her baby during the modeling phase. In each condition (and before each demonstration), the experimenter sat on the opposite side of the table, uncovered the hidden box, and then (as part of a communicative context) looked at the infant, called his or her name, and attracted his or her attention by saying "Look!", making sure that the infant and model were taking part in a joint attention situation.

In the tool successful only condition, the model grasped the ball and lifted the small hidden box out of the slightly larger one by contacting it with the magnet that was hanging from the ball. This event was the core target behavior in all of the experimental conditions. After taking out the smaller box with the ball and magnet, the model grasped the larger box as well and then hid the elements of the apparatus under the table, where she rearranged them and put them back on the table for further demonstrations.

In the hands unsuccessful–tool successful condition, before presenting the very same "lifting with the ball action," the model tried to take the box out with her hand but failed. (This was obviously difficult because the inner box was smaller in height and was inserted tightly into the outer box.) While presenting the failed attempt, the model performed a frown as a sign of her effort. After this failed attempt, the model performed the same successful action with the ball that was presented in the tool successful only condition (see Fig. 1).

In a hand manipulation control condition,[1] the demonstrator touched and pushed the box sideways with two fingers before the same "lifting with the ball action." While presenting this prior manipulation, the model performed the same frown as in the hand unsuccessful–tool successful condition. Thus, in this condition, the box was manipulated twice as in the hands unsuccessful–tool successful condition, accompanied by the same interactional frame, but at the same time this manipulation did not convey information on the efficacy of the hand action.

[1] I thank an anonymous reviewer for suggesting this control condition.

The demonstrator repeated the action sequence three times in each condition, making eye contact with the infant in between these actions and attracting the infant's attention if necessary (calling his or her name and saying "Look!" while alternating his or her gaze between the apparatus and the infant). The demonstrations of the target action sequence were accompanied by an approximately equal amount of verbal interaction in all conditions because attracting the attention of the infants just at the beginning of each demonstration was successful.

Test phase

The infant again was seated on his or her parent's lap, in front of the table with the uncovered hidden box and ball, but this time at a distance that allowed the child to reach it. In the model present condition, the model, who had demonstrated the target action 1 week earlier, sat on the other side and encouraged the infant to play with the apparatus without giving explicit instructions. The infant was told "It's your turn now, you can play!" In the model absent condition, the infant had the opportunity to play with the apparatus in the presence of his or her parent only. The mother was asked to refrain from giving any direct instruction with respect to the modeling phase and apparatus but was instructed to encourage the infant to play with the apparatus (using the very same instruction: "It's your turn now, you can play!").

Baseline condition

To assess the spontaneous production of the target action (lifting with ball or any other attempt to contact the ball and box) in the absence of adult demonstration, a baseline control group of 14 infants was exposed to the apparatus in the presence of an adult experimenter.

Data analysis and scoring

The video recordings of the test phase were scored by two independent observers who were uninformed as to which of the four conditions (tool successful only, hands unsuccessful–tool successful, hand manipulation control, or baseline) the participants belonged. The dependent measures were whether the infants attempted to get the hidden box with their *hand* or with the *ball* (or, if both mean behaviors appeared, their relative order) within a 20-s time window. (This time window was used following the method of Gergely et al., 2002, and Meltzoff, 1988.) The coders needed to decide according to the following criteria: A ball attempt was defined as a visible effort to contact the magnet or the ball with the hidden box, whereas a hand attempt was defined as an obvious trial to take out the little box with the hand. The precise exact execution of the modeled behavior was difficult for the infants; therefore, all of the cases where infants tried to use the ball as a lifter (e.g., contacting the hidden box directly with the ball, gripping the magnet and using it to take out the box), independent of whether their action was successful or not, were coded as imitation of the target behavior. Unsuccessful trials can represent understanding the goal of the modeled behavior, yet they fail to bring about the precise behavior and end state (Call & Carpenter, 2002). There was 98% agreement between the two independent coders, $\kappa = .939$, $p < .001$.

Results

The proportion of infants who performed the target action (the action with the ball) in each condition is presented in Table 1. When the model was present during testing, 93% of infants in the hands unsuccessful–tool successful presentation condition, 54% of infants in the tool successful only condition, and 50% of infants in the hand manipulation control condition reenacted the ball action. When the model was absent during testing, 65% of infants in the hands unsuccessful–tool successful presentation condition, 22% of infants in the tool successful only condition, and 35% of infants in the hand manipulation control condition tried to use the target ball action. In the baseline condition, in the presence of an experimenter, only 14% of infants used the ball as a lifter.

An overall analysis of the proportion of trials imitated was performed on 3 (Presentation Condition: hands unsuccessful–tool successful, tool successful only, or hand manipulation control) $\times$ 2 (Test Condition: model present or model absent) factors with the help of the generalized linear model (a general

Table 1
Percentage of infant imitators in each condition

	Hands unsuccessful–tool successful	Tool successful only	Hand manipulation control	Baseline	Overall
Model present	93% (13/14)	54% (7/13)	50% (6/12)	14% (2/14)	53% (28/53)
Model absent	65% (11/17)	22% (4/18)	35% (4/11)		41% (19/46)
Overall	78% (24/31)	36% (11/31)	43% (10/23)		

Note. The values in parentheses represent the numbers of infant imitators out of all participants.

logistic model for binary variables in SPSS 15). This analysis yielded a main effect of presentation condition, Wald $\chi^2 = 17.125$, $df = 2$, $p < .001$, and a main effect of test condition, Wald $\chi^2 = 6.072$, $df = 1$, $p = .014$.

Pairwise comparison of the presentation conditions revealed stronger imitation tendency in the hands unsuccessful–tool successful condition than in the other conditions: hands unsuccessful–tool successful/tool successful only condition, Bonferroni $p = .001$, hands unsuccessful–tool successful/hand manipulation control condition, Bonferroni $p = .011$. There were no significant differences between the other two presentation conditions.

Pairwise comparison of the test conditions showed a stronger imitation rate in the model present condition than in the model absent condition, Bonferroni $p = .014$. This analysis showed that the presence of the model during the test phase led to more frequent imitative behavior in the experimental conditions.

The impact of adult demonstration on the occurrence of target ball actions was assessed through comparing each experimental group with baseline. There was superior performance of a tool-use behavior in comparison with the baseline condition only in the hands unsuccessful–tool successful condition both when the model was present and when the model was absent during testing, model present/hands unsuccessful–tool successful versus baseline, Bonferroni $p < .001$, model absent/hands unsuccessful–tool successful versus baseline, Bonferroni $p = .015$.

The proportion of infants who performed the prepotent hand action (as a naturally available means behavior to attain the goal) in each condition, irrespective of whether the infants also used the ball as a tool, is presented in Table 2. When the model was present during testing, 36% of infants in the hands unsuccessful–tool successful presentation condition, 85% of infants in the tool successful only condition, and 91% of infants in the hand manipulation control condition used their hands to attain the goal of the action, whereas when the model was absent during testing, 47% of infants in the hands unsuccessful–tool successful presentation condition, 84% of infants in the tool successful only condition, and 73% of infants in the hand manipulation control condition tried to take out the little hidden box with their hands.

The overall analysis of the amount of hand actions (as prepotent actions with respect to goal attainment) was performed on 3 (Presentation Condition: hands unsuccessful–tool successful, tool successful only, or hand manipulation control) $\times$ 2 (Test Condition: model present or model absent) factors and revealed a main effect of presentation condition, Wald $\chi^2 = 19.653$, $df = 2$, $p < .001$. According to post hoc pairwise comparisons, there was a significantly lower level of hand actions in the hands

Table 2
Percentage of infants who performed goal-directed hand action in each condition

	Hands unsuccessful–tool successful	Tool successful only	Hand manipulation control	Overall
Model present	36% (5/14)	85% (11/13)	91% (11/12)	69% (27/39)
Model absent	47% (8/17)	84% (15/18)	73% (8/11)	67% (31/46)
Overall	42% (13/31)	84% (26/31)	82% (19/23)	

Note. The values in parentheses represent the numbers of infants who performed the goal-directed hand action out of all participants.

Table 3
Percentage of infant actors in each group

	Goal-directed hand action	Hand action and then ball action	Ball action and then hand action	Ball action
Model present/Hands unsuccessful–tool successful	7%	–	29% (0)	64% (28%)
Model absent/Hands unsuccessful–tool successful	18%	6% (0)	24% (12%)	35% (6%)
Model present/Tool successful only	46%	15% (8%)	24% (16%)	15% (0)
Model absent/Tool successful only	62%	5% (5%)	17% (5%)	–
Model present/Hand manipulation control	50%	16% (8%)	25% (16%)	9% (9%)
Model absent/Hand manipulation control	47%	–	26% (9%)	9% (9%)

Note. The values in parentheses represent the percentages of successful ball actions with respect to all participants in the condition.

unsuccessful–tool successful condition than in the tool successful only presentation condition, Bonferroni $p < .001$, and also compared with the hand manipulation control condition, Bonferroni $p = .001$, whereas there was no significant difference between the tool successful only and hand manipulation control conditions. Infants tried to use their hands to achieve the goal less frequently in the hands unsuccessful–tool successful condition than in the tool successful only and hand manipulation control conditions.

The proportions of cases where only hand or ball actions were used, or where both hand and ball actions were produced (as well as their relative order), are presented in Table 3. In the model present/hands unsuccessful–tool successful condition, 64% of infants tried to imitate the modeled ball action without trying to use their hands despite the fact that only 28% of the ball actions (with respect to all participants in this condition) were successful; an additional 29% of infants in this condition tried to use the ball and then used their hands. In the model absent/hands unsuccessful–tool successful condition, 35% of infants tried to attain the modeled goal only with the ball, and only 6% of them were successful. These proportions were much lower in the other presentation conditions, with 15% of infants in the model present/tool successful only condition, no one in the model absent/tool successful only condition, and 9% in the hand manipulation control condition (both when the model was present and when the model was absent) producing the ball action first.

Differences in the patterns of using just the ball, using the ball first and then the hand, or using the hand prior to using the ball were tested with respect to the experimental conditions. This analysis revealed significant differences, Kruskal–Wallis test $(\chi 2) = 24.102$, $df = 5$, $p < .001$. Pairwise comparison of test conditions revealed that in the hands unsuccessful–tool successful condition, a greater proportion of infants executed the target ball action first or alone compared with the other presentation conditions: hands unsuccessful–tool successful versus tool successful only, Mann–Whitney $U = 33.0$, $p = .004$; hands unsuccessful–tool successful versus hand manipulation control, Mann–Whitney $U = 23.5$, $p = .001$; tool successful only versus hand manipulation control, Mann–Whitney $U = 72.5$, *ns.* In general, the occurrence of the ball action first or alone in the model present testing condition was more frequent than in the model absent testing condition, Mann–Whitney $U = 611.0$, $p = .008$.

Importantly, in the hands unsuccessful–tool successful condition, there was only one case (6% in the model absent condition) when the order of the appearance of hand and ball actions followed the original order of presentation (the hand action preceding the action with the ball); in the hand manipulation control, two infants (16%) enacted a hand trial prior to a ball action.

In the case of the current apparatus, the new tool-use behavior turned out to be slightly difficult for the infants. The percentage of successful ball actions (regarding all of the participants in each condition) is also presented in Table 3. Overall, the amount of successful ball actions was low (5–28%). Most of the infants in the hands unsuccessful–tool successful condition kept on trying to use the ball as a lifter in alternative forms (e.g., contacting the hidden box directly with the ball, gripping the magnet and trying to use it) rather than using just their hands instead. This pattern of results does not appear in the other conditions, where infants tended to use their hands in addition to using the ball to attain the goal.

Discussion

This study has demonstrated selective "rational" imitation in infants in two testing conditions: the presence or absence of the model during the response phase. We could show different patterns of imitation, namely, selective imitation with respect to the inherent logic of the modes of presentation. The results suggest that in the hands unsuccessful–tool successful condition, infants inferred from the model's initial failed attempt that the subsequent ball-with-magnet action was the most effective means available in the situation to achieve the goal of lifting the box.

This design was different from, but analogous to, the original setup of Gergely and colleagues' (2002) study. In one of the presentation conditions (tool successful only), the rationality or effectiveness of the new means, in this case the unknown tool, was not explicit, and thus not easily interpretable in the situation, because for infants the availability of the model's hands could imply that the prepotent hand action might be the most efficient means to perform the task. Here too, as in the hands-free condition of the original head-on-box study (see Gergely et al., 2002; Király & Gergely, 2002), the model's unexpected choice (i.e., the use of a new tool) could mark the "action with ball" as new and relevant information that the ostensive communicative manifestation conveyed. The unexpected new means was cognitively opaque, but by being presented in a communicative context, it was highlighted as relevant information that is worthwhile to learn and follow.

The other mode of presentation (hands unsuccessful–tool successful) in the current design, on the contrary, emphasized the ineffectiveness of the prepotent hand action compared with the new means. Therefore, the main prediction on the basis of the model of selective imitation would be that the amount of imitation in this case will be even higher than in the tool successful only presentation condition. The reason for this claim is that, besides the manifested significance of the new means (which is conveyed by the ostensive communicative context), the presented negative evidence explains its use by highlighting the intended goal of the action and also making obvious the physical ineffectiveness of a hand action. This makes the new act cognitively transparent for infants. Infants performed the modeled new tool-use behavior more frequently in this presentation condition (both when the model was present and when the model was absent during testing) than in the baseline condition, strengthening the claim that the high level of imitation found in this presentation condition confirmed the above prediction.

In this design, the alternative explanation that imitation occurs only in the case of nonrational unusual actions would not predict any difference in imitation in the two presentation conditions given that the very same unusual target action appeared in them within the same situational constraints for the model and the infants.

The inherent logic of the contrast between the two modes of presentation introduced above might raise an alternative explanation. The main argument of this claim is that in the hands unsuccessful–tool successful condition, the attention of the infants was drawn to the box twice during demonstration. The additional prior hand act (even though it failed) could enhance the coding of the new tool through encountering more manipulation with the box[2]; thus, it is conceivable that infants imitated the ball action more frequently because both the box and the ball became more salient in that condition.

A further clue regarding the possible stimulus-enhancing effect of the demonstration in the hands unsuccessful–tool successful condition is given by recent research focusing on children with autism. Somogyi (2007) used the hidden box paradigm to capture how these children interpret others' actions. The results show that nonverbal low-functioning children with autism, unlike typicals, produce the hand action significantly more frequently in the hands unsuccessful–tool successful condition. This pattern of responses demonstrates that in these children the two-step manipulation of the objects did encourage imitation of both the hand and ball actions. However, these results also show that these children, unlike typicals, did not interpret the hand action as a failed action (indicating that it is not possible to take out the hidden box with the hand).

With the aim of ruling out the possible explanation of stimulus enhancement, a hand manipulation control condition, in which a prior hand manipulation appeared without a clear trial to take out the

[2] I thank an anonymous reviewer for proposing this alternative explanation and suggesting how to control for it.

box, was introduced within the very same communicative context. The proportion of imitators in this condition was significantly less than in the hands unsuccessful–tool successful condition (close to the amount of imitation in the tool successful only condition in both the model present and model absent testing conditions). In contrast, the amount of hand actions was significantly higher in this condition than in the hands unsuccessful–tool successful condition (as much as in the tool successful only condition in both the model present and model absent testing conditions). This pattern of results strengthens the claim that the prior hand attempts can serve as bases for inference for infants. Whereas in the hand manipulation control the prior manipulation only emphasized that the model was able to use her hand, in the hands unsuccessful–tool successful condition the prior failed attempt demonstrated that the hand was ineffective; hence, it was worth following the model's tool-use behavior as a more relevant means.

In addition, the occurrence of the target action was above baseline in the hands unsuccessful–tool successful mode of presentation conditions, indicating that there was imitative learning in these conditions. The above results underline that understanding the physical–causal efficiency of a new and relevant means results in an even higher amount of imitation. Thus, in this study, we could confirm the differential degree of imitation with respect to the effectiveness of a new tool-use behavior.

Interestingly, the presentation of the hands unsuccessful–tool successful and hand manipulation control conditions consisted of two steps (in contrast to the one-step tool successful only condition), so they were more challenging regarding mnemonic competence. Despite this fact, in the hands unsuccessful–tool successful presentation condition we found a higher level of imitation given that the extra information (the additional step) could open a door to better understanding the reason for the action, whereas even in the hand manipulation control the same amount of imitation appeared as in the tool successful only condition. The result that hand actions appeared significantly less often in the hands unsuccessful–tool successful condition, despite the fact that the hand was an "actor" in the presentation, reinforces the explanation that the interpretation of this prior step inhibited the prepotent but ineffective hand action. It also rules out the possibility of blind imitation and underlines the understanding and cognitive transparency of the presentation.

A central result of this study is that a main effect of model presence was found; there was selective imitation in both testing conditions (model present and model absent), whereas there was a decrease in the amount of imitation in the absence of the model during the testing phase. A plausible interpretation of this result is that 14-month-olds do not "blindly" follow the model who acts socially; nevertheless, they profit from her presence during reenactment. First, for young infants, the model as a knowledgeable other functions as the source of information that mediates action learning. Infants at this age drew inferences from the demonstration regarding the efficacy of the available and possible means with respect to goal attainment, and they acted on the basis of this coding of the action. This does not seem to be the case for older infants given that they act more socially, copying the model's behavior with high fidelity irrespective of the relevance of the demonstrated means (Brugger et al., 2007; Horner & Whiten, 2005; Nielsen, 2006). Second, young infants tend to benefit from the presence of the model as a knowledgeable other *and* as a communicative partner during the test phase given that in the course of retrieval the model can modulate the practice of a new tool use by giving feedback and correction on action learning. Thus, although young infants seem to imitate selectively, they still rely on social cues, taking the chance to learn from the model about the world. Consistent with the model of natural pedagogy (Gergely & Csibra, 2005, 2006), a communicative context facilitates the acquisition, and thus the transmission, of new and relevant means.

The presented results add to a growing body of literature demonstrating the importance of selective interpretation behind imitative learning. For example, Nielsen (2006) documented selective imitational performance in 12-month-olds using failed attempts as a source of information for the logical reason of a tool-use behavior. His results emphasized that imitation occurs after a prior failed attempt; otherwise, infants emulatively use their hands to solve the problem. In our case, regarding all of the presentation conditions, there was imitative learning when the model was present during the response phase. However, it was performed at different levels depending on alterations in the cognitive transparency of the actions modeled. Brugger and colleagues (2007) demonstrated that a physically necessary prior step resulted in a more elaborate organization of a two-step event sequence. Similarly, in our study, a prior unsuccessful act (which was not, however, part of the modeled one-step target action) revealed a deeper understanding of the new means behavior. Furthermore, this study confirmed that the selective encoding of the different contextual information has long-term effects in that infants imitated selectively after a longer 1-week delay.

In addition, the results of this study expand our knowledge of the way social cues influence imitative behavior. Previous studies reported that 14- and 16-month-olds not only copied actions that were sufficient to arrive at the end state of the demonstrated action sequence but also copied inefficient actions if they were modeled in a communicative context (Brugger et al., 2007). Furthermore, 18-month-olds were more inclined to copy the specific actions of a model when she was engaged than when she acted aloof during presentation (Nielsen, 2006), and 18-month-olds copied the novel actions of a model regardless of the apparent logic of the demonstration in an ostensive communicative context (Király et al., 2004). In another study, 24-month-olds' tendency to copy the model's specific actions was not influenced by her social disposition (disinterest) during modeling, although their actions were more successful when the model acted socially (Nielsen, 2006). What is common in these studies is that they all investigated the role of a social communicative context during the modeling phase, and thus the period of coding, in imitative learning. Our study revealed that social communicative cues, such as the presence of the model, can play a crucial role in learning through imitation during the retrieval phase as well.

Our results, however, are challenged by the age-related changes that the above-mentioned studies uncovered. Whereas younger infants imitate selectively, copying primarily to satisfy cognitive motivations and learn about events in the world (as in our study), older infants seem to imitate to fulfill social motivations and upgrade shared experience.

As a further interesting piece of evidence, Horner and Whiten (2005) found that 3- and 4-year-olds reproduced both irrelevant and relevant actions with high fidelity, indicating that the availability of causal information did not influence the social learning strategy they employed. Moreover, their tendency to imitate was not influenced by the fact that the model left the room at the time of testing. Thus, it seems plausible that the logic of a model's demonstration and the communicative cues she provides influence differentially how children engage in social learning at different ages.

The difficulty of successfully reproducing the modeled new target behavior is another possible limitation of the current results. Thus, it is still an open experimental question as to how infants would engage in imitative learning in the case of an easy-to- perform new tool-use behavior.

In sum, differential imitation in the three presentation conditions in this study suggests that imitative learning is a selective interpretive process involving the evaluation of the rationality of the means in relation to the situational constraints of the actor. These results also serve as a proof of sensitivity to communicative relevance in imitation. In addition to pointing to what is relevant and new information in the situation, the presentation of a failed attempt in an ostensive communicative context even makes transparent why it is relevant. In our case, the prior failed attempt with the hand explained the model's contrastive choice; it helped to find out the overall goal of the situation (for a similar argument, see Carpenter et al., 2002) and, at the same time, informed the observer about the physical ineffectiveness of the available prepotent means.

Acknowledgments

The research outlined in this article was supported by an FP6 grant ("EDICI") to the research team of the Institute for Psychology, Hungarian Academy of Sciences, led by György Gergely, and a grant from OTKA to Ildikó Király (T 047071). I thank György Gergely for his comments on the manuscript and thank Kata Egyed, Kata Krekó, Krisztina Kupán, Eszter Somogyi, and Rita Kéri for their contributions. Some of the data were presented at the 13th biennial meeting of the International Conference on Infant Studies (ICIS), Toronto, Canada, in 2002.

References

Barr, R., Dowden, A., & Hayne, H. (1996). Developmental changes in deferred imitation by 6- to 24-month-old infants. *Infant Behavior and Development, 19*, 159–171.

Bekkering, H., Wohlschlager, A., & Gattis, M. (2000). Imitation of gestures in children is goal-directed. *Quarterly Journal of Experimental Psychology A, 53*, 153–164.

Bellagamba, F., & Tomasello, M. (1999). Re-enacting intented acts: Comparing 12- and 18-month-olds. *Infant Behaviour and Development, 22*, 277–282.

Brugger, A., Lariviere, L. A., Mumme, D. L., & Bushnell, E. W. (2007). Doing the right thing: Infants' selection of actions to imitate from observed event sequences. *Child Development, 78*, 806–824.

Call, J., & Carpenter, M. (2002). Three sources of information in social learning. In K. Nehaniv & C. L. Nehaniv (Eds.), *Imitation in animals and artefacts* (pp. 211–228). Cambridge, MA: MIT Press.

Call, J., Carpenter, M., & Tomasello, M. (2005). Copying results and copying actions in the process of social learning: Chimpanzees (Pantroglodytes) and human children (Homo sapiens). *Animal Cognition, 8*, 151–163.

Carpenter, M., Call, J., & Tomasello, M. (2002). Understanding "prior intentions" enables 2-year-olds to imitatively learn a complex task. *Child Development, 73*, 1431–1441.

Carpenter, M., Call, J., & Tomasello, M. (2005). Twelve- and 18-month-olds copy actions in terms of goals. *Developmental Science, 8*, F13–F20.

Gergely, G., Bekkering, H., & Király, I. (2002). Rational imitation in preverbal infants. *Nature, 415*, 755.

Gergely, G., & Csibra, G. (2005). The social construction of the cultural mind: Imitative learning as a mechanism of human pedagogy. *Interaction Studies, 6*, 463–481.

Gergely, G., & Csibra, G. (2006). Sylvia's recipe: The role of imitation and pedagogy in the transmission of human culture. In N. J. Enfield & S. C. Levinson (Eds.), *Roots of human sociality: Culture, cognition, and human interaction* (pp. 229–255). Oxford, UK: Berg.

Horner, V., & Whiten, A. (2005). Causal knowledge and imitation/emulation switching in chimpanzees (*Pan troglodytes*) and children (*Homo sapiens*). *Animal Cognition, 8*, 164–181.

Huang, C., Heyes, C., & Charman, T. (2002). Infants' behavioral re-enactment of "failed attempts": Exploring the role of emulation learning, stimulus enhancement, and understanding of intentions. *Developmental Psychology, 38*, 840–855.

Killen, M., & Uzgiris, I. (1981). Imitation of actions with objects: The role of social meaning. *Journal of Genetic Psychology, 138*, 219–229.

Király, I., Csibra, G., & Gergely, G. (2004). The role of communicative-referential cues in observational learning during the second year. Poster presented at the 14th biennial International Conference on Infant Studies, Chicago.

Király, I., & Gergely, G. (2002). Imitation of goal-directed acts in infants is a selective interpretative process. Poster presented at the 13th biennial International Conference on Infant Studies, Toronto, Canada.

Meltzoff, A. N. (1988). Infant imitation after a one week delay: Long term memory for novel acts and multiple stimuli. *Developmental Psychology, 24*, 470–476.

Meltzoff, A. N. (1995). Understanding the intentions of others: Re-enactment of intended acts by 18-month-old children. *Developmental Psychology, 31*, 1–16.

Nagell, K., Olguin, R., & Tomasello, M. (1993). Processes of social learning in the imitative learning of chimpanzees and human children. *Journal of Comparative Psychology, 107*, 174–186.

Nielsen, M. (2006). Copying actions and copying outcomes: Social learning through the second year. *Developmental Psychology, 42*, 555–565.

Schwier, C., van Maanen, C., Carpenter, M., & Tomasello, M. (2006). Rational imitation in 12-month-old infants. *Infancy, 10*, 303–311.

Somogyi, E. (2007). *The development of intentionality in non-verbal children with autism: A comparative study*. Paris: Université Paris V–René Descartes.

Tomasello, M. (1999). *The cultural origins of human cognition*. Cambridge, MA: Harvard University Press.

Uzgiris, I. (1981). Two functions of imitation during infancy. *International Journal of Behavioral Development, 4*, 1–12.

Want, S. C., & Harris, P. L. (2001). Learning from other people's mistakes: Causal understanding in learning to use a tool. *Child Development, 72*, 431–443.

Want, S. C., & Harris, P. L. (2002). How do children ape? Applying concepts from the study of non-human primates to the developmental study of "imitation" in children. *Developmental Science, 5*, 1–13.

dc_1353_16


Volume 34, issue 11, November 2013
ISSN 0891-4222
ELSEVIER
Research in Developmental Disabilities

Research in Developmental Disabilities 34 (2013) 3822–3832

Contents lists available at ScienceDirect

Research in Developmental Disabilities

ELSEVIER

Research in Developmental Disabilities

Understanding goals and intentions in low-functioning autism☆

Eszter Somogyi [a,*], Ildikó Király [b], György Gergely [b], Jacqueline Nadel [c]

[a] Department of Cognitive Psychology, Eötvös Loránd University, Izabella u. 46, H-1064 Budapest, Hungary
[b] Cognitive Development Center, Department of Cognitive Science, Central European University, Nádor u. 9, H-1051 Budapest, Hungary
[c] Centre Émotion, UMR 7593, Pavillon Clérambault, Hôpital de la Salpêtrière, 47, Boulevard de l'Hôpital, 75013 Paris, France

ARTICLE INFO

Article history:
Received 22 March 2013
Received in revised form 23 July 2013
Accepted 30 July 2013
Available online 7 September 2013

Keywords:
Autism
Understanding goals
Intentionality
Deferred imitation

ABSTRACT

We investigated ability to understand goals and attribute intentions in the context of two imitation studies in low-functioning, nonverbal children with autism (L-F CWA), a population that is rarely targeted by research in the domain. Down syndrome children (DSC) and typically developing children (TDC) were recruited to form matched comparison groups. In the two sets of simple action demonstrations only contextual indicators of the model's intentions were manipulated. In the Head touch experiment the model activated a button on a toy by pushing it with the forehead, whereas in the Hidden box experiment the model used a ball with a magnet to lift a box out of its container. Both actions were unusual and non-affordant with regards to the objects involved, none of the children in the baseline condition produced them. L-F CWA imitated the experimenter exactly, regardless of the model's intention. TDC showed appreciation of the model's intention by imitating her actions selectively. DSC reproduced only the intentional action as often as they imitated the experimenter exactly. It is concluded that L-F CWA attributed goals to the observed model, but did not show an appreciation of the model's intentions even in these simplified, nonverbal contexts.

Children interpret and predict others' actions on the basis of the mental states they attribute to the actor. Autism is associated with a specific cognitive deficit in inferring and representing mental states, as documented by seminal studies showing difficulties with false belief tasks (Baron-Cohen, 1995; Baron-Cohen, Leslie, & Frith, 1985; Leslie & Thaiss, 1992; Perner, Frith, Leslie, & Leekam, 1989; Sodian & Frith, 1994) and with pretend play (Wing, Gould, Yeates, & Brierley, 1977). Mental states, however, vary in nature; beliefs, desires, goals, intentions, emotions as well as perceptions have been proposed in literature (Frith, Morton, & Leslie, 1991; Luo, 2011; Premack & Woodruff, 1978; Saxe, Carey, & Kanwisher, 2004; Vivanti et al., 2011). Goals, for instance may be considered as mental states that are more 'transparent' or observable in behaviour than beliefs and desires, at least under a broad mentalising theory (Hamilton, 2009). Still, relatively few studies have investigated goal understanding in autism and most of these studies have involved high-functioning children with autism (H-F CWA). One of the aims of this study therefore was to assess goal understanding on the other end of the autistic spectrum, in low-functioning children with autism (L-F CWA). A further issue with existing reports is that they do not distinguish in all cases between an understanding of a goal as an internal state and the understanding of the visible outcome of a goal directed action, without inferring an intentional mental state. Therefore the other aim of our study was to better

☆ This research was funded by the Co-tutelle Scientific Exchange Programme of the French Government and was also supported by a grant from the Hungarian Science Foundation (OTKA) to the 3rd author (OTKA Grant #: NK 83997).
* Corresponding author. Tel.: +36 205856003.
E-mail addresses: somogyi.eszter@ppk.elte.hu, esomogyi@gmail.com (E. Somogyi).

0891-4222/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.ridd.2013.07.039

control this possibility by using imitation paradigms where the children's responses may indicate for us the different intentional mental states they attribute to the model across the experimental conditions.

1. Goal understanding in high-functioning autism

Studies that explored goal understanding in H-F CWA show that although understanding beliefs and desires is clearly impaired, the ability to read another person's intended goals is preserved. In fact, there are not many direct tests of whether individuals with autism understand action goals, but much information can be gathered from studies of imitation and experiments that explored how they interpret the actions of animated figures. We shall first review these two groups of studies.

Imitation research shows that CWA, high-functioning or with mild to moderate mental handicap for a few studies, can successfully imitate actions with clear goals, thus showing some understanding of goals. For instance, they reproduce object-use actions (Beadle-Brown, 2004; Charman & Baron-Cohen, 1994; Stone, Ousley, & Littleford, 1997), recognise and reproduce the goal of others' hand actions (Avikainen, Wohlschläger, Liuhanen, Hänninen, & Hari, 2003; Hamilton, Brindley, & Frith, 2007), and also perform well on nonverbal gesture recognition (Hamilton et al., 2007; Smith & Bryson, 1994) or gesture memory tasks (Rogers, Bennetto, McEvoy, & Pennington, 1996). Autistic children's imitation behaviour seems to be driven by goals' saliency, with more imitation in cases where the action has a clear and interesting outcome (a light or sound) compared to cases without an outcome (Ingersoll, Schreibman, & Tran, 2003). Hamilton et al. (2007) add to these results the finding that goal understanding is in fact an island of intact functioning in autism, in contrast to these children's poor performance on theory of mind tasks. In this study, three action tasks were proposed to CWA with moderate mental retardation, assessing goal-directed imitation, mirror imitation and grasp planning. Although CWA did not succeed as controls in the theory of mind tasks, there were no differences between the two groups in goal imitation.

Two studies, also using imitation as a measure, have asked whether H-F CWA expect agents to use the most efficient or rational means possible to reach their goals. Evaluating the effectiveness of an observed means is an important step towards reading others' intentions. Although these investigations are not conclusive regarding the rationality principle (Gergely, Bekkering, & Király, 2002; Gergely & Csibra, 2003), they provide further evidence showing that H-F CWA understand goals. For instance, Rogers, Young, Cook, Giolzetti, and Ozonoff (2010) compared the imitation of functional and non-functional acts on objects (to shake a maraca or to shake a potato masher) and found that CWA imitated the functional acts. They imitated less in the non-functional conditions; however, this does not necessarily mean that they did not understand these goals. Perhaps the children were simply less inclined to imitate in these conditions for two reasons, which we explain here in more detail because they are both relevant regarding the choice of tasks for our present study. Firstly, non-functional acts are less rational in the sense that they do not correspond to the common use of objects (and are therefore more social in nature – as shaking a potato masher may be an invitation to play or to share a pleasant moment). The absence of imitation in these tasks may simply reflect a preference for the imitation of functional actions with objects that highlight means–end relations (shaking a maraca corresponds to the object's function and brings about a salient effect). Secondly, in this particular case, in order to imitate a non-functional action, children need to suppress the action scheme that is activated by the object's affordances and common usage (which, in the case of the potato masher would be to make mashing movements). Perhaps a failure to imitate in these situations also reflects a failure to inhibit the usual action scheme. Studies that similarly report enhanced performance on meaningful imitation compared to meaningless imitation or the imitation of an action's style used pantomime or gestural imitation as measures where again the goal of the action was less manifest in the sense that it did not bring about a salient change in the environment (Hobson & Hobson, 2008; Rogers et al., 1996; Stone et al., 1997). The lack of imitation in the non-functional or meaningless conditions in these studies therefore does not necessarily reflect a failure to interpret actions being that are non-rational with reference to their goals. However, they provide clues regarding autistic children's imitative behaviour and show their preference for imitating goal-directed, functional actions with objects that stress means–end relations.

The second group of studies on goal understanding in H-F CWA examined how they interpret the actions of animated geometric shapes. The results show that they derive the goals of an action from the situation's physical parameters or the agent's kinematic properties just like controls (Abell, Happe, & Frith, 2000; Castelli, 2006; see also Castelli, Frith, Happe, & Frith, 2001 for a neuroimaging study with adults). In Abell et al.'s (2000) study the animations showed two triangles moving around the screen according to one of three conditions, where the different types of motions could be described either in terms of mental states, goal-directed actions or non-deliberate actions (such as random movement). Before each presentation, subjects were cued with character roles; for example, the two triangles were a mother and a child. In this study, H-F CWA gave descriptions of the goal-directed animations (e.g. fighting, chasing) and the random animations (e.g. floating, drifting) that were as accurate as the controls. They used mentalistic descriptions (e.g. tricking, being jealous) less than typical controls, frequently referring to mental states that were inappropriate to the animation. In a similar study by Castelli (2006), children saw a circle at the bottom of a U-shaped valley rolling up and down the slopes and getting closer to (but failing to actually reach) a target (another circle resting at the top of either side of the valley). The task was to decide about the final goal of the moving circle by clicking (with the computer mouse) on one of the five marked locations along the slopes. H-F CWA were as able as controls to infer the agent's intended goal, even though the target was never reached.

All the above studies indicate that goal understanding is indeed an island of intact functioning in H-F CWA. We can see, however, that further investigations involving L-F CWA are needed in order to confirm that ability is generally preserved in ASD. Only one study, by Nadel et al. (2011), explored observational learning in L-F CWA. Results show that they are able form long-term representations of actions involving several sub-goals and can also reproduce these actions if they have an opportunity to practice. Another issue with these studies is that they mostly show that autistic children understand and infer the visible outcome of a goal-directed action, but do not indicate whether they also infer an intentional mental state underlying the goal. In everyday life, the observable outcome of a motor act is most often not sufficient to understand the action; the actor's intention also needs to be inferred, based on the referential and communicative cues conveyed by the agent's behaviour and the broader context in which the action occurs (Gergely & Csibra, 2003; Gómez, 2009).

2. Intention understanding in high-functioning autism

Only few studies have explored specifically intention understanding in autism and again, these have mostly been done with children on the higher end of the autistic spectrum. In fact, only one study examined whether L-F CWA recognize the imitative intention of another person (Nadel et al., 2000). In this experiment, using a revisited version of the still face paradigm, it was found that, although children do not react to a first still face episode, they react to a second one as a violation of intentional exchanges. Studies involving H-F CWA tested intention inference in situations where the observable outcome did not match the original intention (own intention or another person's intention), as in the case of accidents or experimental situations where the outcome of an action is manipulated.

Two such studies investigated whether individuals with autism have difficulty recalling and processing their own intentions, but results are not consistent. A study by Phillips, Baron-Cohen, and Rutter (1998) examined whether H-F CWA children could recognise when the outcome of their own action was caused intentionally or not in a series of tasks where the outcome was controlled. They found that CWA were more likely than typical children to say they did mean to hit an unintended target (reporting intention to hit a target when the outcome was successful, but not when it was unsuccessful), suggesting a difficulty in monitoring and remembering their own intentions. Russell and Hill (2001) reported no difference between CWA who had mild to moderate intellectual disability and controls in both the shooting game task and a novel 'drawing' task where children were asked to report their own intended action when the final outcome was unexpected. However, it is possible that subjects' response in the drawing task was prompted by the direct question of whether they thought or meant to draw what turned out to be unexpected ('Did you think you were drawing an X or a Y?'/'Did you mean to draw an X or a Y?').

Two other studies used the unfulfilled intentions paradigm (Meltzoff, 1985), where children witness an incomplete or failed act involving an object and are later asked to act on the same objects. For example, in one of these tasks, the model holds a dumbbell shaped pull-apart toy and attempts to pull it apart, but does not succeed because her hand slips off the object. In the first study, reported by Aldridge, Stone, Sweeney, and Bower (2000), 2–4 year-old CWA were found to imitate intended actions on the objects, even though they never saw the completed the intended act. Similar results were found in a better controlled study by Carpenter, Pennington, and Rogers (2001) who showed that 2.5–5 year-old CWA produced as many target acts as controls after observing the model demonstrate the failed action. Both studies concluded that CWA were able to guess the model's intentions by observing failed attempts. The difficulty with the unfulfilled actions paradigm, however, is that the reproduction of a model's intended behaviour after observing the model's failed attempts may reflect affordance learning alone, without any actual understanding of the adult's intentions (Huang, Heyes, & Charman, 2002; Huang, Heyes, & Charman, 2006). In fact, one of the control conditions Carpenter et al. (2001) included in their study addresses this issue. In a manipulation condition, the experimenter modelled a set of similar actions that were different from the target action but were performed in the same general area of the object. In this condition CWA produced as many target acts as in the unfulfilled condition, an effect that was not found for children with developmental delays. Thus, it is possible that any random manipulation of the object was enough to make CWA produce the target action through stimulus enhancement.

D'Entremont and Yazbek (2007) avoided this problem with an experimental setting where the manipulated objects' constraints did not provide cues on the target action since the same action was presented as intentional (therefore target action) in some conditions and accidental in others. Only verbal cues marked the experimenter's intention: 'There' for intentional actions and 'Whoops' for accidental ones. The authors found that, unlike typically developing controls, H-F CWA did not respond on the basis of the experimenter's intentions. They tended to imitate the intentional acts as well as the irrelevant 'accidental' acts performed by the demonstrator, thus failing to select the intended or effective, rational acts in the stream of the demonstrator's behaviour. The authors conclude that, consistently with the findings in Carpenter et al.'s (2001) study, CWA were guided by stimulus enhancement, that is, the model's actions highlighted affordances of the objects and the children subsequently performed the actions that the objects afforded. It should be noted that the mean CARS scores of the participants (29 points) in this study did not reach clinical cut-off for autism; therefore a replication of this pattern is needed with children who clearly have autism.

We can observe that the last three studies (Aldridge et al., 2000; Carpenter et al., 2001; D'Entremont & Yazbek, 2007) all proposed actions that could be derived from the affordances of the objects involved, such as: grasping knobs to pull a toy apart (Meltzoff task) or using accessories like switches, buttons, handles and loops (D'Entremont & Yazbek, 2007). The issue with these objects and parts is that they are all used in everyday life and are therefore loaded with associated action schemes that, given their preference for functional actions, may be difficult to inhibit for CWA. For instance, having seen a dial turned, CWA may have more difficulties inhibiting this highly functional action, independently of whether it was presented intentionally or by accident. All three studies (Aldridge et al., 2000; Carpenter et al., 2001; D'Entremont & Yazbek, 2007) report that CWA, unlike controls, did not take into account the model's intention when imitating her actions, instead, they often showed exact copying behaviour. If this is a learning strategy specific to CWA, then exact copying should appear in tasks where the presented object-directed actions are not common everyday actions and could only have been acquired from the demonstration in the laboratory. Also, if exact copying is specific to ASD, then this strategy should appear on the lower end of the spectrum.

A recent study by Vivanti et al. (2011) added the use of eye tracking to behavioural measures in order to investigate the very mechanisms that scaffold action understanding in H-F CWA. Consistently with above, their data show that H-F CWA, like typically developing controls, take into account the environment in which the action occurs and consider information about both the agent's behaviour and the constraints of the situation (see also Bedford et al., 2012, for intact early gaze following in infants later diagnosed with ASD). They also showed typical usage of the agent's emotional expressions to infer his or her intentions. However, they presented subtle atypicalities in the way they responded to an agent's direct gaze and showed impairments in their ability to attend to and interpret referential cues such as a head turn for understanding an agent's intentions. Even though this study again involved only a select subgroup of older, H-F CWA, disentangling these contributors to action understanding is relevant to interpreting the results of our present investigation.

3. Aims of the present study

From this review of literature, two main issues emerge that we wished to address:

(1) Studies show that goal attribution is preserved in H-F CWA. If this ability is generally preserved in ASD, then L-F CWA should also be able to attribute goals (either by inferring the visible outcome of a goal-directed action or by inferring an intentional mental state underlying the goal).
(2) Only few studies (Aldridge et al., 2000; Carpenter et al., 2001; D'Entremont & Yazbek, 2007) tackled the ability to understand a goal as an internal, intentional state and not simply as the outcome of an action. These report that H-F CWA do not take into account the experimenter's intentions when imitating her actions, instead they often showed exact copying as a learning strategy. If this pattern is specific to ASD, (a) then it should be confirmed by responses in tasks where functionality or usual affordances cannot bias the children's responses and (b) L-F CWA should also respond with this strategy.

To answer these issues, we used two imitation tasks, where the model's intent was conveyed by easily observable elements (e.g., situational features or behaviour). The presented object-directed actions were not common functional actions and could only have been acquired from the demonstration in the laboratory. The actions we presented to the children (using the forehead to activate a button or using a ball as a lifter) were not only unusual, but also required the inhibition of existing motor schemas associated with the objects. Both paradigms were adopted from research on the typical development of intentionality (Gergely et al., 2002; Király, 2009); they were nonverbal, involved interesting novel objects, and required only a short attention span. The first task was the 'Head touch' paradigm devised by Gergely et al. (2002) to test goal inference in typically developing infants. The second task, which we refer to as the 'Hidden box' paradigm, was devised by Király (2009) to explore goal inference in situations in which the model's intentions are conveyed by negative behavioural evidence. Participants' deferred imitative responses were the outcome measure in both paradigms. This choice was based on the assumption that deferred imitation requires individuals to store demonstrated actions in memory and intentionally select those actions to reproduce from among several alternatives.

For our study, we recruited L-F CWA, who are known to have severe communication problems (Noens & van Berckelaer-Onnes, 2004). These children's language profile develops unevenly as compared with H-F CWA or children who have an intellectual deficit without autism. For instance, L-F CWA often achieve an expressive language level that is above their receptive language level, whereas comparison groups show the opposite profile (Hudry et al., 2010; Kjelgaard & Tager-Flusberg, 2001; Maljaars, Noens, Scholte, & van Berckelaer-Onnes, 2012). In L-F CWA specifically, language abilities are correlated with joint attention and symbol understanding (Charman et al., 2003; McDuffie, Yoder, & Stone, 2005), which are both considered as precursors to goal understanding and intentionality. In order to avoid the confounding effect of language abilities, we decided to recruit nonverbal children for our study.

4. Method

4.1. Participants

4.1.1. Experiment 1: Head touch

For the Head touch experiment, we recruited 28 CWA (25 boys and 3 girls) from two schools for CWA in Paris. All participants met full DSM-IV (Diagnostic and Statistical Manual of Mental Disorders, 4th edition; American Psychiatric Association, 1994) criteria for autistic disorder. The children's diagnosis was previously established by various health care professionals before the children entered the schools. The schools' professionals assessed the severity of the children's autism with the Childhood Autism Rating Scale (CARS) (Schopler, Reichler, & Renner, 1988) upon admission and each year. We selected children with severe autism for this study, all the participants had CARS scores between 38 and 60 points (mean 45 points), which is associated with severe and low-functioning autism (Mayes et al., 2012). The children were also selected for their nonverbal behaviour, as reported in their diagnosis and confirmed by the school's psychologist and teachers. Their

Table 1
Descriptive data of the clinical samples taking part in the two conditions of the Head touch experiment.

Experimental condition	Diagnosis	Mean mental age (months)	Std. deviation	*N*
Hands free	Autism	30.47	8.839	15
	Down syndrome	22.44	9.761	9
	Total	27.46	9.820	24
Hands occupied	Autism	30.00	9.204	15
	Down syndrome	32.71	11.101	7
	Total	30.86	9.662	22
Total	Autism	30.23	8.869	30
	Down syndrome	26.94	11.305	16
	Total	29.09	9.788	46

ages ranged from 6.7 to 16.2 years, with a mean age of 11.3 years. Their mental ages, as defined with the Psychoeducational Profile Revised or PEP-R (Schopler, Reichler, Bashford, Lansing, & Marcus, 1990), ranged from 17 to 46 months, with a mean age of 26 months. The PEP-R (Schopler, Lansing, Reichler, & Marcus, 2005) is designed to evaluate the development of communication and motor skills, and the presence of maladaptive behaviours in children with ASD aged 2 years to 7.5 years (developmental age). We also recruited 16 DSC (7 girls and 9 boys) from a medical and research institute for children with genetic illnesses in Paris. All of the children in this group were diagnosed with Down syndrome via chromosome study. Their chronological age ranged from 1.4 to 5.5 years, with a mean age of 3.3 years. Developmental age was defined by the Brunet-Lézine test (Brunet-Lézine, 2001) and ranged from 14 to 46 months, with a mean developmental age of 27 months. The group of TDC was composed of 18 14-month-old infants (8 girls and 10 boys) recruited from the research participant pool at the Institute of Psychology of the Hungarian Academy of Sciences. All children's parents received an information sheet about the experiments and provided signed consent. Children were excluded from participating in the study if they suffered from epilepsy, serious visual or motor problems, West syndrome, or any other serious medical condition associated with autism or Down syndrome. Children with comorbid Down syndrome and autism were also excluded. We recruited a Down syndrome comparison group to control for the effect of mental retardation; DSC have been demonstrated to have an unimpaired ability to attribute intentions and mental states (Baron-Cohen et al., 1985). We included TDC in this experiment in order to validate an adapted version of Gergely et al.'s (2002) original paradigm; the lamp box was replaced with a soft toy to make it safer for use with CWA.

No significant difference in mean mental age between the two clinical groups was observed ($F = 1.96$; $df = 3$; $p = 0.13$). Please see Table 1 for the descriptive data of the two clinical samples taking part in the Head touch experiment. We did not compare chronological age because DSC were matched with CWA by mental rather than chronological age, and were consistently younger; in DSC, mental age corresponds reliably with chronological age (Anderson et al., 2007; Lord & Luyster, 2006).

4.1.2. Experiment 2: Hidden box

For the Hidden box experiment, we recruited 14 children with autistic disorder (2 girls and 12 boys) from the same schools as for Experiment 1. The children's diagnosis and the assessment of the severity of their symptoms were made in the same way as for Experiment 1. All participants met full DSM-IV (Diagnostic and Statistical Manual of Mental Disorders, 4th edition; American Psychiatric Association, 1994) criteria for autistic disorder and had severe, low-functioning autism, with CARS scores between 38 and 60 points (mean 47 points). The children were again nonverbal, as reported in their diagnosis and confirmed by the school's psychologist and teachers. Exclusion criteria were also the same as for Experiment 1. Chronological age ranged from 3.5 to 17.75 years, with a mean age of 8.3 years. Mental age, as defined with the Psychoeducational Profile Revised or PEP-R (Schopler et al., 1990), ranged from 27 to 46 months, with a mean of 40 months. Twenty DSC (11 girls and 9 boys) were recruited to form the clinical control group. Of the 20 children, 9 were recruited from the same institute as the DSC in the Head touch experiment, 11 were from a school for mentally disabled children. Chronological age in DS group ranged from 2.2 to 6.8 years, with a mean of 4.2 years. Developmental age, as defined with the Brunet-Lézine test, ranged from 17 to 46 months, with a mean of 31 months.

Analysis revealed significant differences in mental age between groups ($F = 3.28$; $df = 3$; $p = 0.04$). Tukey's post hoc test revealed a significant difference in mean mental age between CWA in the Hand Action Possible condition and the DSC in the Hand Action Unsuccessful condition ($p = 0.05$, CWA had a higher mean). Since the objective of the study was to compare children's performances within a clinical diagnosis or within an experimental condition (i.e., not to compare the performances of CWA and DSC across different experimental groups), we retained the original group assignments. Please see Table 2 for the descriptive data of the two clinical samples taking part in the Hidden box experiment. Again, for the reasons explained above, we did not compare the chronological ages.

4.2. Baseline group

Prior to the experiments, to assess spontaneous production of the target actions (touching with head or lifting with ball) in the absence of adult demonstration, we exposed a baseline control group of 10 children with severe autism to the

Table 2
Descriptive data of the clinical samples taking part in the two conditions of the Hidden box experiment.

Experimental condition	Diagnosis	Mean mental age (months)	Std. deviation	*N*
Hand action unsuccessful	Autism	36.25	6.319	8
	Down syndrome	30.27	9.951	11
	Total	32.79	8.929	19
Hand action possible	Autism	39.78	5.848	9
	Down syndrome	31.44	7.552	9
	Total	35.61	7.830	18
Total	Autism	38.12	6.153	17
	Down syndrome	30.80	8.746	20
	Total	34.16	8.418	37

experimental toys (5 children for each apparatus). These children were recruited from the same schools for CWA, diagnosis and the assessment of the severity of their symptoms were made in the same way as for Experiments 1 and 2.

4.3. Apparatus

4.3.1. Experiment 1: Head touch

For the Head touch experiment, we adapted Gergely et al.'s (2002) paradigm to be safer for use with CWA by replacing the lamp box with a stuffed bee. Upon gentle pressure on the button on its belly, the bee makes a friendly noise accompanied by a light effect. The toy is shown in Fig. 1.

4.3.2. Experiment 2: Hidden box

For the Hidden box experiment (see Fig. 2), we used exact replicas of Király's (2009) objects. Children were presented (1) a small red box with (2) a slightly smaller box hidden inside and a (3) magnet attached to a ball by a short string; this was used to lift the smaller box.

4.4. Procedure

During the experiments, the child was seated at a small table opposite the model. The experimental objects were arranged on the table prior to the child's arrival, and covered with a cloth. The child's parent or caregiver sat behind the child or took the child in her or her lap if necessary. Parents and caregivers were instructed not to interact with the child during the model's demonstration. All sessions were videotaped for later analysis.

4.4.1. Experiment 1: Head touch

The procedure for the Head touch study was identical to that used by Gergely et al. (2002). In the Hands Occupied condition, the model pretended to be cold and asked for her shawl. An assistant wrapped the shawl tightly around the model's shoulders, covering her arms and hands. The model then told the child that she was going to show him or her something interesting. Holding the shawl from underneath, the model bent forward from the waist and pressed the button on the bee's belly with her forehead, eliciting the noise and light effect. The model repeated the action three times, at one-second intervals. The Hands Free condition proceeded in the same manner as did the Hands Occupied condition, with one important exception: the experimenter's hands were free and she placed them on the table with one hand on either side of the object, making it clear that they were not occupied. Each child was assigned to one of the two conditions. After the demonstration, children were provided a short break during they engaged in free play (in fact, 7 DSC had their regular check-up with their doctor during this time- as this regular visit consisted mostly of observing the child during play in the doctor's

Fig. 1. Head touch paradigm apparatus.

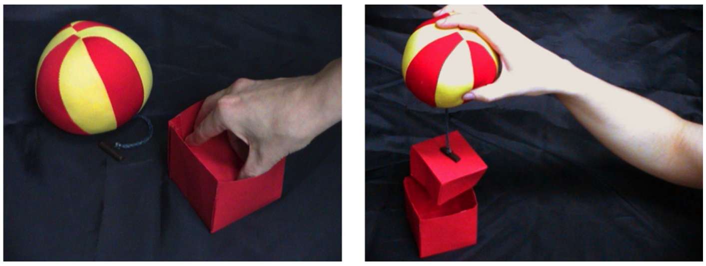


Fig. 2. Hidden box paradigm apparatus. In the photo on the left, the model demonstrates an unsuccessful hand action. In the photo on the right, no negative evidence is provided and the model simply uses the ball to lift the smaller box.

room and asking standard questions from the parents, there was no reason to analyse the responses of these children separately). Following the break, the child was returned to the original room and given the opportunity to play with the toy. The instruction was "Here is the bee; you can play with it now!"

4.4.2. Experiment 2: Hidden box

Our Hidden box study replicated Király's (2009) procedure. In the Hand Action Possible condition, the model held the ball and used the magnet (novel tool) hanging from the ball to connect with the small box and lift it out of the slightly larger box. The model repeated the demonstration three times, at one-second intervals. In the Hand Action Unsuccessful condition, prior to the 'lifting with the ball' action, the model used her hand to attempt to remove the smaller box from the larger box, but failed; her fingers repeatedly slipped off the smaller box and she was unable to hold it. Following the failed efforts, the model successfully completed the ball and magnet action presented in the Hand Action Possible condition. The model repeated the sequence of failed action followed by successful action three times, at one-second intervals. The child was then provided with a short break for free play; as in the Head touch study. After the break, the child was returned to the original room and given the opportunity to play with the toy. The instruction was "Here is the toy; you can play with it now!"

4.5. Scoring

Our method of scoring the children's responses was comparable to that used in most imitation research (see Table 3). If the child did not act on the toy or clearly refused to play with it a score of 0 was given. If the child explored the toy with hands, but did not produce target-relevant acts, a score of 1 was given. Score 2 was given when the child produced a hand action: pressing the toy bee or taking out the small box from the container. Score 3 was given when the child produced the head action on the toy bee or took out the small box from the container using the ball. Interrater agreement, as calculated with Cohen's Kappa statistic ($\kappa = 0.82$), was very good. The dependent variables used for the statistical analysis were: hand action on toy bee, hand action on small box, head action on toy bee and ball action on small box.

5. Results

5.1. Baseline results

In the baseline group, none of the CWA used their forehead to activate the bee or attempted to use the ball to connect with the magnet on the Hidden box. Fisher's exact tests confirmed that baseline responses differed significantly from the responses of children in the experimental conditions. The number of children who produced the head action in the Head touch experiment was significantly greater than in the baseline group ($p = 0.007$). The number of children who used the ball was also significantly greater in the Hidden box experiment than in the baseline group ($p = 0.005$).

5.1.1. Experiment 1. Effect of diagnosis in the Head touch experiment

Statistical analysis on the amount of imitators in each condition (Hands Free and Hands Occupied conditions) revealed a significant difference between TDC's responses ($\text{Chi}^2 = 4.22$; $df = 1$; $p = 0.04$). The frequency with which typical children imitated the model's head action decreased significantly when the model's hands were occupied during the demonstration. These results are consistent with Gergely et al.'s (2002) findings on selective imitation in infancy, and validate the use of the stuffed bee as an experimental object.

Table 3
Scoring method for the Head touch and the Hidden box experiments.

Score	Response
0	Does not act on toy or clearly refuses to play with it
1	Explores toy with hands, but does not produce target-relevant acts
2	Produces hand action: pressing toy bee or taking out box
3	Produces head action or takes out box with the ball

Table 4
Responses of children with autism, children with Down syndrome and typically developing children in the Head touch experiment.

Condition	Diagnosis	Head action
Hands free	Autism	57
	Down syndrome	67
	Typical	64*
Hands occupied	Autism	71
	Down syndrome	71
	Typical	15*

Note: The values represent percentages of children who re-enacted the head action.
* $p < 0.01$.

Log-linear analyses were conducted to explore the effect of clinical diagnosis (independent variables: autism, Down syndrome or typical) and experimental condition (independent variables: Hands Free or Hands Occupied) on the children's responses (dependent variable: reproduction of the head action). Significant differences were revealed when comparing the responses of the typical group with both clinical groups (for differences between the typical and the autism group: G2 = 10.7, df = 4, $p < 0.05$; for differences between the typical and the DS group; G2 = 14.6, df = 4, $p < 0.01$). When comparing the responses of the two clinical groups, no significant effects or interactions were observed (G2 = 3.11, df = 4, $p = 0.54$). There was no main effect of experimental condition, indeed.

Subsequent chi-square tests as follow up analyses were conducted to examine the pattern of differences between typical and clinical groups. We compared the responses of the typical group with the autism group for both conditions and found a significant effect of autism on the reproduction of the head action in the Hands Occupied condition ($Chi^2 = 8.65$; $df = 1$; $p = 0.03$), but not in the Hands Free condition ($Chi^2 = 0.21$; $df = 1$; $p = 0.65$). Comparing the responses of the typical group with the DS group for both conditions, again the effect of diagnosis was found in the Hands Occupied condition ($Chi^2 = 9.55$; $df = 1$; $p = 0.00$) and there was no effect in the Hands Free condition ($Chi^2 = 1.04$; df = 1; $p = 0.31$).

All in all, these results show that clinical diagnosis had a significant effect on the reproduction of the head action in the Hands Occupied condition. CWA and DSC produced the head action significantly more frequently than typicals in the Hands Occupied condition. The diagnosis of autism, however, did not have a specific effect, CWA and DSC responded in the same way. Table 4 summarises the results obtained for the three groups of children in the two conditions of the Head touch paradigm.

We also investigated hand actions across experimental groups and obtained similar results to Gergely et al. (2002): all (100%) of the TDC who produced the head action also touched the toy with their hands. This is an expected effect: manual manipulation is the prepotent instrumental response associated with achieving an effect. Similarly, the majority of the CWA and DSC who reproduced the head action (73% and 67%, respectively) produced the hand action first. Fisher's exact tests revealed no significant differences between the three experimental groups in this respect ($p = 0.68$).

5.1.2. Experiment 2. Effect of diagnosis in the Hidden box experiment

In this experiment, two action steps were demonstrated in one of the conditions, the Hand Action Unsuccessful condition (a hand action that served as negative evidence and the target ball action), following Király (2009). For this reason, not only (1) the reproduction of the ball action, but also (2) the production of the hand action, and (3) their sequence was analysed, since Király (2009) reports selective production in typical infants for both actions. In her study, infants in the Hand Action Unsuccessful condition performed the target ball action more frequently, together with inhibiting the prepotent hand action. Thus, there was a selective effect of demonstration condition regarding the *sequence* of these responses. From this perspective, the pattern of hand and ball actions can reflect whether the model's overall intention was detected (inhibiting the prepotent hand action together with performing the target action), or the exact copying of behaviour occurred (performing the demonstrated hand action together with the target action). For this reason, we involved all three variables in the present analyses as well.

Log-linear analyses were performed to explore the effect of clinical diagnosis (independent variables: autism or Down syndrome), experimental condition (independent variables: Hand Action Unsuccessful or Hand Action Possible) on the children's responses. We conducted three separate analyses with the following dependent variables:(1) reproduction of the demonstrated ball action only, (2) reproduction of the hand action only and (3) reproduction of the hand and ball action in sequence. No significant interactions between diagnosis and experimental condition were observed for (1) the ball action alone (G2 = 2.28, df = 4, $p = 0.68$). This is in striking contrast with Király's (2009) results; Király reported that typically developing 14-month-olds demonstrated selectivity in frequency of ball actions across the two conditions, and were less likely to reproduce the use of the novel tool when no negative evidence about the more rational hand action was presented. The interaction between diagnosis and experimental condition was significant, however, for reproduction of (2) the hand action alone (G2 = 9.7, df = 4, $p = 0.008$) and (3) the sequence of hand and ball actions (G2 = 21.74, df = 4, $p < 0.001$).

Subsequent Fisher's exact tests as follow up analyses revealed a significant effect of experimental condition in the autism group for the hand action ($p = 0.003$) and for the hand and ball sequence ($p = 0.0005$). CWA touched the smaller box with their hands and produced the sequence of means actions significantly more frequently than DSC in the Hand Action Unsuccessful

Table 5
Imitative reproduction of modelled behaviour in children with autism and children with Down syndrome in the Hidden box experiment.

Condition	Diagnosis	Hand action	Both hand and ball actions
Hand action possible	Autism	13**	0**
	Down syndrome	22	11
	Typical (as reported in Király, 2009)	85	15
Hand action unsuccessful	Autism	100**	100**
	Down syndrome	45	36*
	Typical (as reported in Király, 2009)	36	0

Note. The values represent percentages of children who produced the hand action or both means actions.
* $p < 0.01$.
** $p < 0.001$.

condition. No effect of experimental condition was observed in the Down syndrome group for the hand action ($p = 0.27$) or for the hand and ball sequence ($p = 0.22$); children in this group produced the hand action and the sequence of means actions with equal frequency in the two experimental conditions. Table 5 summarises the two clinical groups' results in the two conditions of the Hidden box paradigm.

6. Discussion

The primary objective of the current study was to explore interpretations of a model's behaviour in simplified social situations in a group of nonverbal, L-F CWA. We wished to explore (1) whether the ability to attribute goals is preserved in L-F CWA, just as in H-F CWA (Abell et al., 2000; Castelli, 2006; Hamilton et al., 2007; Rogers et al., 2010) and (2) whether L-F CWA (a) will show ability to attribute intentions in tasks where functionality or usual affordances cannot bias the children's responses or (b) respond by copying exactly the experimenter's actions, a strategy described by studies involving H-F CWA (Carpenter et al., 2001; D'Entremont & Yazbek, 2007).

We devised imitation tasks involving object-directed actions that were not common everyday actions and could only have been acquired from the demonstration in the laboratory. The experimenter's intention across these tasks was marked by nonverbal cues.

Our results for the typical infants repeat Gergely et al.'s (2002) finding that 14-month-olds are sensitive to contextual features that provide information about the model's circumstances and about the rationale for an unusual action. Infants' imitative performance in such situations is guided by a selective interpretive process that involves evaluation of the relative effectiveness of the action as a function of the model's constraints. CWA and DSC frequently reproduced the unusual head action in the Head touch paradigm and the novel ball action in the Hidden box paradigm, indicating that they interpreted the model's behaviour as a teleological, means-end action. In the Head touch paradigm CWA and DSC also reproduced the prepotent hand action in order to achieve the objective, demonstrating that they identified both the means and the end-state as the model's goals. This finding shows that the ability to attribute goals is preserved in L-F CWA, confirming our first hypothesis, namely that goal attribution is generally preserved in ASD.

Although CWA and DSC both demonstrated a teleological interpretation of the model's behaviour, our results indicated that they did not take the model's circumstances into consideration and did not use context to interpret the model's unusual and non-affordant actions in the same manner as did TDC. In the clinical groups, imitative behaviour did not show the same selective pattern across demonstration conditions as in the typical group. Current research on contextual constraints in selective imitation raises the possibility that, for typically developing infants, communicative cues during a behavioural demonstration convey information about a model's intended goal (Csibra & Gergely, 2009; Király, Csibra, & Gergely, 2012). Our results demonstrate that the use of ostensive-communicative cues during the demonstration did not make the situation more transparent for CWA and DSC. Although children in these two groups succeeded in learning the novel means presented, they did not interpret the model's behaviour in terms of her intended goal.

This replicates the results reported in D'Entremont and Yazbek's study (2007), where the authors concluded that, (consistently with Carpenter et al.'s (2001) inferences), H-F CWA were guided by stimulus enhancement, that is, the model's actions highlighted affordances of the objects and the children subsequently performed the actions that the objects afforded. In our study, the demonstrated actions were not consistent with the objects' affordances (to touch a button with the head or to use a ball as a lifter), and CWA still imitated all the actions in the stream of the demonstrator's behaviour. Regarding the first part of our second hypothesis, we can therefore conclude that even in better-controlled tasks where functionality or usual affordances could not trigger the children's responses, L-F CWA did not show ability to attribute intention.

DSC clearly engage in social interaction, and the lack of inferential processing and selectivity observed in DSC is therefore probably not attributable to an avoidance of cues. Rather, our evidence that DSC do not learn from failed attempts suggests that they are unable to harness the information presented, possibly as a result of memory or inferential processing problems.

The ceiling effect obtained in the Head touch experiment may be attributable to the simplicity of the experimental setting, where the saliency of the head action may have averted the children's attention from the cues indicating the model's intention. We controlled this possibility in the Hidden box experiment, where we increased the explicitness of the model's intention. In this setting, rather than using a situational constraint (tying the model's hands), an additional explicit behavioural step (a failed action) was used to indicate the reason for the model's non-affordant action. In addition, the Hidden box experiment created identical situational constraints for the model and the child, which can also be expected to make the situation more transparent for the child. Again, children in the two clinical groups reproduced the model's non-affordant action (the ball action) with equal frequency across the two experimental conditions. This indicates that it was not the saliency of the actions presented that lead L-F CWA to imitate these regardless of the model's intention. Instead, their responses reflected a specific imitative strategy: after witnessing the adult model's failed efforts, unlike TDC, they included the failed attempt in their response and reproduced the hand action, as well as the exact hand and ball sequence significantly more in the failed attempt condition than in the condition where no failed attempt was demonstrated. We did not observe similar selectivity for the reproduction of the hand action in DSC; that is, the children reproduced the action with equal frequency in the two conditions. Regarding the last part of our second hypothesis, we can therefore conclude that copying exactly the experimenter's actions is strategy that CWA, H-F or L-F, generally employ to manage social situations.

It is interesting to note that, although the addition of a behavioural step could render the underlying reasons for an overall action more opaque, a contrary effect was observed in Király's (2009) study of typically developing infants. Here, the additional step helped typical children interpret the model's unaffordant action with the ball, and they did not retain this 'indicator' action (although a prepotent one) when it was their turn to interact with the experimental objects. The pattern observed in L-F CWA was quite the opposite.

Why do CWA, H-F or L-F use exact imitation to manage social situations? Rather than analyzing the context – behavioural, situational, or mental – of an observed behaviour and selecting the most efficient means to achieve an observed goal, why do they reproduce the entirety of a model's demonstrated actions? As pointed out by Gergely et al. (2002), without the ability to take into account the context of a presented behaviour, most social situations become cognitively opaque or difficult to interpret. The strategy that CWA applied across our experimental conditions seems appropriate for addressing this difficulty. That is, reproducing the entire sequence of witnessed behaviour is a redundant but safe method of producing relevant and efficient action (Gergely & Csibra, 2006). It remains possible though, that similarly to H-F CWA in Vivanti et al.'s study (2011) L-F CWA do take into account the environment in which the action occurs, but are unable to analyse or interpret in a second step the referential cues observed in the situation.

Three main conclusions can be drawn from the results of the present study. First, consistently with the studies showing that in high-functioning autism there is an intact ability to understand others' goals (Hamilton, 2009), we show that children at the other end of the spectrum, L-F CWA imitate goal-directed actions. This finding implies that they perceive other people as goal-directed agents whose behaviour is driven by the desire to achieve an end state. L-F CWA not only reproduced the most efficient means action to achieve the objective, but also reproduced the model's unusual means action. This indicates that the L-F CWA in this study differentiated between the two methods presented for achieving the goal; however, the model's ostensive-communicative cues did not guide the children in determining which method was more congruent with the model's intended goal.

Second, the results presented here indicate that nonverbal L-F CWA do not generate a hierarchical interpretation of the means act as a function of the goal and of the model's circumstances or constraints. In other words, CWA's imitative performance was not influenced by factors that account for selectivity in TDC. One possible explanation is that CWA have deficits in inhibition and difficulty applying top-down selection rather than primed motor programmes activated through motor resonance. We suggest that CWA use alternative strategies to cope with social situations. The strategy revealed by the experimental paradigms in the present study consists of reproducing the full set of behaviours witnessed, without selectivity. This strategy may allow CWA to respond appropriately in social situations.

Third, that our sample of CWA readily participated in the study and responded to the experimental conditions demonstrates that deferred imitation tasks are indeed an appropriate means to evaluate the strategies that this population – otherwise very difficult to evaluate – uses to navigate their environment.

References

Abell, F., Happe, F., & Frith, U. (2000). Do triangles play tricks? Attribution of mental states to animated shapes in normal and abnormal development. *Cognitive Development, 15*(1), 1–16.

Aldridge, M. A., Stone, K. R., Sweeney, M. H., & Bower, T. G. R. (2000). Preverbal children with autism understand the intentions of others. *Developmental Science, 3*(3), 294–301.

Anderson, D. K., Lord, C., Risi, S., DiLavore, P. S., Shulman, C., Thurm, A., et al. (2007). Patterns of growth in verbal abilities among children with autism spectrum disorder. *Journal of Consulting and Clinical Psychology, 75*(4), 594–604.

Avikainen, S., Wohlschläger, A., Liuhanen, S., Hänninen, R., & Hari, R. (2003). Impaired mirror-image imitation in Asperger and high-functioning autistic subjects. *Current Biology, 13*(4), 339–341. http://dx.doi.org/10.1016/S0960-9822(03)00087-3

Baron-Cohen, S. (1995). *Mindblindness: An essay on autism and theory of mind*. Cambridge: MIT Press.

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a theory of mind? *Cognition, 21*(1), 37–46.

Beadle-Brown, J. (2004). Elicited imitation in autistic children and adults – The effects of different types of actions. *Journal of Applied Research in Intellectual Disabilities, 17*(1), 31–48.

Bedford, R., Elsabbagh, M., Gliga, T., Pickles, A., Senju, A., Charman, T., et al. (2012). Precursors to social and communication difficulties in infants at-risk for autism: gaze following and attentional engagement. *Journal of Autism and Developmental Disorders, 42*(10), 2208–2218. http://dx.doi.org/10.1007/s10803-012-1450-y

Brunet-Lézine. (2001). *Échelle de développement psychomoteur de la première enfance révisée (revised scale of psychomotor development of infancy)*. Paris: Éditions du centre de psychologie appliquée.

Carpenter, M., Pennington, B. F., & Rogers, S. J. (2001). Understanding of others' intentions in children with autism. *Journal of Autism and Developmental Disorders, 31*(6), 589–599.

Castelli, F. (2006). The Valley task: Understanding intention from goal-directed motion in typical development and autism. *British Journal of Developmental Psychology, 24*, 655–668.

Castelli, F., Frith, U., Happe, F., & Frith, C. D. (2001). Autism and the perception of intentionality in moving geometrical shapes. *NeuroImage, 13*(6 (Suppl 1)), 1035.

Charman, T., & Baron-Cohen, S. (1994). Another look at imitation in autism. *Development and Psychopathology*, 6: 403–413.

Charman, T., Baron-Cohen, S., Swettenham, J., Baird, G., Drew, A., & Cox, A. (2003). Predicting language outcome in infants with autism and pervasive developmental disorder. *International Journal of Language and Communication Disorders, 38*(3), 265–285. http://dx.doi.org/10.1080/136820310000104830

Csibra, G., & Gergely, G. (2009). Natural pedagogy. *Trends in Cognitive Sciences, 13*(4), 148–153.

D'Entremont, B., & Yazbek, A. (2007). Imitation of intentional and accidental actions by children with autism. *Journal of Autism and Developmental Disorders, 37*(9), 1665–1678. http://dx.doi.org/10.1007/s10803-006-0291-y

Frith, U., Morton, J., & Leslie, A. M. (1991). The cognitive basis of a biological disorder: Autism. *Trends in Neurosciences, 14*(10), 433–438.

Gergely, G., Bekkering, H., & Király, I. (2002). Rational imitation in preverbal infants. *Nature, 415*, 755.

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

Gergely, G., & Csibra, G. (2006). Sylvia's recipe: The role of imitation and pedagogy in the transmission of human culture. In N. J. Enfield & S. C. Levinson (Eds.), *Roots of human sociality: Culture, cognition, and human interaction* (pp. 229–255). Oxford: Berg Publishers.

Gómez, J. C. (2009). Embodying meaning: Insights from primates, autism, and Brentano. *Neural Networks, 22*(2), 190–196. http://dx.doi.org/10.1016/j.neunet.2009.01.010

Hamilton, A. (2009). Research review: Goals, intentions and mental states: Challenges for theories of autism. *Journal of Child Psychology and Psychiatry, 50*(8), 881–892. http://dx.doi.org/10.1111/j.1469-7610.2009.02098.x

Hamilton, A., Brindley, R., & Frith, U. (2007). Imitation and action understanding in autistic spectrum disorders: How valid is the hypothesis of a deficit in the mirror neuron system? *Neuropsychologia, 45*(8), 1859–1868.

Hobson, R. P., & Hobson, J. A. (2008). Dissociable aspects of imitation: A study in autism. *Journal of Experimental Child Psychology, 101*(3), 170–185. http://dx.doi.org/10.1016/j.jecp.2008.04.007

Huang, T.-C., Heyes, C., & Charman, T. (2002). Infants' behavioural reenactment of failed attempts: Exploring the roles of emulation learning, stimulus enhancement, and understanding of intentions. *Developmental Psychology, 38*, 840–855.

Huang, T.-C., Heyes, C., & Charman, T. (2006). Preschoolers' behavioural reenactment of failed attempts: The roles of intention-reading, emulation and mimicry. *Cognitive Development, 21*, 36–45.

Hudry, K., Leadbitter, K., Temple, K., Slonims, V., McConachie, H., Aldred, C., et al. (2010). Preschoolers with autism show greater impairment in receptive compared with expressive language abilities. *International Journal of Language and Communication Disorders, 45*(6), 681–690. http://dx.doi.org/10.3109/13682820903461493

Ingersoll, B., Schreibman, L., & Tran, Q. H. (2003). Effect of sensory feedback on immediate object imitation in children with autism. *Journal of Autism and Developmental Disorders, 33*(6), 673–683.

Király, I. (2009). The effect of the model's presence and of negative evidence on infants' selective imitation. *Journal of Experimental Child Psychology, 102*(1), 14–25.

Király, I., Csibra, G., & Gergely, G. (2012). Beyond rational imitation: Learning arbitrary means actions from communicative demonstrations. *Journal of Experimental Child Psychology0:* http://dx.doi.org/10.1016/j.jecp.2012.12.003

Kjelgaard, M. M., & Tager-Flusberg, H. (2001). An investigation of language impairment in autism: Implications for genetic subgroups. *Language and Cognitive Processes, 16*(2-3), 287–308. http://dx.doi.org/10.1080/01690960042000058

Leslie, A. M., & Thaiss, L. (1992). Domain specificity in conceptual development: Evidence from autism. *Cognition, 43*, 467–479.

Lord, C., & Luyster, R. (2006). Early diagnosis of children with autism spectrum disorders. *Clinical Neuroscience Research, 6*(3-4), 189–194.

Luo, Y. (2011). Do 10-month-old infants understand others' false beliefs? *Cognition, 121*(3), 289–298. http://dx.doi.org/10.1016/j.cognition.2011.07.011

Maljaars, J., Noens, I., Scholte, E., & van Berckelaer-Onnes, I. (2012). Language in low-functioning children with autistic disorder: Differences between receptive and expressive skills and concurrent predictors of language. *Journal of Autism and Developmental Disorders, 42*(10), 2181–2191. http://dx.doi.org/10.1007/s10803-012-1476-1

Mayes, S. D., Calhoun, S. L., Murray, M. J., Morrow, J. D., Yurich, K. K. L., Cothren, S., et al. (2012). Use of the Childhood Autism Rating Scale (CARS) for children with high functioning autism or Asperger syndrome. *Focus on Autism and Other Developmental Disabilities, 27*(1), 31–38. http://dx.doi.org/10.1177/1088357611406902

McDuffie, A., Yoder, P., & Stone, W. (2005). Prelinguistic predictors of vocabulary in young children with autism spectrum disorders. *Journal of Speech, Language, and Hearing Research, 48*(5), 1080–1097. http://dx.doi.org/10.1044/1092-4388(2005/075)

Meltzoff, A. (1985). Immediate and deferred imitation in fourteen- and twenty-four-month-old infants. *Child Development, 56*(1), 62–72.

Nadel, J., Aouka, N., Coulon, N., Gras-Vincendon, A., Canet, P., Fagard, J., et al. (2011). Yes they can! An approach to observational learning in low-functioning children with autism *Autism, 15*(4), 421–435. http://dx.doi.org/10.1177/1362361310386508

Nadel, J., Croué, S., Mattlinger, M.-J., Canet, P., Hudelot, C., Lecuyer, C., et al. (2000). Do children with autism have expectancies about the social behaviour of unfamiliar people? A pilot study using the still face paradigm. *Autism, 4*(2), 133–145.

Noens, I., & van Berckelaer-Onnes, I. (2004). Making sense in a fragmentary world: Communication in people with autism and learning disability. *Autism, 8*(2), 197–218. http://dx.doi.org/10.1177/1362361304042723

Perner, J., Frith, U., Leslie, A. M., & Leekam, S. R. (1989). Exploration of the autistic child's theory of mind: knowledge, belief, and communication. *Child Development, 60*(3), 688–700.

Phillips, A., Baron-Cohen, S., & Rutter, M. (1998). Understanding intention in normal development and in autism. *British Journal of Developmental Psychology, 16*(3), 337–348.

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences, 4*, 515–526.

Rogers, S., Bennetto, L., McEvoy, R., & Pennington, B. (1996). Imitation and pantomime in high-functioning adolescents with autism spectrum disorders. *Child Development, 67*(5), 2060–2073.

Rogers, S. J., Young, G. S., Cook, I., Giolzetti, A., & Ozonoff, S. (2010). Imitating actions on objects in early-onset and regressive autism: Effects and implications of task characteristics on performance. *Development and Psychopathology, 22*(1), 71–85. http://dx.doi.org/10.1017/s0954579409990277

Russell, J., & Hill, E. L. (2001). Action-monitoring and intention reporting in children with autism. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 42*, 317–328.

Saxe, R., Carey, S., & Kanwisher, N. (2004). Understanding other minds: Linking developmental psychology and functional neuroimaging. *Annual Review of Psychology, 55*(1), 87–124. http://dx.doi.org/10.1146/annurev.psych.55.090902.142044

Schopler, E., Lansing, M. D., Reichler, R. J., & Marcus, L. M. (2005). *Examiner's manual of psychoeducational profile* (Vol. 3). Austin, TX: Pro-ed Incorporation.

Schopler, E., Reichler, R. J., Bashford, A., Lansing, M. D., & Marcus, L. M. (1990). *The psychoeducational profile revised (PEP-R)*. Austin: Pro-Ed.

Schopler, E., Reichler, R. J., & Renner, B. R. (1988). *The Childhood Autism Rating Scale (CARS)*. Los Angeles: Western Psychological Services.

Smith, I. M., & Bryson, S. E. (1994). Imitation and action in autism: A critical review. *Psychological Bulletin, 116*(2), 259–273.

Sodian, B., & Frith, U. (1994). Deception and sabotage in autistic, retarded and normal children. *Journal of Child Psychology and Psychiatry, 24*, 591–605.

Stone, W. L., Ousley, O. Y., & Littleford, C. D. (1997). Motor imitation in young children with autism: What's the object? *Journal of Abnormal Child Psychology, 25*(6), 475–485.

Vivanti, G., McCormick, C., Young, G., Abucayan, F., Hatt, N., Nadig, A., et al. (2011). Intact and impaired mechanisms of action understanding in autism. *Developmental Psychology, 47*(3), 841–856.

Wing, L., Gould, J., Yeates, S. R., & Brierley, L. M. (1977). Symbolic play in severely mentally retarded and in autistic children. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 18*(2), 167–178.

dc_1353_16

Journal of Experimental Child Psychology 116 (2013) 471–486

Contents lists available at SciVerse ScienceDirect

Journal of Experimental Child Psychology

journal homepage: www.elsevier.com/locate/jecp

Beyond rational imitation: Learning arbitrary means actions from communicative demonstrations

Ildikó Király [a,*], Gergely Csibra [b], György Gergely [b]

[a] Department of Cognitive Psychology, Eötvös Loránd University, Budapest H-1064, Hungary
[b] Department of Cognitive Science, Central European University, Budapest H-1051, Hungary

ARTICLE INFO

Article history:
Available online 15 March 2013

Keywords:
Rational imitation
Relevance-guided imitation
Teleological stance
Natural pedagogy
Social Learning
Ostensive communication

ABSTRACT

The principle of rationality has been invoked to explain that infants expect agents to perform the most efficient means action to attain a goal. It has also been demonstrated that infants take into account the efficiency of observed actions to achieve a goal outcome when deciding whether to reenact a specific behavior or not. It is puzzling, however, that they also tend to imitate an apparently suboptimal unfamiliar action even when they can bring about the same outcome more efficiently by applying a more rational action alternative available to them. We propose that this apparently paradoxical behavior is explained by infants' interpretation of action demonstrations as communicative manifestations of novel and culturally relevant means actions to be acquired, and we present empirical evidence supporting this proposal. In Experiment 1, we found that 14-month-olds reenacted novel arbitrary means actions only following a communicative demonstration. Experiment 2 showed that infants' inclination to reproduce communicatively manifested novel actions is restricted to behaviors they can construe as goal-directed instrumental acts. The study also provides evidence that infants' reenactment of the demonstrated novel actions reflects epistemic motives rather than purely social motives. We argue that ostensive communication enables infants to represent the teleological structure of novel actions even when the causal relations between means and end are cognitively opaque and apparently violate the efficiency expectation derived from the principle of rationality. This new account of imitative learning of novel means shows how the teleological stance and natural pedagogy—two separate cognitive adaptations to interpret

* Corresponding author.
E-mail address: kiralyi@caesar.elte.hu (I. Király).

http://dx.doi.org/10.1016/j.jecp.2012.12.003

instrumental versus communicative actions—are integrated as a system for learning socially constituted instrumental knowledge in humans.

Introduction

The principle of rationality as a fundamental factor governing action selection in knowledge-based systems was introduced by Newell (1982): "If an agent has knowledge that one of its actions will lead to one of its goals, then the agent will select that action" (p. 102). The rationality principle has also been proposed to be the central inferential principle in Gergely and Csibra's (2003) theory of human infants' teleological action interpretation mechanism. This core system is a cognitive adaptation to represent instrumental actions in terms of their teleo-functional properties. It implements human infants' naive theory of rational action in the domain of instrumental agency by representing actions as efficient means to bring about specific goal states in the world. The main tenets of the rationality principle are that (a) actions serve to bring about future goal states and (b) goal states are realized by the most efficient action available to the actor within the constraints of the situation (Gergely & Csibra, 2003). Teleological reasoning (just like mentalistic or practical reasoning about actions) relates three aspects of action interpretation—goals, actions, and situational constraints—in a systematic manner by the "rationality assumption": Given information about any two of the three elements, one can infer (and predict) what the third element ought to be (Csibra, Bíró, Koós, & Gergely, 2003).

Using violation of expectation and eye-tracking paradigms, ample evidence confirms that infants can make inferences about observed actions with the help of teleological reasoning (Gergely, Nadasdy, Csibra, & Bíró, 1995; see also Biro, Csibra, & Gergely, 2007; Biro, Verschoor, & Coenen, 2011; Csibra, 2007; Csibra, Gergely, Bíró, Koós, & Brockbank, 1999; Csibra et al., 2003; Gredebäck & Melinder, 2010; Gredebäck & Melinder, 2011; Hernik & Southgate, 2012; Kamewari, Kato, Kanda, Ishiguro, & Hiraki, 2005; Sodian, Schoeppner, & Metz, 2004; Southgate & Csibra, 2009; Verschoor & Biro, 2012; Wagner & Carey, 2005; Woodward & Sommerville, 2000).

If infants expect other agents to act rationally by choosing to perform the most efficient means available to the goal, one would expect infants themselves to rely on the same principle of rationality to guide their own choices of instrumental actions as well. This prediction, however, was apparently contradicted by the results of Meltzoff's (1988) seminal imitation study, in which infants chose to reenact a model's unusual and subefficient head action to illuminate a light box instead of just using their hands to induce the same effect. Optimizing to "least effort," operating the light box by touching it with one's hand seems more efficient (and hence more rational) than bending forward from the waist to use one's forehead to achieve the same end. To address this puzzle, Gergely, Bekkering, and Király (2002) developed a modified version of Meltzoff's (1988) imitation paradigm to test whether efficiency evaluations could modify infants' action choice and production by introducing a new context condition in which the demonstrator's hands were occupied when she performed the unfamiliar head action to operate the touch lamp (*hands-occupied* condition). Whereas in the original *hands-free* condition 69% of infants reenacted the head action (replicating Meltzoff's results), the number of "imitators" dropped significantly to only 21% in the hands-occupied condition. Gergely and colleagues (2002) referred to this phenomenon of context-sensitive learning of novel means actions as "rational imitation." Since the original demonstration, the finding of selective rational imitation has been replicated several times and shown to generalize across a range of different task contexts with 12- and 14-month-olds (Buttelmann, Carpenter, Call, & Tomasello, 2008; Király, 2009a; Schwier, van Maanen, Carpenter, & Tomasello, 2006; Zmyj, Daum, & Aschersleben, 2009).

The explanation of context-sensitive and selective reenactment of novel actions in terms of rational imitation, however, has been challenged on several grounds. One alternative account developed by Paulus, Hunnius, Vissers, and Bekkering (2011a, 2011b) suggests that the phenomenon is attributable to the interaction between automatic motor resonance elicited by the observed actions and the limited motor capabilities of infants. Another contrasting view was proposed by Beisert, Zmyj, Liepelt, Jung, Prinz and Daum (2012), who considered the selectivity of infants' action imitations to differences in attentional factors that are assumed to be induced by the different levels of saliency of the head action when observed in the hands-free versus hands-occupied contexts. Finally, our current account also argues (on different grounds) that the original explanation of selective head touch imitation purely in terms of the application of the rationality principle fails to capture all of the relevant aspects of the phenomenon. Below we advance and empirically test a new proposal that combines teleological action understanding (Gergely & Csibra, 2003) with the theory of natural pedagogy (Csibra & Gergely, 2009; Gergely & Jacob, 2012) to provide a more satisfactory explanatory account of the nature of selective imitation and learning of communicatively demonstrated novel means actions (for an earlier formulation of this hypothesis, see Gergely & Csibra, 2005, 2006). Because the original demonstration of selective rational imitation and the different alternative accounts proposed have all concentrated on the head touch paradigm, we decided to use the very same task to test our new proposal as well in order to make a systematic comparison of the various theoretical accounts possible.

The original rational imitation explanation as presented by Gergely and colleagues (2002; see also Buttelmann et al., 2008) focused on infants' differential performance in the new hands-occupied condition and attributed the decreased likelihood of reenacting the unfamiliar head action to infants' evaluation of the rationality of the demonstrated act to bring about the goal given the context of the demonstrator's situational constraints. But why did the majority of infants reenact the novel head action in the hands-free condition? Here both the model's and infant's own hands were free, so infants could have (rationally) opted for emulating the goal by performing the more efficient "hand action" available to them to contact the light box rather than reenacting the awkward and clearly suboptimal nonrational head action.[1] And, as a matter of fact, they did so; all infants in Gergely and colleagues' (2002) study (see also Paulus et al., 2011b) performed at least one hand action to bring about the goal, and they typically chose to perform the hand action as their first response to operate the light box. Even more striking, and in spite of that, 69% of the infants in the hands-free condition went on to reenact the novel head action of the model.

The rational imitation proposal (Gergely et al., 2002; see also Buttelmann et al., 2008) suggests that the reproduction of this apparently nonrational action reflected infants' assessment of the novel act as manifesting some unknown reason that must justify the action as rational and so they reproduced the odd head action as a way of figuring (learning) what the agent's (rational) reason for his action might have been. The main problem with this proposal as it stands is that it essentially makes the idea of appealing to the rationality principle unfalsifiable; when infants did not reproduce the demonstrated action (in the hands-occupied condition), it was treated as evidence of the application of the rationality principle, and when they did reproduce it (in the hands-free condition), it was also interpreted as evidence for the operation of the same inferential principle (cf. Gergely & Jacob, 2012).

Our new approach to the selective imitation phenomenon attempts to explain not only why infants refrained from imitating the novel action when it seemed rational but also why they reproduced it when it was obviously not efficient. Our account relies on the theory of natural pedagogy (Csibra & Gergely, 2009, 2011). Gergely and Csibra (2005, 2006) argued that children's tendency to reproduce novel actions of others reflects the functioning of an adaptive learning system, which allows for the fast transfer of relevant knowledge between individuals. In this view, imitation of behavior does not itself serve a causal role per se in learning; nevertheless, the reenactment of the novel action does provide evidence that learning has taken place and about what infants have learned. According to natural pedagogy theory, ostensive communicative signals, such as eye contact and child-directed speech (Csibra, 2010), indicate to children that the information about to be communicated provides them with the opportunity to acquire some new and relevant knowledge. From this perspective, the results of Gergely and colleagues' (2002) modified head touch study can be reinterpreted as follows. When ostensive signals are presented to infants and the model produces her subsequent actions deliberately, children interpret the demonstrated action as a manifestation of a communicative intention of the model rather than as a merely instrumental action. The ostensively induced expectation that

[1] Note that in Meltzoff's (1988) study, in the no-modeling baseline condition there was no infant who used the head to light up the lamp. This result confirms that the head action is a rather unusual means to perform on the lamp (see also Zmyj et al., 2009, for similar results with a novel lamp setup).

informative new and relevant knowledge is about to be manifested would suggest that the subsequently demonstrated action is important to acquire regardless of the fact that it seems not to be the most efficient way to achieve its apparent goal, that is, regardless of the opacity of the model's choice of the demonstrated action to bring about the effect. This account is also supported by further empirical evidence indicating that infants (Brugger, Lariviere, Mumme, & Bushnell, 2007; Nielsen, 2006) and even adults (Wang, Newport, & Hamilton, 2010) show more imitation in the presence of ostensive communicative signals such as eye contact.

When proposing that action imitation in communicative contexts often reflects learning, we do not intend to suggest that this epistemic function is limited to learning about the physical world only. In fact, when children learn about the function and manner of use of novel artifacts, they learn not only about the causal dispositional properties of the physical object world but also about relevant and shared normative dispositional properties of the social world around them. (An obvious characteristic of many, but by no means all, cultures is the use of cutlery for eating, which is the socially accepted and sanctioned normative manner of consuming food even though eating with hands is clearly much easier and is universally preferred by children. Moreover, different cultures have developed a variety of culture-specific artifacts specialized for eating as well as opaque but normative manners of using them to satisfy the need to consume food). In fact, artifacts are cultural products serving a variety of functions whose demonstrated use may manifest shared social knowledge not only of hidden dispositional affordance properties of the artifact (and its kind) but also of social norms and conventional functions that their use may involve or serve (Rakoczy, Warneken, & Tomasello, 2008; Tomasello, 1999). Note that the ostensively induced "basic epistemic trust" (Gergely, Egyed, & Király, 2007) and the social motivation to acquire shared cultural knowledge (including social conventions, manners of use, and traditions) from communicative action manifestations result in infants' spontaneous propensity to acquire *cognitively opaque* forms of actions demonstrated to them. The epistemic motivation to learn about opaque but socially shared cultural norms and conventions can easily lead to "overimitation" (the general tendency evidenced by older children to faithfully reenact relevant as well as apparently irrelevant steps of demonstrated action sequences; see Lyons, Damrosch, Lin, Macris, & Keil, 2011; Lyons, Young, & Keil, 2007). For example, a recent study of Kenward (2012) shows that preschoolers imitate actions that they already discovered to be unnecessary to achieve an outcome because they conceive them as norms.

But how do children know which elements of a demonstrated action sequence constitute instrumental and/or social norms to be acquired? We propose that communicative demonstrations elicit the presumption of relevance in their addressee (cf. Sperber & Wilson, 1986). When the model deliberately produces her instrumental actions while engaging in ostensive communication with infants despite the fact that her action is causally opaque (i.e., seemingly not the most efficient way to achieve the goal state), her demonstration activates the presumption of relevance in infant learners. This presumption guides the interpretation of the demonstration by searching for a communicative content that appears to be in some way relevant to recipients. If the action involves a novel artifact, the most relevant information for novice learners is (a) its function and (b) its mode of usage as dictated by instrumental or social norms. It is important to note that because of the opaque nature of much cultural knowledge in human societies (Gergely & Csibra, 2006), naive learners cannot expect to derive "b" from "a" by applying the rationality principle, but the principle can still be employed to disregard action elements of the observed demonstration that are justifiable by situation-specific physical constraints.

Thus, we propose the following reinterpretation of the selective imitation phenomenon demonstrated by Gergely and colleagues' (2002). Because of the ostensive signals that accompany the action demonstration, infants construe it as a communicative action rather than a purely instrumental action (Gergely & Jacob, 2012; Southgate, Chevallier, & Csibra, 2009). The fact that the action is performed on a novel artifact and produces a salient effect will suggest to infants that the relevant information to be acquired is the function and usage of the artifact. In the hands-occupied condition, they learn a final goal state (artifact function: the lamp is illuminated) and a subgoal (usage: the lamp is operated by contact) that brings about the final goal. The relation between these two goals is arbitrary because infants have no way to assess the causal mechanisms by which the subgoal produces the final goal. Infants then learn this arbitrary relation as the informative content communicated by the action manifestation. In the hands-occupied context, there is nothing more to explain about (or learn from) the demonstration because the model (in accordance with the rationality principle) performs the most efficient action available to her to produce the subgoal (to achieve the lamp being contacted). Therefore, infants who observed the head touch in the hands-occupied condition and acquired the subgoal to achieve direct contact with the lamp will touch the lamp with their hands because this is the most efficient action available to them to bring about the subgoal. In the hands-free condition, however, the rationality principle would be violated by the demonstrated head touch action if its goal state were assumed to be the same subgoal (i.e., to make contact with the lamp) performed in order to light up the lamp (the final goal). This would leave the agent's choice of performing the suboptimal head touch action as the means action unaccounted for. Thus, the observed action is reanalyzed at a finer level by the introduction of an additional finer grained arbitrary means–end relation, namely, the normative further subgoal specifying that the subgoal of contacting the lamp should be achieved by the use of the head. This second arbitrary relation is then added to the inferred informative content of the communicative demonstration, that is, to the manifested relevant new knowledge to be acquired. Therefore, when infants have an opportunity to operate the artifact in the hands-free context condition, they retrieve not one but two subgoals that they acquired and reproduce *both* of them by emulation (through the use of the most appropriate motor actions available to them). Thus, insofar as both arbitrary means–end relations are stored in and reactivated from their memory, infants will use their hands to efficiently realize the first subgoal (i.e., making contact with the lamp) but will also reenact the head action to efficiently bring about the second, more fine-grained subgoal (i.e., to realize making contact with the lamp by using the head).

The current experiments tested predictions derived from this proposal. In Experiment 1, we investigated whether social–communicative signals are necessary to elicit the reenactment of a teleologically unjustified action. If the account described above is correct, observation of a noncommunicative action should not induce the presumption of relevance and the acquisition of arbitrary means actions. Experiment 2 tested whether information about the overall goal state (i.e., the function of the novel artifact) is necessary to be present in the demonstration, as our proposal hypothesized, and whether the function of the learning is primarily epistemic or social.

Experiment 1

Although various alternative theories have been advanced to account for the underlying mechanisms driving imitative learning, apart from the current approach, none of them proposes that the communicative context of action demonstration plays a qualitative role in inducing selective imitation of novel means in different action contexts. One alternative account proposes that infants' observation of an action with a salient effect makes infants interpret it as a goal-directed action, and infants are motivated to imitate goals (Bekkering, Wohlschläger, & Gattis, 2000; Carpenter, Call, & Tomasello, 2005). Another approach suggests that imitation occurs as a consequence of motor resonance of previously encoded action–effect associations that are activated by the observed actions (Paulus et al., 2011a; Paulus et al., 2011b). Yet another alternative proposes that action imitation is modulated by the differential amount of attention evoked by the variable saliency of the target action performed in different contexts (Beisert et al., 2012). None of these alternatives, however, predicts a differential pattern of action imitation as a function of whether the target action is communicatively demonstrated to infants or is simply observed from a third-person perspective without communication. In contrast, if reenactment depends on the interpretation of the observed action, which in turn is modulated by ostensive communicative signals (Southgate et al., 2009), selective imitation is expected only in communicative contexts. Experiment 1 contrasted these alternative hypotheses by repeating Gergely and colleagues' (2002) study both in a communicative context (as in the original version) and in a noncommunicative third-person observation situation.

Method

Design

We investigated the effect of two independent variables on infants' tendency to imitate. The first independent variable was whether the model's action was observed when presented communicatively to infants (Communicative context) or from a third-person observational perspective when it was intentionally performed in a noncommunicative context (Incidental Observation context). The second independent variable was the mode of presentation of the target action (Hands Free vs. Hands Occupied), as in Gergely and colleagues' (2002) study. These two factors were crossed in a factorial design, creating four groups of participants.

Participants

A total of 75 14-month-old infants were recruited through advertisements in local newspapers. Of these infants, 6 were excluded from the final sample because of technical error ($n = 2$), parental interference ($n = 1$), or fussiness ($n = 3$). The remaining 69 children were randomly assigned to one of the four conditions. Mean age, sex distribution, and number of infants in each condition are presented in Table 1.

Apparatus

The tool on which the target action was modeled for infants was a small, circular, translucent "magic lamp" mounted on a box (27 × 19 × 4.5 cm). The lamp reacted to a gentle push by lighting up and remained illuminated until it was released. The lamp was placed on a small table in between the model and infants during the modeling phase and was presented to infants at the same location during the test phase. The sessions were monitored and videotaped from behind a one-way mirror.

Procedure

The procedure was composed of a modeling phase and a test phase.

Modeling phase: Communicative context. The infant was seated on the parent's lap in front of the table with the magic lamp, covered with a cloth. The distance from the table was approximately 1 m, which prevented them from reaching the apparatus. The mother was instructed not to interact with the infant during the modeling phase. The experimenter sat down at the other side of the table, uncovered the magic lamp, looked at the infant, and called the child's name, making sure that the infant paid attention. Then she shuddered and told another experimenter who was present in the laboratory that she was cold and asked for a blanket. After the blanket was handed to her, she wrapped it around her shoulders. In the Hands Free condition, she left the blanket hanging on her shoulders. She then placed her visibly free hands on the table on either side of the magic lamp. In the Hands Occupied condition, she wrapped the blanket around her shoulders and held it tightly with both hands. In both conditions, the model then bent forward from the waist and lit up the lamp by touching it with her forehead. She

Table 1
Descriptive statistics, looking times to the model, and the proportion of imitators in each condition.

Experiment/condition	N	Mean age (weeks)	Sex distribution (male/female)	Looking time to the model (s) [SD]	Infants performing the head action (successful goal attainment)	
					N	Percent
Experiment 1						
Communicative						
Hands Free	17	60.94	7/10	32.0 (4.9)	11 (8)	64.7
Hands Occupied	17	61.02	8/9	36.3 (5.5)	4 (3)	23.5
Incidental Observation						
Hands Free	17	60.53	5/12	35.6 (4.9)	5 (3)	29.4
Hands Occupied	18	61.20	8/10	33.2 (4.5)	8 (5)	44.4
Experiment 2						
No Effect	14	61.41	8/8	31.0 (4.6)	1 (0)	7.1
Model Present	13	61.20	7/6	34.6 (5.4)	9 (5)	69.2
Model Absent	18	60.07	9/9	33.5 (5.5)	9 (6)	50.0

repeated this action three times, making eye contact with the infant in between these actions and getting the infant's attention if necessary.

Modeling phase: Incidental Observation context. In this context, the experimenter did not interact with the infant. The child and the parent were seated in a restricted playing area of the laboratory by Experimenter A, and the parent was instructed to play with the child. The restricted playing area was a blanket spread on the floor, and the parent was asked to keep the infant there. From this playing zone, the child could clearly see the table where the modeling would take place, which was approximately 2 m away. After a short delay, Experimenter B (the model) entered the room and went straight to the table without looking at or speaking to the infant. She uncovered the magic lamp and rang a buzzer to get the infant's attention. The mother was asked not to look at the model except when the infant was pointing to her. When the infant paid attention to the model, a confederate from behind a one-way mirror indicated to the model that she could perform the target action by switching on a hidden light-emitting diode (LED). Experimenter B then wrapped the blanket around her shoulders, leaving her hands free or occupied (depending on the condition), and then modeled the target action three times. The hidden LED alerted the model when the infant was watching her without needing to look at the child. If needed, the model operated the buzzer again to get the infant's attention. After performing the target action three times, the model covered the magic lamp and left the room without looking at the infant.

Test phase. The test phase followed the modeling phase by 10 min, which the child spent outside the laboratory. In the Communicative context, the model led the child and the parent to the apparatus, encouraged the infant to play with it, and stayed in the room. In the Incidental Observation context, Experimenter A (not the model) seated the child next to the apparatus, encouraged the infant to play with it, and then left the room.

Data analysis and scoring

The video records of the test phase were scored by two independent observers who were uninformed as to the conditions in which children participated. The dependent measure was whether infants attempted to perform the head-on-box action within a 20-s time window. An attempt was defined as either touching the lamp with the head or leaning forward in such a way that an infant's head approached the lamp within 10 cm or less (see Meltzoff, 1988). The two coders' evaluation of infants' performance was in 100% agreement.

Results

The overall time spent looking at the demonstration was measured to check whether infants attended to the demonstrator during the modeling phase equally in each condition. The mean looking times are presented in Table 1. An analysis of variance (ANOVA) found no significant difference among the four conditions in the time spent looking at the demonstrator, $F(3,65) = 2.476$, $p = .069$.

The number of infants imitating the target action is presented in Table 1. An overall analysis of the proportion of imitators was performed by a generalized linear model (i.e., a general logistic model for binary variables) along 2 (Presentation Condition: Hands Occupied vs. Hands Free) $\times$ 2 (Context: Communicative vs. Incidental Observation) factors. This analysis yielded no main effect of presentation condition (Wald $\chi^2 = 1.354$, $df = 1$, *ns*), and there was no effect of context either (Wald $\chi^2 = 4.564$, $df = 1$, *ns*). However, the Presentation Condition $\times$ Context interaction was significant (Wald $\chi^2 = 6.326$, $df = 1$, $p = .012$).

To explore this interaction, we performed separate Fisher exact tests in the two contexts. In the Communicative context, modeling the action in the Hands Free condition produced more imitation than in the Hands Occupied condition (Fisher's exact $p = .018$). No such effect was found in the Incidental Observation context (Fisher's exact $p = .489$).

In addition, we performed a follow-up analysis to check whether the reproduction of the novel head touch action occurred as a result of more attention to the demonstration. We compared the overall looking time at the model during the modeling phase between infants who were "imitators" and those who were "nonimitators" during the test phase. We found that there was no significant difference between the looking times of these groups (imitators: $M = 34.42$ s, $SD = 4.91$; nonimitators: $M = 34.12$ s, $SD = 5.60$), $t(66) = 0.219$, $p = .827$. A closer look at the sequence of actions performed by infants revealed that, without exception, all of them used their hands first to attain the final goal of lighting up the lamp. This was true for both the Communicative and Incidental Observation contexts and for both the Hands Free and Hands Occupied presentation conditions.

Because the variability of the particular manner of reenactment of the head touch actions reflects the degree of fidelity of imitative motor responses, and hence can be indicative of the underlying mechanism that mediates social learning, the specific forms of head touches were also coded in this sample. Interestingly, 10 of 28 imitators (36%) performed two or even three different forms of head actions during the testing session. Only 3 infants (11%) produced a faithful motor copy of touching the lamp with the forehead, although 2 of them also used other parts of their head to light up the lamp (contacting it with an ear or with the face). Looking at the first instances of head touches only, there were only 3 high-fidelity imitative responses (touches with the forehead), 10 touches with the mouth (4 of them followed by other forms of head touches), 4 touches with the nose, and 3 contacts with one of the cheeks, and in 8 cases only the head approached the lamp box within 10 cm (without contacting it). This variability of the head contact actions revealed that the reenactment of head touches showed low fidelity when compared with the model's demonstrated forehead action.

We also tested the empirical validity of the assumption that the head touch action is a suboptimal means for infants to bring about the final goal. We counted the number of head touches that were successful in lighting up the lamp (see Table 1); overall, 67,8% of infants who reenacted the target act managed to attain the goal by this means, whereas goal attainment through the hand actions was always successful (100%). This clearly indicates that the novel head touch was a more awkward and less efficient means action for infants to perform than to touch the lamp by the familiar hand action.

Discussion

Our results in the Communicative context replicated the original findings of Gergely and colleagues (2002) and, thus, strengthen the interpretation that infants can adapt their reenactment behavior to the justifiability of the goal-directed actions performed by a model (see also Buttelmann et al., 2008; Király, 2009a; Schwier et al., 2006; Zmyj et al., 2009). Importantly, infants in the Hands Occupied condition did not reenact the observed means action. Rather, they achieved the same final goal by the most efficient action available to them, that is, their hands.

Our findings also show that selective imitation of the head action, and thus the reenactment of a cognitively opaque and arbitrary means, appeared only in the Communicative context. Infants in the Incidental Observation context did not perform more head touches in the Hands Free condition than in the Hands Occupied condition. Our procedure ensured that they had equal visual access to the head touch actions as infants in the Communicative context, and our measure of their looking time at the demonstrator's actions confirmed that they attended to the target actions equally in both conditions. The results demonstrate that infants learned about the affordance properties of the novel object in both conditions (as shown by their equal success in illuminating the lamp). However, there is no evidence to suggest that in the Incidental Observation context infants also acquired the obviously arbitrary head touch means action that the model chose to operate the lamp.

Furthermore, infants' selective low-fidelity reenactment seems to reflect a hierarchically organized goal-emulative strategy. In fact, the results suggest that infants inferred that the relevant information that the model's communicative demonstration intended to manifest for them in the Hands Free condition consisted of two goals that are hierarchically organized. The final goal is to bring about an interesting effect (illuminate the box by exploiting its affordance properties), and the relevant subgoal is the means by which the effect should be brought about (contact the box with the head). Thus, the differential pattern of reenactment of the exact same motor behavior observed in the Communicative versus Incidental Observation context raises doubts about the resonance-based automatic motor copying account of imitative learning (Paulus et al., 2011b). These findings also seem hard to account for in terms of the assumed differential salience of the head action in the Hands Free versus Hands

Occupied conditions (Beisert et al., 2012) because the very same action contexts were tested in both the Communicative and Incidental Observation conditions.

Note also that even in the condition where the head action was reliably reproduced, infants' reenactments remained relatively low fidelity. The model demonstrated touching the lamp with her forehead, which infants reproduced by contacting the lamp with virtually any parts of their heads. This suggests that infants encoded the intended subgoal at a relatively abstract level (establish contact between the lamp and the head) and emulated that subgoal nonimitatively by producing a variety of alternative head actions (Csibra, 2007). This variability of low-fidelity renditions of the demonstrated head touch action is also hard to account for in terms of the direct matching-induced motor resonance approach to imitation (Paulus et al., 2011b).

Experiment 2

The results of Experiment 1 are consistent with the proposal that infants learn the head touch action as a means to achieve the final goal, but they do not constitute a proof of this assumption. Our proposal is that infants use the interpretive scheme of the teleological stance (Gergely & Csibra, 2003) to infer the related elements of an ongoing action sequence and organize them around the attainable goal state. To test this hypothesis, in Experiment 2 we violated the availability of well-formed goal state information in the modeling situation by presenting the head action so that it did not induce a perceivable effect. If the minimal interpretability of an overall goal serves as an anchor for learning (and potentially for generalization), and for hierarchically relating a subgoal to it, infants should not be able to interpret the behavior as an instrumental action without supporting information about its goal and no acquisition of the novel behavior is expected.

Another question concerns the motivation behind infants' acquisition of arbitrary means actions. In Experiment 1, we covaried the presence or absence of communicative signals during the modeling phase with the presence or absence of the experimenter during the test phase. Thus, it is possible that the differential tendency to reproduce the novel means action in the two contexts was attributable to the presence of the model during test rather than to her communication signals during demonstration. Indeed, it has been proposed that imitation serves a social function by letting infants express their affiliation to others (Nielsen & Blank, 2011; Over & Carpenter, 2012). If this is the case, infants may reenact the newly acquired means action not because they have learned a novel behavior with instrumental value but rather to demonstrate to the model that they have acquired it. We attempted to address this issue by varying the experimental setup of the test phase as follows. In one condition the model who had demonstrated the target behavior was present during testing, and in the other condition no model or communicative partner was present during testing, only the infant's parent. Crucially, the novel action was demonstrated in an ostensive manner in both conditions.

Method

Design

There were three experimental conditions: Model Present, Model Absent, and No Effect. In all three conditions, the head action was demonstrated to infants in the Hands Free version and in a communicative manner.

Participants

A total of 50 14-month-olds were recruited through advertisements in local newspapers. Of these infants, 5 were excluded from the final sample because of technical error ($n = 1$), parental interference ($n = 1$), or failure to come back for the test phase of the study ($n = 3$). The remaining 45 children were assigned to one of the three conditions. Mean age, sex distribution, and number of infants in each condition are presented in Table 1.

Apparatus

The same apparatus was used as in Experiment 1

Procedure

In the Model Present and Model Absent conditions, infants were brought to the laboratory twice, with a 1-week delay in between. The first session consisted of the modeling phase, and the second session (1 week later) was the test phase of the study. In the No Effect condition, a short (10-min) delay was used, as in Experiment 1.

Our rationale to introduce these different delay intervals was motivated by several considerations. First, we wanted to present empirical proof for the claim that the same effect of selective learning that has been demonstrated following a shorter (10-min) delay can be similarly induced after a long (1-week) delay as well. Second, a central purpose of Experiment 2 was to disentangle whether it is the epistemic function or the social-affiliative function of imitation that is the dominant determinant of infants' performance in this particular task. To compare the relative contribution of the epistemic motive (as tested by the Model Absent condition) versus the social-affiliative motive (and/or the effect on recall of the presence of the demonstrator as an additional mnemonic cue, as tested by the Model Present condition), we used the same delay interval in both conditions. Note that if imitation serves primarily the epistemic function of learning, using the longer (1-week) delay would not be expected to influence the pattern of action reenactment, whereas if one assumes the primacy of the social-affiliative function, infants' imitative performance would be more likely to be reduced after a long delay due to the possible decrease in salience of the memory of the experimenter with whom infants were interacting during the demonstration phase. For this reason, we chose to use the longer (1-week) delay to compare the Model Present and Model Absent conditions.

On the other hand, the No Effect condition was designed with the primary purpose to test whether the selective reenactment of the head action induced in the communicative demonstration condition of Experiment 1 was constrained by the interpretability of the head touch as a goal-directed instrumental action or whether it was determined solely by the facilitative influence of the ostensive demonstration context to induce imitative behavior. This required that we test for the relative degree of imitation of the two types of actions by using the same delay interval in the No Effect condition as the one we used for testing the imitation of the head action *with* an observable effect in Experiment 1, so we employed the same (10-min) delay in the No Effect condition of Experiment 2 as well.

Modeling phase: Model Present and Model Absent. This phase was identical to the modeling phase of the Hands Free condition in the Communicative context of Experiment 1.

Modeling phase: No Effect condition. The demonstration was identical to the Hands Free condition in the Communicative context of Experiment 1 with the exception that, when the model bent forward, she stopped short of touching the lamp by approximately 2 cm and did not light it up. This action was repeated three times and was accompanied by the usual communicative signals.

Test phase: Model Present condition. The infant again was seated on the parent's lap in front of the table with the uncovered magic lamp, but this time at a distance that allowed the child to reach it. The model who had demonstrated the head touch action 1 week earlier sat on the other side and encouraged the infant to play with the apparatus without giving explicit instructions.

Test phase: Model Absent condition. The infant had the opportunity to play with the props in the presence of his or her parent only. The mother was asked to refrain from giving any direct instruction with respect to the modeling phase.

Test phase: No Effect condition. The model who demonstrated the head touch action 10 min earlier sat on the other side of the table and encouraged the infant to play with the apparatus without giving explicit instruction.

Data analysis and scoring

The same procedure was used to analyze the recording as in Experiment 1. Only 2 infants' performance was evaluated differently by the coders, and these cases were resolved by repeated scoring until agreement was reached.

Results

The overall time spent looking at the demonstration by infants is presented in Table 1. An ANOVA revealed that there was no significant difference among the three conditions in this measure, $F(2,42) = 1.726$, $p = .190$.

Number and proportion of infants who performed the target action are presented in Table 1. We compared the performance in the three conditions with each other in a chi-square test and found that they were significantly different from each other, $\chi^2(2) = 11.396$, $df = 2$, $p = .003$. We also compared the performance in the three conditions pairwise; the frequency of target actions did not differ significantly between the Model Present and Model Absent conditions (Fisher's exact $p = .462$), whereas the frequency of target action was lower in the No Effect condition than in the Model Present and Model Absent conditions (Fisher's exact $ps = .001$ and .019, respectively).

We also compared children's performance in the No Effect condition to that in the Hands Free condition presented in the Communicative context (Experiment 1) to check the potential impact of difference in delay on imitative tendencies. Here again the frequency of target action was lower in the No Effect condition than in the Hands Free condition presented in the Communicative context (Fisher's exact $p = .002$).

To directly test whether changing the delay from 10 min to 1 week had any effect on imitative reenactment tendencies, we also compared children's performance in the Model Present condition of Experiment 2 with that in the Communicative context–Hands Free condition of Experiment 1. There was no significant difference between the reenactment of target behavior in the above conditions (Fisher's exact $p = 1.000$), suggesting that learning of the opaque head action was not influenced by the different length of delays in the two experiments.

Discussion

Our findings suggest that although ostensive communicative signals are important, by themselves they are not sufficient to trigger action imitation. When the communicatively demonstrated head action was observed without a consequent visible effect being brought about, infants ceased to reenact the model's head action. This is consistent with the proposal that infants attempt to organize their interpretation of the demonstrated behavior into a hierarchical teleological structure of an instrumental goal-directed action that is anchored by the artifact function. This result is in line with the empirical findings of Paulus and colleagues (2011a), who found that the modeled action must be followed by a salient action effect in order to be imitated. Paulus and colleagues argued that the activation of the infants' own motor response can be linked to the representation of the action effect only if the latter is salient, and when this happens an association is automatically established between the activated motor program and the object producing the salient effect. However, the results of Experiment 1 are incompatible with this account because they provide no evidence that such an automatic association has been formed in the Incidental Observation condition. We believe that the role of the action effect is more likely to be that of specifying the relevant information about the function of the artifact (the final goal that can be achieved with it), which can provide an anchor that the further elements of the communicatively demonstrated action can be related to as subgoals.

Our findings are in accordance with the results of Lyons and colleagues (2007, Experiment 2B), who found that 4-year-olds do not overimitate demonstrated actions in which the contact principle was violated. The authors interpreted this finding as supporting their hypothesis that core assumptions of naive physics need to be satisfied to bring into play an automatic causal interpretation of the behavior as an instrumental act for the sake of learning (Lyons et al., 2007).

Furthermore, we found that infants tend to use a novel means action as a subgoal to attain a final goal both when the model who had demonstrated the behavior to them was present and when she was absent during the reenactment phase. This finding lends support to the claim that although it is crucial for infants to receive the novel information to be learned from a communicative partner, the subsequent reenactment of the acquired means action can be elicited without social cueing. This supports the view that the acquisition of the novel action was driven by an epistemic motive; the equal likelihood of its reenactment even in the absence of a social partner reflects the fact that the action has been learned as a culturally relevant novel instrumental means that ought to be used to operate the novel artifact. This result also confirms that the absence of head touch reproduction in the Incidental Observation context of Experiment 1 was not due to the absence of the model during the test phase.

At the same time, finding no significant difference between the proportion of reenactment of the novel means act in the Model Present versus Model Absent testing conditions seems unexpected at first glance given the results of Király (2009a), who, using a different task, found that when tested in the presence of the model, infants were more likely to attempt to reenact a previously demonstrated, rather complex tool use procedure to attain a goal (93% attempted tool use) than when they were tested without the demonstrator's presence (65% attempted tool use). This discrepancy, however, is plausibly attributable to factors stemming from the significantly higher difficulty for infants to perform the complex tool use action demonstrated in Király's (2009a) study (practically none of the imitators succeeded in achieving the goal by their attempted reproduction of the novel tool use) than to perform the head touch action in the current paradigm (67.8% of imitators were successful in lighting up the lamp through reproducing the novel head touch action). In our view, the increased proportion of imitative attempts to reenact the novel tool use in the presence of the demonstrator can also be plausibly attributed to infants' epistemic motive to learn relevant and new information from the communicative demonstrator.

Recently, Nielsen and Blank (2011) reported a study in which children showed an increased likelihood of imitating previously observed opaque and arbitrary parts of an action sequence to achieve a goal when tested in their demonstrator's presence as compared with the presence of another model who had demonstrated to them a more efficient—and simpler—version of the goal approach that did not include the unnecessary target actions in question. The 4-year-old participants in that study produced irrelevant actions at a significantly *lower rate* when given the apparatus by the efficient adult demonstrator than when the apparatus was handed over to them by the adult who had demonstrated the irrelevant actions to them in the first place. We accept the authors' interpretation that this finding demonstrates that social affiliative motives can indeed influence and increase imitation of others, at least in 4-year-olds. However, it should be noted that the results also provide positive evidence for the influence of the epistemic motive underlying children's imitation of the demonstrated opaque means actions; after all, they did reenact the demonstrated irrelevant actions *even when* interacting with the efficient demonstrator, although they admittedly (and quite understandably) did so with lower frequency than when tested by the inefficient model.

These findings, together with the current results, suggest that although the presence of a social model can in some circumstances increase the rate of imitation, it is not a necessary condition for the learning and imitative reenactment of novel means actions from communicative demonstrations by others.

General discussion

We have offered a reinterpretation of the phenomenon known as rational imitation in the literature and tested predictions drawn from this interpretation in two experiments. Under our proposal, the mechanism behind this phenomenon is not imitation (in the sense of motor copying actions), nor does it reflect the application of the principle of rationality (in the sense of optimal instrumental action selection). Rather, our account explains this phenomenon as an interplay among (a) specific forms of human communication, (b) the learning of hierarchical teleological structures of artifact function and use, and (c) emulative manifestation of acquired knowledge. The results of the experiments presented in this article confirmed the predictions drawn from this account.

Ostensive communication has been proposed to induce expectation of the opportunity to acquire new and relevant knowledge from the source of communication in human children (Csibra & Gergely, 2009). We believe that many earlier studies that ostensibly investigated the mechanisms of imitation actually studied how children interpret communicative action demonstrations. So far, not many experiments have contrasted communicative versus noncommunicative demonstrations, but the ones that have done so found different patterns of imitation in these conditions (Brugger et al., 2007;

Nielsen, 2006; Southgate et al., 2009). Our results confirm these findings. However, unlike Nielsen (2006) and Brugger and colleagues (2007), who emphasized the role of the "social" nature of demonstrations in general, the proposal we defend points to the role of *child-directed communication* as the crucial factor influencing reenactment and learning of novel skills and means actions demonstrated. Because human social interaction is normally communicative in nature, these two kinds of account predict action reproduction in similar situations. However, the absence of reenactment in the absence of obvious artifact function and the undiminished tendency to reproduce a novel means action in the absence of a social partner in Experiment 2 are more consistent with the primarily epistemic than social function of observational action learning.

Our results also speak against proposals that (a) consider only the motor aspects of action demonstrations in explaining imitative behaviors (Paulus et al., 2011a; Paulus et al., 2011b) and (b) try to interpret the findings of the original head touch study in terms of varying attentional distractiveness of the different modes of presentation (Beisert et al., 2012) because both the motor components and the relative salience and distractiveness of the demonstrations were similar across contexts and conditions in Experiment 1.

We have proposed that when children observe an object-directed instrumental action in the context of ostensive communication, they interpret the action in a different manner. If the manifested new behavior is a goal-directed action resulting in a well-defined outcome state (effecting a change of state in the world), infants try to interpret it in terms of their teleological representational schema, but they "suspend" the rationality requirement that the action needs to be the most efficient means available in the situation (cf. Gergely & Jacob, 2012). They do so because they expect relevant information (Sperber & Wilson, 1986), and one way that the content of the communication could be relevant for them is if it reveals arbitrary means–end relations about the novel artifact that infants could not have discovered on their own (lacking relevant causal knowledge). Thus, when children observe, in a communicative context, a goal-directed action that cannot be justified by invoking the principle of rationality, they interpret it as an arbitrary subgoal to be fulfilled in the service of attaining the final goal. Indeed, when children have already acquired means–end knowledge about an artifact (which, therefore, is no longer novel), they do not learn a nonefficient means action performed on it even in an ostensive communicative context (Pinkham & Jaswal, 2011). Similarly, they do not learn such an action when it is not addressed to them (i.e., in an Incidental Observation context) because this action does not carry much expected relevance for them.

It is important to emphasize that communication does not make infants learn just any arbitrary action. As Experiment 2 demonstrated, children must be able to assign to the action a well-formed teleological interpretation (Gergely & Csibra, 2003) in order to acquire it. Although the action may be arbitrary, it is not an arbitrary goal in itself but rather an arbitrary *means* toward some final goal. If some observations (e.g., turning on a light by a head action when hands are available) do not fit in this model, the explanatory attempt fails. However, the ostensive communicative context induces a search for a finer grained action explanation that extends the hierarchical representation of the teleological schema by sanctioning the inclusion of a more specific subgoal.[2]

Finally, when infants are given the opportunity to handle the novel artifact, they retrieve this hierarchical representation and perform actions that reproduce the goals and subgoals stored in their memory. Such a process is essentially a kind of emulation driven by the motivation to achieve and learn (sub)goals rather than to reenact actions (cf. Csibra, 2007; on the notion of goal-directed imitation, see Bekkering et al., 2000; Carpenter et al., 2005). This is consistent with the claim that in other contexts, imitative action reproduction may serve social functions (Over & Carpenter, 2012), and it is also possible that the primacy of social-affiliative function of imitation emerges later in development (see Nielsen & Blank, 2011).

We are aware of the potential limitations of the fact that our theoretical proposals are based on studies with infants of a limited age range and from the detailed examination of the conditions inducing selective imitation, which, however, relies predominantly on using one specific task only—the

[2] For a recent logical formalization of the action interpretation mechanism that children employ, see Varga and van Lambalgen's (2011) model, which conceptualizes this system as a "closed world" reasoning process that involves both action interpretation and action planning.

head touch paradigm. Although this fact represents a challenge for the generalizability of our findings, there is increasing new evidence indicating that the phenomena we have uncovered can be successfully demonstrated and further investigated by using a variety of other tasks as well (see Brugger et al., 2007; Schwier et al., 2006; Southgate et al., 2009). As an example, Nielsen (2006) found that 12-month-olds copied a specific novel tool use when its relevance for goal attainment was made explicit to them by providing negative evidence about the applicability of the alternative prepotent action available in their motor repertoire. However, when no such additional evidence was made accessible to them and the tool use was demonstrated in isolation, they did not learn and imitate the novel means action. Such results underline the need for further research to explore and better understand the effects of demonstrating different types of target actions and their informative contribution in modulating infants' reenactment tendencies in different contexts. Furthermore, some recent imitation studies with 2- and 3-year-olds now clearly indicate that selective imitation is not an isolated and transient developmental phenomenon that is restricted only to an early phase of infant cognitive development and social learning (e.g., Király, 2009b; Williamson, Meltzoff, & Markman, 2008). Such results make us confident that future research using a larger variety of task domains as well as a wider range of age groups will lead us to a fuller understanding and appreciation of the central role that ostensive communication and demonstrative manifestations play in making the efficient cultural transmission and stabilization of relevant and shared cultural knowledge, even if cognitively opaque, possible in human social groups.

Indeed, our reinterpretation of action reproduction in communicative contexts explains a host of findings in the literature but raises a question about the ultimate purpose of this learning process. Why do infants and children learn nonefficient instrumental actions from communicative demonstrations when they could discover such actions themselves and may even find more efficient means to the same goals (Pinkham & Jaswal, 2011)? We think that the answer to this question lies in the inherent opacity of culturally accumulated means–end knowledge, which is often embodied in human artifacts (Gergely & Csibra, 2006). Such knowledge is difficult to acquire by individual learning or by observational social learning, but benevolent adults could facilitate such learning by communicatively demonstrating it to children.

Together with other findings, our results show that infants do expect to learn from child-directed communication. If they do so, they are rational in the evolutionary sense; they learn more efficiently than they would without relying on adults' communication. Nevertheless, as the current results attest, the mechanisms of such learning cannot be explained solely by reliance on the principle of rationality of instrumental actions alone (cf. Gergely & Jacob, 2012).

Acknowledgments

This research was partly funded by an Advanced Research Grant from the European Research Council (OSTREFCOM). The European Union and the European Social Fund provided financial support to the project under the Grant agreement (TÁMOP 4.2.1/B-09/1/KMR-2010-0003). The authors thank the infants and their parents for their participation as well as Kata Egyed, Kata Krekó, and Krisztina Kupán for their help in data collection.

References

Beisert, M., Zmyj, N., Liepelt, R., Jung, F., Prinz, W., & Daum, M. M. (2012). Rethinking "rational imitation" in 14-month-old infants: A perceptual distraction approach. *PLoS One, 7*, e32563.

Bekkering, H., Wohlschläger, A., & Gattis, M. (2000). Imitation of gestures in children is goal-directed. *Quarterly Journal of Experimental Psychology A, 53*, 153–164.

Biro, S., Csibra, G., & Gergely, G. (2007). The role of behavioural cues in understanding goal-directed actions in infancy. *Progress in Brain Research, 164*, 303–322.

Biro, S., Verschoor, S., & Coenen, L. (2011). Evidence for a unitary goal-concept in 12 month old infants. *Developmental Science, 14*, 1255–1260.

Brugger, A., Lariviere, L. A., Mumme, D. L., & Bushnell, E. W. (2007). Doing the right thing: Infants' selection of actions to imitate from observed event sequences. *Child Development, 78*, 806–824.

Buttelmann, D., Carpenter, M., Call, J., & Tomasello, M. (2008). Rational tool use and tool choice in human infants and great apes. *Child Development, 79*, 609–626.

Carpenter, M., Call, J., & Tomasello, M. (2005). Twelve- and 18-month-olds copy actions in terms of goals. *Developmental Science, 8*, F13–F20.
Csibra, G. (2007). Action mirroring and action interpretation: An alternative account. In P. Haggard, Y. Rosetti, & M. Kawato (Eds.), *Sensorimotor foundations of higher cognition: Attention and performance XXII* (pp. 435–459). Oxford, UK: Oxford University Press.
Csibra, G. (2010). Recognizing communicative intentions in infancy. *Mind & Language, 25*, 141–168.
Csibra, G., Bíró, S., Koós, O., & Gergely, G. (2003). One-year-old infants use teleological representations of actions productively. *Cognitive Science, 27*, 111–133.
Csibra, G., & Gergely, G. (2009). Natural pedagogy. *Trends in Cognitive Sciences, 13*, 148–153.
Csibra, G., & Gergely, G. (2011). Natural pedagogy as evolutionary adaptation. *Philosophical Transactions of the Royal Society B, 366*, 1149–1157.
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*, 237–267.
Gergely, G., Bekkering, H., & Király, I. (2002). Rational imitation in preverbal infants. *Nature, 415*, 755.
Gergely, G., & Csibra, G. (2003). Teleological reasoning in infancy: The naive theory of rational action. *Trends in Cognitive Sciences, 7*, 287–292.
Gergely, G., & Csibra, G. (2005). The social construction of the cultural mind: Imitative learning as a mechanism of human pedagogy. *Interaction Studies, 6*, 463–481.
Gergely, G., & Csibra, G. (2006). Sylvia's recipe: The role of imitation and pedagogy in the transmission of human culture. In N. J. Enfield & S. C. Levinson (Eds.), *Roots of human sociality: Culture, cognition, and human interaction* (pp. 229–255). Oxford, UK: Berg.
Gergely, G., Egyed, K., & Király, I. (2007). On pedagogy. *Developmental Science, 10*, 139–146.
Gergely, G., & Jacob, P. (2012). Reasoning about instrumental and communicative agency in human infancy. In: J.B. Benson, (Series Ed.) & F. Xu, & T. Kushnir (Vol. Eds.), Rational constructivism in cognitive development (pp. 59–94). San Diego: Elsevier.
Gergely, G., Nadasdy, Z., Csibra, G., & Biro, S. (1995). Taking the intentional stance at 12 months of age. *Cognition, 56*, 165–193.
Gredebäck, G., & Melinder, A. (2010). Infants' understanding of everyday social interactions: A dual process account. *Cognition, 114*, 197–206.
Gredebäck, G., & Melinder, A. (2011). Teleological reasoning in 4-month-old infants: Pupil dilation and contextual constraints. *PLoS One, 6*, e26487.
Hernik, M., & Southgate, V. (2012). Nine-month-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*, 714–722.
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*, 303–320.
Kenward, B. (2012). Over-imitating preschoolers believe unnecessary actions are normative and enforce their performance by a third party. *Journal of Experimental Child Psychology, 2*, 195–227.
Király, I. (2009a). The effect of the model's presence and of negative evidence on infants' selective imitation. *Journal of Experimental Child Psychology, 102*, 14–25.
Király, I. (2009b). Memories for events in infants: Goal relevant action coding. In T. Striano & V. Reid (Eds.), *Social cognition: Development, neuroscience, and autism* (pp. 113–128). Oxford, UK: Wiley–Blackwell.
Lyons, D. E., Damrosch, D. H., Lin, J. K., Macris, D. M., & Keil, F. C. (2011). The scope and limits of overimitation in the transmission of artefact culture. *Philosophical Transactions of the Royal Society B, 366*, 1158–1167.
Lyons, D., Young, A., & Keil, F. (2007). The hidden structure of overimitation. *Proceedings of the National Academy of Sciences of the United States of America, 104*, 19751–19756.
Meltzoff, A. N. (1988). Infant imitation and memory: Nine-month-olds in immediate and deferred tests. *Child Development, 59*, 217–225.
Newell, A. (1982). The knowledge level. *Artificial Intelligence, 18*, 87–127.
Nielsen, M. (2006). Copying actions and copying outcomes: Social learning through the second year. *Developmental Psychology, 42*, 555–565.
Nielsen, M., & Blank, C. (2011). Imitation in young children: When who gets copied is more important than what gets copied. *Developmental Psychology, 47*, 1050–1053.
Over, H., & Carpenter, M. (2012). Putting the social into social learning: Explaining both selectivity and fidelity in children's copying behavior. *Journal of Comparative Psychology, 126*, 182–192.
Paulus, M., Hunnius, S., Vissers, M., & Bekkering, H. (2011a). Bridging the gap between the other and me: The functional role of motor resonance and action effects in infants' imitation. *Developmental Science, 14*, 901–910.
Paulus, M., Hunnius, S., Vissers, M., & Bekkering, H. (2011b). Imitation in infancy: Rational or motor resonance? *Child Development, 82*, 1047–1057.
Pinkham, A. M., & Jaswal, V. K. (2011). Watch and learn? Infants privilege efficiency over pedagogy during imitative learning. *Infancy, 16*, 535–544.
Rakoczy, H., Warneken, F., & Tomasello, M. (2008). The sources of normativity: Young children's awareness of the normative structure of games. *Developmental Psychology, 44*, 875–881.
Schwier, C., van Maanen, C., Carpenter, M., & Tomasello, M. (2006). Rational imitation in 12-month-old infants. *Infancy, 10*, 303–311.
Sodian, B. T., Schoeppner, B., & Metz, U. (2004). Do infants apply the principle of rational action to human agents? *Infant Behavior and Development, 27*, 31–41.
Southgate, V., Chevallier, C., & Csibra, G. (2009). Sensitivity to communicative relevance tells young children what to imitate. *Developmental Science, 12*, 1013–1019.
Southgate, V., & Csibra, G. (2009). Inferring the outcome of an ongoing novel action at 13 months. *Developmental Psychology, 45*, 1794–1798.
Sperber, D., & Wilson, D. (1986). *Relevance: Communication and cognition*. Oxford, UK: Blackwell.

Tomasello, M. (1999). *The cultural origins of human cognition*. Cambridge, MA: Harvard University Press.
Varga, A., & van Lambalgen, M. (2011). Infants' flexible closed world reasoning and imitation-for-learning. In F. Guerin, J. Alexander, D. Kazakov, & G. Tsoulas (Eds.), *Computational models of cognitive development: Proceedings of AISB 2011* (pp. 13–23). London: Society for the Study of Artificial Intelligence and the Simulation of Behaviour.
Verschoor, S., & Biro, S. (2012). The primacy of means selection information over outcome selection information in infants' goal attribution. *Cognitive Science, 36*, 714–725.
Wagner, L., & Carey, S. (2005). 12-month-olds represent probable endings of motion events. *Infancy, 7*, 73–83.
Wang, Y., Newport, R., Hamilton, A. F., & de, C. (2010). Eye contact enhances mimicry of intransitive hand movements. *Biology Letters, 7*, 7–10.
Williamson, R. A., Meltzoff, A. N., & Markman, E. M. (2008). Prior experiences and perceived efficacy influence 3-year-olds' imitation. *Developmental Psychology, 44*, 275–285.
Woodward, A. L., & Sommerville, J. A. (2000). Twelve-month-old infants interpret action in context. *Psychological Science, 11*, 73–77.
Zmyj, N., Daum, M. M., & Aschersleben, G. (2009). The development of rational imitation in 9- and 12-month-old infants. *Infancy, 14*, 131–141.

Relevance or Resonance:

Inference based selective imitation in communicative context

Ildikó Király*, Katalin Egyed* and György Gergely**

* Department of Cognitive Psychology, Eötvös Loránd University

**Cognitive Development Centre, Central European University

Corresponding author: Ildikó Király

Department of Cognitive Psychology,

Eötvös Loránd University,

46 Izabella Street, Budapest,

Hungary, H-1064

kiralyi@caesar.elte.hu

Abstract

The Rational imitation approach introduced early imitative learning as a selective, non-automatic, and inference-guided process (Gergely et al., 2002). This view has been challenged by Paulus, Hunnius, Vissers & Bekkering (2011a, 2011b, 2013) who argue that imitative learning is non-inferential and mediated by motor resonance through automatic matching by the mirror neuron system.

Paulus et al. (2011a) reported some modified versions of the head touch study to demonstrate that the selective imitation effect could be accounted for in terms of whether the action could be mapped onto an already existing motor scheme that the infant could perform.

Here we defend the view of selective imitation within a novel framework that builds on natural pedagogy, and answer this challenge on empirical grounds. We present new studies that modifies further two of the Paulus et al.'s procedure in such a way that the motor resonance account and the natural pedagogy account generates contrary predictions. In Study 1, the model presented the head touch with her hands free *and* held upwards either in a communicative context or in a non-communicative setting for 14-month-old infants. Different proportion of infants imitated the very same action, as a function of the demonstration context. In Study 2, the model kept balls in her hands while she put her hands next to the table, but either she demonstrated that she needed to hold the balls (hands occupied with balls condition;) or she demonstrated that she could leave the balls in the plates next to the lamp (hands free with balls). The proportion of imitators differed significantly in the two conditions, despite the identicalityof motor behavior in the slightly different situations. In Study 3 the model presented both the head and the hand action to light up the lamp. Infants in this case refrained themselves from performing the head action, despite the fact that it was modelled for them with hand support.We argue that our results provide support for the inferential relevance and rationality-sensitive account of selective imitation.

Key words: selective imitation, communicative cues, rationality principle/ inference based approaches

Relevance or Resonance: inference based selective imitation in Communicative Context

Introduction

The idea that imitation plays a crucial role in cultural knowledge transmission has been proposed in a variety of disciplines and extensively investigated both in comparative and developmental psychology (Tomasello, 1999; Richerson & Boyd, 2005; see Gergely & Csibra, 2006, for a review). There is broad agreement about three general propositions: (i) Humans have evolved specialized cognitive mechanisms adapted for cultural learning (such as their capacity to learn and imitatively re-enact actions they observe others perform); (ii) In contrast, inter-generational transmission of cultural skills in non-human primate groups is mostly accomplished by outcome-focused emulative learning processes rather than faithful imitative copying of novel actions, and iii) human imitative capacities are likely to have been adapted to serve several distinct – epistemic (Gergely & Csibra, 2005) *vs.* social (Uzgiris, 1981; Nielsen, 2006; Over & Carpenter, 2012) functions during evolution.

In contrast, there is much controversy and on-going debate concerning the nature of the psychological and neuronal brain mechanisms that sub-serve imitative cultural learning processes in humans. There is disagreement between currently proposed alternative models as how to account for the acquisition of *new* behavioral skills by imitative re-enactment on the one hand, and for their flexible (but functionally adequately constrained) generalization and selective reproduction in appropriate novel contexts on the other.

There has been two major contrasting approaches to account for the mechanisms that sub-serve imitative learning of cultural skills in humans: motor resonance-based accounts assuming direct and automatic perceptual-motor action matching and priming processes *versus* relevance-based inferential accounts assuming context-sensitive interpretive processes that guide and constrain selective re-enactment of observed actions. In this paper we shall first briefly describe these theoretical alternatives, then we shall address the recent empirical challenge launched by proponents of motor resonance theory (Paulus et al., 2011a,2011b.) against the inference-based interpretive approach to imitative learning (Gergely, Bekkering, & Kiraly, 2002). Finally, we introduce our answer for this challenge on empirical grounds.

Relevance-based inferential models of selective imitation and cultural learning.

Alternatively, *inference-based interpretive models* of action understanding and imitative learning were suggested, as an example Tomasello and colleagues (Call, Carpenter, Tomasello, 2005) proposed to differentiate *'true' imitative learning* from 'blind' mimicry (i. e., resonance-based automatic motor copying) arguing that only the former can account for the infant's capacity for "reproducing the adult's actual behavioral strategies in their appropriate functional contexts, which implies an understanding of the intentional state underlying the behavior" (Tomasello et al., 1993, p. 497). In this view, it is the human infant's cognitive understanding of the other's mental intentions and reasons behind his or her "rational choice of behavioral strategy" that allows the infant to determine "which aspects of the behavior are relevant for reproduction" (Tomasello, 1996, p. 323).

An empirical challenge to resonance-based direct matching theories of imitative learning came from Gergely, Bekkering and Király's (2002) demonstration of *context-sensitive selective imitation* using a modified version of Meltzoff's original (1988) 'head touch' study. Gergely et al. ran two groups of 14-month-olds varying across groups the relevant situational constraints ('hands-free' vs. 'hands-occupied' context conditions) in which the demonstrator performed the very same 'head touch' target action to illuminate the touch-lamp.

Gergely et al., (2002) found that infants re-enacted the novel ‘head action’ selectively in the two conditions in spite of the fact that it was modeled to them equally saliently in both. In the ‘hands-free’ condition 69 % of the 14-month-olds re-enacted *imitatively* the novel head action (replicating Meltzoff’s, 1988, result). In contrast, in the ‘hands-occupied’ condition only 21 % reproduced the modeled head-touch action, while the majority of the infants only used their hands to light up the lamp, producing a – more efficient - emulative response to achieve the goal.

Gergely et al.’s (2002) study provided challenging evidence indicating that the imitative re-enactment of the novel head-touch action is due to a top-down, context-sensitive, and inferentially-driven response selection process rather than being the result of automatic priming, motor resonance and consequent imitative ‘copying’ of the modeled action.

The Gergely et al. (2002) selective imitation paradigm has since then been extended in a number of informative ways. Zmyj et al., (2009) using the same task with a modified ‘hands occupied’condition where the model’s hands were tied to the table replicated Gergely et al.’s selective imitation finding already *in 12-month-old infans*. Accordingly, Schwier et al. (2006) using a *different but analogous task* , the house with chimney task, confirmed the availability of context-sensitive selective imitation in 12-month-olds,. Király (2009) used contrastive action demonstration contexts to induce inference-based selective imitation of *a novel tool use* task, the ball-as a lifter action in 14-month-olds. In a *comparative study* Buttelmann et al., (2008) used three different types of novel goal-directed actions involving the use of new tools to test for context-sensitive selective imitation in 14-month-old human infants and in apes. For each types of actions they found evidence for context-sensitive selective imitation in human infants “who used a tool more often when a demonstrator freely chose to use it than when she had to use it” (due to situational constraints) while “the apes generally used the tool equally often in both conditions”.

These replications closely followed the logic of the Gergely et al. (2002) selective imitation paradigm. However, there is also converging evidence from other studies using different procedures that indicate selective, inference-based imitation in infants. Nielsen (2006) found that in a situation where the context of presenting a prior failed attempt to operate an apparatus simply by hand clarified the reason for the demonstrator's subsequent use of a new tool to achieve the desired effect, 12-month-olds followed the specific tool use actions of the model only when they were given prior evidence of the reason to do so – otherwise they focused on reproducing the outcome of the demonstrated actions by emulation. Brugger et al., (2007) showed that infants' choice as to what to selectively imitate depended largely on their knowledge of the causal relationship between means and outcomes. They reported that 14- to 16-month-old infants were more likely to imitate the first action of a two-step event sequence when it was physically necessary to generate the effect.

Rational imitation account reformulated

Natural pedagogy theory (Csibra & Gergely, 2009, Gergely, Egyed and Király, 2007) as a general frame for early cultural learning offers a novel insight to the abovementioned body of selective imitation studies. Based on its primary assumptions it was hypothesized that the ostensive communicative cuing context may be a necessary precondition to induce context-sensitive inferences in the infants directing them to identify the new and relevant aspects of the action demonstration to be learned and selectively re-enacted. This hypothesis was tested by replicating the Gergely et al. (2002) study in either a 2nd-person ostensive communicative demonstration context or a 3rd-person non-communicative observation context. It was found that 14-month-olds selectively re-enacted the head touch action only in the ostensive communicative demonstration conditions. In contrast, when simply observing in a 3rd-person non-communicative situation an adult intentionally performing the unusual head-touch action, infants did not imitate the novel head-action in either context conditions ('hands free' *vs.* 'hands occupied'), but emulated the goal instead by touching the light box with their hands only (Király, Csibra & Gergely, 2013).

Moreover, other studies also revealed confirming evidence concerning the role of sociality in inducing selective imitation. In the study of Nielsen (2006), the group of 18-month-olds were more likely to copy the actions of the model when she acted socially, than when she acted aloof during modeling. Furthermore, 24-month-olds copied the actions of the model irrespective of the model's behavior but they were more likely to produce the end result of these actions when the model acted socially. Brugger et al (2007) also found that 14- to 16-month-old infants were more likely to imitate the action when it was socially cued, irrespective of its causal efficacy in the situation.

According to this current interpretation (Gergely & Csibra, 2006, 2009; Király et al., , 2013,) in the 'hands free' condition infants learned and imitatively re-enacted the head-touch action because they interpreted the model's *deliberate and contrastive choice* to present them this unfamiliar (and apparently sub-efficient) manner of operating the touch-lamp as a communicatively intended 'pedagogical' manifestation to demonstrate that – despite of its apparent teleological opacity - the novel head-touch action represents *culturally relevant and shared* knowledge that the infant should acquire.

<u>Motor resonance-based theories of imitative learning</u>.

A group of influential theories share the basic assumption that imitation of observed actions is an automatic process sub-served by a direct perceptual-motor action matching mechanism. These models assume that through the activation of the direct matching mechanism any observed action that maps onto a corresponding motor action scheme of the observer's – already existing - motor repertoire will induce some amount of automatic activation of that motor scheme. The consequent 'motor resonance' of the observer's matching motor action scheme is the hypothesized source of the human tendency to automatically re-enact an imitative motor 'copy' of the observed action. The main theoretical approaches that share the above assumptions include Meltzoff's "just-like-me" theory of human imitation (Meltzoff, 1996), supposing that infants have an innate predisposition to "identify" with others perceived as "just-like-them", and they "have an inbuilt drive to 'act like' their conspecifics" (Meltzoff, 1996, p. 363, see also Meltzoff, 2002); the Mirror Neuron System theory of imitative learning (Rizzolatti & Craighero, 2004), claiming that perceptually induced direct activation of the observer's matching motor action scheme functions as an automatic motor 'simulation' of the other's observed behavior through which the observer is assumed to gain direct (non-inferential) access and understanding of the 'goal' of the other's action; and the Action-Effect Theory that arguing for a cognitive system with stimuli and responses that are represented in a commensurable format (Brass et al. 2000, Craighero et al. 2002, Wohlschlager & Bekkering 2002; see also Prinz 2002).

The relevance-based inferential models reviewed above were recently challenged on empirical grounds by proponents of the resonance-based direct matching theory of imitative learning (Paulus et al., 2011a.). Paulus and his colleagues made the cogent observation that in Gergely et al.'s (2002) selective imitation paradigm there are possibly significant differences between the two context conditions in the model's particular 'bodily posture' during the demonstrated 'head touch' action that may provide the basis of a lower-level (non-inferential) account of the selective imitation results. They noticed that in the 'hands free' condition the model's hands are resting on the table, while in the 'hands occupied' condition the demonstrator's arms are folded across her chest when performing the unusual head-touch action. This observation indeed reveals a potential confound in the Gergely et al. study. Paulus el al. argue that 14-month-olds are not yet able to bend forward from waist and perform the head-touch action unless they support their body weight by leaning on their hands that they place on the table. (The authors report that all of the 14-month-olds in their study who reproduced the head-action acted this way.) Therefore, this motor performance constraint may have rendered these young infants unable to produce an imitative motor copy of the model's action as it was demonstrated in the 'hands-occupied' condition, while in the 'hands-free' condition the head-action was performed *with* apparent hand support so the infants could (and did) perform the corresponding imitative head-action (using their hands to support their body). Paulus et al. argue that this perceptual difference between how the modelled head-actions were performed in the two context conditions may in itself be sufficient to explain Gergely et al.'selective imitation result without recourse to any of the hypothesized higher-level inferential processes.

This is so because according to motor resonance theory, as long as the observed action is similar enough - and can therefore be directly mapped onto - an already existing action scheme in the infants' motor repertoire , the consequent motor resonance of the infant's corresponding action schemes will result in automatic imitative re-enactment of the modeled action. However, that the bending forward without any hand support manner of performing the head-touch action will induce little motor resonance in the infants as this behavior is not part of their motor repertoire, so no imitative re-enactment is predicted for the 'hands occupied' condition.

To empirically contrast and test the competing alternative accounts, Paulus et al. (2011a) designed three new modified versions of the Gergely et al. (2002) head-touch study while they also (successfully) replicated its original two ('hands free' and 'hands occupied') context conditions. Two of their new studies (their '*hands up*' and '*button*' conditions) were designed to be functionally equivalent to the original 'hands free' condition of Gergely et al. in so far as they also included evidence that the model's hands were free (but not used) when she demonstrated the head-action.

In their commentary on the experiments of Paulus et al., 2011b, Buttellmann and Zmyj (2012) argue that in the 'button' condition one cannot know whether infants (1) perceive the hands as being free; and (2) understand the function of buttons. Yet, a closer look at Paulus et al.'s (2011a button condition, and all conditions in 2011b) studies, they claim, could reveal that alternative explanations that are in line with the rational-imitation-account hold for infants' performance in these new conditions. As we agree on the critical re-analysis of Buttellmann and Zmyj (2012), we concentrate on the hands-up and balls condition of Paulus et al, 2011a.

Paulus et al's 'hands up' condition differred from Gergely et al.'s 'hands free' condition only in that during the head touch action the model's 'free' hands were not placed on the table (to avoid the hand-support confound). Instead, the demonstrator held both of her (supposedly 'freely available') hands rigidly up in the air and continued to hold them in the same relative position as she was bending forward from waist to touch the lamp with her head (see Table. 1). The rigidly 'held-up-hands' were moving simultaneously and contingently together with the model's upper body and head as she was bending forward to touch the light box during the head action.

So according to the inference-based account – argue Paulus et al. – infants should equally infer form the perceptual evidence that is also available in these new conditions that the model freely and deliberately chose to demonstrate the sub-optimal head-action instead of performing the equally available (familiar and more efficient) hand action.

In contrast, no such prediction follows from the motor resonance account as the head-touch action was performed by the model in both conditions *without hand support*, therefore, the direct matching account predicts little motor resonance and so no consequent imitation of the head-touch action for either the 'hands up' or the 'button' conditions. The authors report significantly reduced levels of imitating the head touch action in both of these new conditions claiming support for the motor resonance theory.

Paulus et al.'s third new context manipulation, their '*hands on balls'* condition was designed to achieve the opposite effect: as in this case the head action was demonstrated *with the hands placed on the table* (well, more exactly, *on two balls that the model's hands grasped and placed on the table* where she continued to hold them with her grasping hands even while performing the 'head-action', see Table 2). This 'hand posture and body configuration' is perceptually similar enough – argue Paulus et al. - to the 'bend over *with manual body support'* kind of head-touch action that the infants are in fact able to perform. It is assumed, therefore, that the corresponding motor action schemes in the infant's repertoire are going to be activated by direct matching and the consequent motor resonance will lead to automatic imitation of the head-touch action (with the infants hands supporting her bending body). This prediction was confirmed by the results of Paulus et al.'s 'hands on balls' study.

Furthermore, the authors again argue that from the point of view of the inference-based approach the 'hands on balls' condition should be considered functionally equivalent to the 'hands occupied' condition of the Gergely et al. study. This prediction was, however, not confirmed as the majority of the infants did imitate the head action in the 'hands on balls' condition (as predicted by the motor resonance account).

We agree that in each of their critical conditions their modified demonstrations succeeded in removing the purported artefact they identified in Gergely et al.'s procedure (i. e., the presence of apparent body support provided by the hands). However, we believe that the very same task modifications by which Paulus et al.achieved this desired methodological outcome have inadvertantly induced an unintended side-effect as well: they obliterated the kind of informative cues that allowed infants in Gergely et al.'s 'hands free' condition to assign a functionally relevant interpretation to the action sequences observed. In other words, Paulus et al. have presented infants with *overly ambiguous action demonstrations* that lacked the necessary ostensive-communicative, contextual and temporal parsing cues that the infant could have relied on to identify the relevant action context for interpreting the demonstrated target action and to infer the new and relevant information that the demonstrator intended to manifest for them to learn. Therefore, we suggest that - in spite of Paulus et al.'s explicit intention to the contrary - none of their new tasks managed to provide a fair test for the relevance-based inferential theory of selective imitation.

We argue that in order to learn novel actions infants need to be able to identify the relevant target actions. Moreover, for this, infants need to rely on the active inferential guidance provided by the demonstrator's ostensive communicative gestures and manner of action manifestations (cf. Gergely, 2007; Gergely & Csibra, 2006, 2011; Király et al., 2013). In our studies we introduce modifications to Paulus et al.'s 'Hands Up' and 'Balls' conditions in order to guide infants - with the help of ostensive-communicative signals - to 'read' the relevant action components in the relevant context with the help of communicative signals.

<u>Study 1. Disambiguating Paulus et al.'s 'hands up' condition - The 'palms in air' study :</u>

The *'palms in air'* demonstration context we employed in our first experiment is a modified version of Paulus et al.'s *'hands up'* context condition. The 'palms in air' procedure was designed to be perfectly comparable with that of the 'hands up' condition in so far as in both cases the model's hands were clearly and visibly held up in air throughout the demonstration phase and, in particular, at the moment when the model made her choice to present the head touch means action to light up the lamp. According to Paulus et al. the continuous visibility of the hands in the air has provided direct perceptual evidence for the infant to infer that the model's hands were 'free' in their 'hands up' condition. In this regard, therefore, the 'palms in air' demonstration condition is clearly comparable to Paulus et al.'s 'hands up' condition. This ensured that the hands could not be seen as providing manual support for the model's body while she was bending forward to touch the lamp with her head (controlling for the hypothesized 'hand position' artefact in the Gergely et al. study equally in the two conditions).

The 'palms in air' demonstration context differed significantly, however, from Paulus et al.'s 'hands up' condition both in the kind of hand action presented and in the action context in which the target head action was performed. After having sat down in front of the table and the touch-lamp, the model in the 'palms in air' condition demonstrated *two different salient actions separately in a sequence.* She first presented a *hand action* extending her two hands towards the touch lamp on the table while turning her palms upwards in mid air (see Fig 2B and C). This hand action corresponds to the kind of semi-conventionalized ostensive referential manual gesture we often use for 'showing' or 'highlighting' a referent object or event for the other as relevant to attend to. (In every-day communication this demonstrative manual gesture is often accompanied by some verbal referential expression such as 'Here!' or 'Voilá!' in French: something we've adopted in our ostensive communicative demonstration condition, see below).

Then, after a slight pause, the model proceeded to perform a *second action with her head*: she bended forward from waist and lit up the lamp by touching it with her forehead. During the performance of the head touch action her hands remained stationary in their previous position (held with palms up in mid-air). So in contrast to Paulus et al.'s 'hands up' condition, the 'palms in air' demonstration context provided clear temporal and contextual segmentation cues to help infants interpret the model's hand gesture as a separate action demonstration that was independently performed from the subsequent head touch means action.

We presented two groups of 14-month-olds with this action sequence in two demonstration conditions: in a *2nd-person ostensive communicative context* and in a *3rd-person non-communicative observation context*. In the communicative condition, apart from providing temporal action segmentation cues, the demonstrator also addressed the infant through ostensive referential gestures and presented the action demonstrations in an ostensive manifestative – slightly exaggerated 'motionese'- manner. This provided infants with the necessary highlighting and temporal parsing cues to guide them to separately interpret the initial hand action demonstration as an ostensive referential manual gesture. We hypothesized that the presence of these ostensively provided informative cues – similarly to Gergely et al.'s original 'hands free' condition – will direct the infant to parse and interpret the hand gesture as forming part of the relevant action demonstration context, rather than being part of the demonstrated head touch target action itself (as we argued was the case in Paulus et al.'s 'hands up' condition).

This allowed us to test the central assumption of Paulus et al.'s motor resonance account directly: if the observed head action is performed without apparent manual hand support, infants will not be able to imitate it (as they lack the corresponding motor scheme to perform the bending forward head touch action without hand support). In contrast, the relevance-guided inferential account predicts for the 'palms in air' condition that guided by the ostensive signals and temporal parsing cues present, infants will be able to infer the new and relevant information manifested for them by the ostensive demonstration of the unusual head touch means action and will be able to learn and re-enact it[1].

In the 3rd-person non-communicative observation context of the 'palms in air' study, the demonstrator presented exactly the same action sequence to a different group of 14-month-olds without, however, accompanying it with any ostensive communicative gestures. We predict that without the presence of such ostensive communicative cues, there will be no imitative re-enactment of the demonstrated – but cognitively opaque and teleologically sub-efficient – head touch action. No such differential prediction can be derived from the direct matching motor resonance account as both the communicative and the non-communicative observation conditions present exactly the same action sequences to the infants.

Method

Participants

Twenty-six 14-month-old infants were recruited through advertisements in local newspapers. Two of them were excluded from the final sample because of parental interference (n = 1) or fussiness (n = 1). The remaining 24 children were assigned into one of the two experimental conditions (12-12).

Apparatus

The novel head touch means action was modeled to the infants using a small, circular, translucent 'magic touch-lamp', mounted on a box (sized 27 x 19 x 4,5 cm). When being

[1] Note that according to this account infants' re-enactment of the head-touch action may not – and need not – involve the production of a faithful motor 'copy' of the 'head touch without hand support' action of the model. Infants may have to emulate the demonstrated goal of contacting the lamp with the head in so far as they – unlike the model – may need to use their hands to support their body during the re-enacted head touch action.

gently pressed the touch-lamp lit up and remained lit as long as the contact was maintained. The lamp was placed on a small table in between the model and the infant during the modeling phase. The sessions were monitored and videotaped.

Procedure

Similarly to the Paulus et al. (2011) study, after the modeling phase there was only a short delay before administering the test phase.

Modeling phase. The infants were seated in their parents' lap in front of a table on which the touch-sensitive 'magic lamp' was placed being covered up with a cloth. There was about 80 cm distance between the subject and the table so the infants were not able to reach the apparatus. Mothers were instructed not to interact with their infant during the modeling phase. The experimenter sat down at the other side of the table facing the infant and uncovered the magic lamp. Then she performed the same sequence of actions in both of the two demonstration context conditions (the 2^{nd}-person ostensive communicative' and the '3^{rd}-person non-communicative observation context' conditions).

2^{nd}-person Ostensive Communicative Demonstration Context condition:

Here the demonstrator first addressed the baby using a variety of ostensive communicative signals: she looked and smiled at the infant, greeted him or her by calling the baby's name in motherese. Then she told the infant: 'Look! I'll show you something!... Voilá!', and while saying 'Voilá' she performed an ostensive referential manual gesture (the 'palms in air' action) to highlight and direct the infant's attention to the magic lamp on the table. This manual presentation gesture involved raising her hands in air extending them towards the magic lamp while turning her palms upwards. After a short pause, she proceeded to demonstrate the head touch target action: she bent forward from waist and lighted up the lamp by touching it with her forehead. While performing the head action her hands remained stationary in their previous position (being held in mid-air with palms up). Thus, the model's hands were *not* used to support the model's body as she bent forward to perform the head touch action (similarly to Paulus et al.'s 'hands up' condition). She repeated the demonstration of the head touch action three times, making eye contact with the infant in between these actions and calling the infant's attention if necessary.

3rd-person Non-communicative Observation Context condition:

In this condition the experimenter did not interact with the infant at all while performing exactly the same sequence of actions as in the communicative demonstration condition. The infant was already seated in his/her parent's lap in front of the table, when the experimenter entered the room from the other side. She went straight to the table and sat down without looking at the infant or saying anything to him or her. She uncovered the light-box and proceeded to perform the two actions in a sequence: she first raised and extended her hands in the direction of the touch lamp turning her palms upward, then she bent forward and lit the lamp by touching it with her forehead (while continuing to hold her hands in the air). After modeling the target action three times, the experimenter covered up the lamp and left the room.

Test phase. The test phase followed the modeling phase immediately. In the Communicative demonstration context condition, the model pushed the lamp across the table to the infant and said 'It is your turn now! You can try it!'. She encouraged the infant to play with the touch lamp and stayed in the room throughout the test phase. In the Non-communicative observation context, after the model left the room, a second experimenter came in and pushed the lamp to the infant encouraging him or her to play with it. Infants were given 60 seconds to play with and explore the lamp.

In this experiment, we co-varied the presence or absence of communicative signals in the modeling phase with the presence or absence of the original experimenter in the test phase. We are aware of this constraint, but as this confound was controlod for in a previous study (Király et al., 2013) , and was found that infants tended to use equally the novel means-action, the head-touch when the model who taught them the behavior was present and when she was absent during the re-enactment phase, we used this scenario for the sake of ecological validity (it would have been pragmatically strange for the infants to face the originally non-communicative experimenter, reappearing as a communicative partner instantly).

Data analysis and scoring.

The video records of the test phase were scored by two independent observers who were uninformed as to which of the conditions the participant belonged to. The dependent measure was whether the infant attempted to perform the head-on-box action within a 60 s time window. An attempt was defined as either touching the lamp with the head, or approaching the lamp with the head (e.g. leaning forward) within 10 cm or less (see Meltzoff, 1988). We also coded the direction of the approach of the target action: the potential ways of approaching the lamp was: leaning forward or lifting up the lamp. The two coders' evaluation of the participants' performance was in 97 % agreement (Kappa=0,94).

Results

The proportion of infants who performed the target action are presented in table 1.

Please insert Table 1. here.

We compared the performance in in the two conditions. The frequency of target action re-enactment was lower in the in the 'Non-communicative observation context' condition than it was in the 'Communicative demonstration context' (Fisher exact p = .05). Calculating the odd ratio confirmed (OR=5.431) that infants in the 'Communicative context' condition were more likely to re-enact the head action in comparison to the group of infants in the 'Non-communicative context' condition. Like in former studies (Gergely et al, 2002; Paulus et al. 2011a), at least one hand action preceded the head action in the 92 % of cases. The frequency of hand actions was 7.8 for one head touch within the first 60 seconds. Interestingly, the head touches appeared in different forms in contrast to the modeled behavior[2].: most importantly in 30 % of infants who performed the head action (3 infants in the Hands up in communicative context and 1 infant in the Hands up in Non.communicative context) lifted up the lamp to their heads instead of leaning forward to it. Moreover altogether in 30 % of cases there was no contact between the approaching head and lamp (2 infants lifted up the lamp to the head but did not contacted it, the other 2 infants bend forward but did not make contact with the lamp in the Hands up in communicative context condition). These results confirm that infants performed voluntarily chosen variations of the originally observed behavior, rather than re-enacting a matched, similar motor behavior.

Discussion

The aim of this study was to proide a contrastive test for Paulus et al.'s hypothesis based on the direct matching model of motor resonance theory and the reformulated inference based

[2] This chance observation could have occurred because the velcro we used here to mount the lamp on the box was somewhat weaker than the earlier used one, so infants could lift up the lamp if they wanted to.

account (Gergely and Csibra, 2006. Király et al. 2013). Motor resonance theory would predict that if the observed head action is performed by the model without apparent manual hand support, infants will not be able to imitate it irrespectively of the presence or absence of an ostensive communicative demonstration context. This prediction is contradicted by our finding that in the communicative demonstration condition 75% of the infants re-enacted at least once the novel head touch means action demonstrated to them (replicating Gergely et al.'s (2002)original result). Furthermore, the motor resonance theory is not compatible with the selective nature of the imitative re-enactment induced: i.e., for the fact that such a high proportion of imitative re-enactment was present only in the ostensive demonstration condition while there was sigificantly lower degree of head actions produced (33%) in the non-communicative observation condition. However, since in both conditions exactly the same actions (and effects) were presented to the infant, the amount of motor responance they can induce should be identical (as should be the amount of imitative copying) given the direct matching account of motor resonance theory.

The selective pattern of results in the different demonstration conditions provide further empirical evidence for the natural pedagogy account of learning (Gergely & Csibra, 2005, 2006), namely, that the presence of the demonstrator's ostensive and referential communicative signals addressing the infant, is a critical factor that is necessary to induce the imitative re-enactment of the novel – though apparently sub-efficient means action. This finding confirms similar results obtained by Király et al., (2013) who replicated the original Gergely et al. (2002) 'head touch' paradigm in a 2nd-person ostensive communicative context vs. a 3rd-person non-communicative observation condition. They found similarly high level of imitation of the 'head touch' action only in the 'hands free' condition and only in the ostensive communicative demonstration context.

A potential limitation of this study, however, could be that ostensive communicative contextonly induced more attention in infants, and that in itself could account for the difference in the amount of imitators in the two conditions. To rule this alternative explanation out, we introduced a control study, the ’head and hand’ condition. In this, the model acted just like in the original hands free condition (putting her hands to the sides of the light-box), with the only difference that she lighted up the lamp with one of her hands as well (either before or after the head action). This manipulation has no impact on the potential process of direct matching of actions - the hands serves as support for the body during performing the head action, so motor resonance theory should predict high level of imitation in this case. From an inference based theoretical angle, however, we can suppose that if an adult voluntarily demonstrate within the same ostensive context an easily performable, prepotent action and a new, strange action as well, infants could identify the already known, efficient mean to attain a goal, and they would tend to choose the one that is easier to perform for themselves.

Study 2: The Head AND Hand action control condition .

Our main suggestion is that when children observe, in a communicative context, a goal-directed action that cannot be justified by invoking the principle of rationality, they interpret it as an arbitrary sub-goal to be fulfilled in the service of attaining the final goal. However, in the present experiment we build on a further assumption that when children have already acquired means-end knowledge about an artifact (which is therefore no longer novel), they do not learn a non-efficient means action performed on it, even in an ostensive communicative context (Pinkham & Jaswal, 2011).

Method

Participants

Fourteen infants (14-month-olds) were tested in the Head and Hand Control Condition.

Apparatus

The apparatus used was identical to that of Study 1.

Procedure

Modeling phase. Here the model demonstrated *two* actions (in ostensive-communicative context), while she also demonstrated that her hands were free (lying on the table next to the box with the lamp). This condition thus was exactly the same as the 'hands-free' condition of Gergely et al (2002), with the only difference that here the model demonstrated the prepotent hand action as well, and both actions resulted in lighting up the magic lamp. The Head action was identical to the one employed in the previous experiments, while the Hand action consisted of touching the top of the lamp, and lighting it up, by hand. Both actions were modeled twice, in alternating order, with the first action counterbalanced across participants.

Test phase.The test phase in this condition was similar to the test phase of the communicative context condition of Study 1.

Results

Briefly, in the 'Head and Hand' condition only 8 % of infants (1 child) imitated the head touch (see Table 1). Since the aim of the present study was to control for a potential confound of Study 1, we compared the results of the present study to the results of the two conditions of Study 1. The frequency of target action re-enactment was lower in the 'Head and Hand' condition than it was in the Hands up in Communicative context' condition of Study 1 (Fisher exact p = .001 , OR=48,1), and the frequency of target actions did not differ significantly in the 'Hands up-in Non-communicative context' condition of Study 1 and in the 'Head and Hand' condition (Fisher exact p =.148).

Discussion

The results of the 'Head and Hand' condition helps us to rule out the possible objection that the differential degree of re-enactment of the new head-touch behavior in the two 'palms in air' conditions of Study 1 occured just because there was ostensive (and so attention grabbing) cues in one, while there was no such cues in the other. In the case of the 'Head and Hand' condition the head touch was presented within a communicative framework, (could receive much attention), yet infants did not tend to perform it afterwards, rather the proportion of imitators were closer to the proportion of imitators in the Palms in air - no communicative context condition. This result is speaking for the inference based assumption: infants get to know within the demonstration situation that the hand action is an established and efficient way to attain the goal, while they are also manifested that the head action is a novel alternative to attain the same goal. Moreover, infants were demonstrated that both of the presented means (hand action and head action as well) are equally *relevant* to attain the same goal as. Our interpretation is in accordance with the findings and interpretation of Pinkham and Jaswal (2011) - who proved that when children themselves could acquire means-end knowledge about an artifact previosuly to the demonstration of a novel means action, infants do not learn a non-efficient means action performed on it, even in an ostensive communicative context. As the hand action is a more efficient, manipulator, and it was identified as a means in the modelling situation, they prefered this action to light up the lamp and consequently there was no justification in the situation to learn the suboptimal novel means- the head touch.

On the other hand, despite the fact that both the hand action and the head action could entail motor resonance easily (in this condition the models hands served as support for the body during the head-touch), infants did not show a tendency to perform the head action, debating the low -level motor- copying assumption of motor resonance theory.

Study 3- Disambiguating the 'balls' condition

Paulus et al.'s (2011) '*balls'* condition was designed so that it should be considered functionally equivalent to the 'hands occupied' condition of the Gergely et al (2002) as hands *were occupied with balls*, while as in this case the head action was demonstrated *with the hands placed on the table*, this hand and body configuration was perceptually similar enough to the 'bend over *with manual body support'* sort of head-touch action that the infants are able to perform. Let us, however, have a closer look on what was *manifested* for the infants in this demonstration condition. Again, we would like to highlight the role of ostensive communicative and temporal parsing cues that could guide infants' inferences on what was shown for them in the situation.

In the 'balls' condition of Paulus et al. there were two softballs lying on the table next to the lamp,.During the demosntration, after taking a seat, the experimenter started to play with these softballs for approximately 8 seconds. The experimenter kept one softball in each hand and put her hands next to the lamp. Then, the procedure followed exactly the procedure of the 'hands free' condition with the only difference that the experimenter was holding the two softballs in her hands during the experiment.

In this condition also, we find ambiguous what is demonstrated: are the hands really occupied by holding the balls or are they in fact free, just resting on the balls? Remember that as part of the demonstration context, the model - just before presenting the head action - played with the balls, freely using her hands. It is possible so, that this demonstration in itself was enough to induce the inference, that the model could use her hands whenever she wanted, but she chose to use her head instead to light up the box.

Our attempt to disambiguate this condition was a simple modification of the demonstration condition that could help infants to parse and interpret the ongoing action: in the 'Hands free keeping balls' condition we followed exactly the procedure of the 'Balls' condition in the Paulus et al study, except that 1. the two balls were lying on two little plates next to the table, 2. and after the model put her hands next to the light-box keeping balls, she lifted up her hands without the balls for 2-3 seconds, the balls stayed in the plates, and could not roll away, then the model re-grasped them. After this short event, the model performed the head action with her hands resting on the table, keeping balls. In this context it was made explicit during the demonstration that the hands are free. In the Hands occupied with balls' condition there was no plates next to the light-box, and the model performed the very same action sequence like in 'Hands free keeping balls' condition. So, when she lifted up her hands for a little, the balls could roll away, so she had to reach and stop them by re-grasping. In this situation the unavailability of the hands were unambigously manifested. In both of the situations, the model provided support for her bending forward action with her hands (keeping balls) so, according to motor resonance theory there should be no difference in the amount of imitators in the two conditions. In contrast the different situational constraints presented in ostensive communicative context could highlight for infants that the model made a contrastive choice - approving the relevance of the new head action - only in the 'Hands free keeping balls'. If this assumption is true, the amount of imitators should differ in the two conditions.

Method

Participants

Thirty 14-month-old infants were recruited through advertisements in local newspapers. Three of them were excluded from the final sample because of fussiness (n = 1), technical error (n=1), and parental inference (n=1). Participants were randomly assigned to two experimental conditions, as a result finally 14 infants were tested in *hands free keeping balls condition* and 13 infants were tested in the *hands occupied with balls* condition.

Apparatus

The apparatus and setting was identical to the apparatus introduced in Study 1. Additionally two small coloured plastic balls (diameter 5 centimeters) were used. Furthermore, in the *hands free keeping balls condition* there was two small white plates (diameter 9 centimeters) next to each side of the box. The sessions were monitored and videotaped.

Procedure

In this study we employed a short delay between the modeling and test phases, like in Study 1, and our conditions were all presented in ostensive-referential context.

Modeling phase. The general procedure was identical to the procedure introduced in study 1. In the '*Hand occupied with balls'* condition, after taking a seat, the model started to play with two little balls, throwing them up for 8-10 seconds. Then the experimenter put the balls on the table next to each side of the box. In this condition, the model invisibly let the balls roll away and just after it she re-grasped them. The model rested her hands on the table keeping the balls in her hands during the demonstration. Then the procedure followed exactly the general demonstration of the head touch: the model bent forward from waist and lit up the lamp by touching it with her forehead.

In the '*Hands free - keeping balls' condition***,** as well as in the former condition, after taking a seat, the model started to play with two little balls, throwing them up for 8-10 seconds. Then the experimenter put the balls on the table next to each side of the box. However, in this condition there was two little plastic plates next to each side of the box, thus the experimenter placed the balls into these plates. After putting down the balls the model lifted up her hands for 2-3 seconds and grasped again the balls and kept them in the hands during the demonstration. The balls could move on the plates, but they could not roll away. Here again, the procedure followed exactly the general demonstration of the head touch: the model bent forward from waist and lit up the lamp by touching it with her forehead. Note that in both conditions the balls could roll during the short presentation of lifting up the hands, so both presentations contained equal amount of movements, and were equally salient perceptually.

*Test phase***.** The test phase followed the modeling phase immediately in both conditions. The model pushed the lamp across the table in front of the infants, and told 'It is your turn now! You can try it!' She encouraged the infant to play with it and stayed in the room. Infants were given 60 seconds to play with and explore the lamp.

Data analysis and scoring

The video records of the test phase were scored by two independent observers who were uninformed as to which of the conditions the participant belonged to. The dependent measure was whether the infant attempted to perform the head-on-box action within a 60 s time window (like in study 1). The two coders' evaluation of the participants' performance was in 92 % agreement (Kappa = 0.85).

Results

Number and proportion of infants who performed the target action are presented in Table 2.

Please insert Table 2. here.

We compared the performance in the two conditions to each other and it was revealed that the frequency of target action was lower in the ‘Hand occupied with balls’ condition than it was in the ‘Hands free keeping balls’ condition (Fisher exact p = .054).; Odd ratio (OR=5,177) examination revealed that the probability of performing a head touch is more likely in the ‘Hands free keeping balls’ than in the ‘Hand occupied with balls’ condition.

Hand actions preceded head action in 94 % of cases. The frequency of hand actions was almost 6 hand action for a head touch Moreover, like in study 1, the head touches did not follow the modeled head touch with high fidelity: intriguingly, in 50 % of the cases infants lifted up the lamp to their heads instead of leaning forward to it (4 infants in the Hands Free keeping balls condition and 2 infants in the Hands Occupied with balls performed the head action this way). Also in 58 % of cases (out of which 25 % - 3 infants were lifters) there was no contact between the approaching head and the lamp.

Discussion

With a simple modification to the 'balls' condition of Paulus et al., we made the situation unambiguous whether the hands are really occupied with another, parallel action (Hands occupied with balls' condition), or the hands are really free, continuously available ('Hands free keeping balls' condition). Given that infants were presented with different constraints, these situations are similar to the original structure of the 'hands-free' and 'hands-occupied' conditions of Gergely et al. (2002), but the motor behavior performed in the two contrastive conditions are identical. We found that in the 'Hands free keeping balls' condition infants tended to imitate the target new behavior, 65 % of infants re-enacted the head action, replicating the results of Paulus et al (and also the original results of 'hands-free' condition of Gergely et al., 2002). In contrast, in the 'Hands occupied with balls' condition only 23 % of infants imitated the target action, (replicating the results of the 'hands occupied' condition of Gergely et al., 2002). Since the motor behavior was identical in the two conditions, motor resonance account (or any low-level approaches based on perceptual dissimilarities) could not explain the different pattern of re-enactment, and thus these results bring to light the role of inferential processes beneath imitation.

General Discussion

The main objective of the present paper was to provide a fair empirical contrast for the motor resonance and the relevance based account for the phenomena of selecive imitation with the help of situations in which demonstration cues were available to help infants in interpreting the target behavior while behavioral, motor components were kept contstant.

Our studies revealed selective imitative patterns as a function of different demonstration contexts despite the fact that the modeled target behaviors and related components were very similar or identical in our conditions. These findings challenge the interpretations of motor resonance or automatic direct matching processes since these approaches cannot grasp the influence of contextual cues on imitative patterns in case of similar modeled actions. Indeed, these result speak for the supposition of inferences induced by contextual cues that guide imitation in infants.

Our supposed interpretation of inference based selective imitation is that infants do not automatically produce a matching motor program (with high fidelity) but they encode the goal of the situation *and* they retrieve a behavior that is effective in its attainment. In addition, it is proposed that natrural pedagogy modulates what is learnt in the situation (Gergely,& Csibra, 2005, 2009, Király, Csibra and Gergely, 2013): communication can enrich the encoding of the goal with signaling the particular means (or features of it) as a culturally relevant way to goal-attainement that is manifested as worth-to be learnt.

The findings , hence, promote the refinement of the previous theoretical position of rational imitation (in accordance with the objections of Paulus, 2012201). Paulus (2012), analyzing the theory of rational action (Gergely & Csibra, 2003) and the proposal that infants engage in rational imitation (Gergely et al., 2002), has argued that this theory is based on several assumptions, such as that infants are able to evaluate others' action capabilities, that they possess conceptual body representations, and that infants engage in counterfactual reasoning. Yet, Paulus suggests that there is no evidence that these abilities are in place in young infants. Furthermore, applying Hacker's (2010) considerations, he has criticized the introduction of a cognitive principle as an explanation of human behavior as being a theoretical fallacy.

Our refined account could answer the criticism on the original rational imitation account put forward by Paulus (2012). In our novel account it is underlined similarly to the second point of Paulus (2012) that the problem with the original assumption (Gergely et al, 2002) is that it attributed the operation of the exact *same inferential principle* behind *different (opposite*) behavioral patterns. Namely, in our view the main problem with the original proposal as it stands was that it essentially made the idea of appealing to the rationality principle un-falsifiable: When infants did not reproduce the demonstrated action (in the 'hands occupied' condition), it was treated as evidence of the application of the rationality principle, and when they did reproduce it (in the 'hands free' condition), it was also interpreted as evidence for the operation of the same inferential principle (cf. Gergely and Jacob, 2012). The refined model we argue for in the present paper elaborates the role of inferential processes beyond action analysis, and argues that ostensive communication enables infants to represent the goal structure (overall goal and subgoal of specific means) of novel actions even when the causal relations between means and end are cognitively opaque. It is proposed that the presumption of relevance guides the interpretation of the demonstration by searching for a communicative content that appears in some way relevant to the recipient. It is still suggested that the principle of efficiency is employed, to compute and disregard action elements of the observed demonstration that are justifiable by situation-specific physical constraints. So this refined theory of teleological explanation of selective imitation suggest that infants do not have to engage in counterfactual reasoning, (neither do they have to be able to evaluate others' action capabilities, and they do not have to possess conceptual body representations): we suggest that after the observation of a novel goal-directed action, (1) infants can identify the goal (otcome) of the action that they encode; (2) when they invited to re-enact, young children recall the encoded goal (or goals / otcomes), and (re-)enact to attain the same goal-state. Most importantly, they turn to encode the *novel* *means as a subgoal* when it is signaled as *novel* information in ostensive communicative context.

In our view, to provide an adequate explanatory model of the role of imitative re-enactment in human cultural learning, any viable theory must be able to account for two significant empirical properties of the way human infants acquire novel skills from observing them performed by others in their social environment. The first problem is how to account for the remarkable species-unique ability that makes even pre-verbal infants capable of *fast-learning, long-term retention and delayed (but functionally appropriate) re-enactment of novel means actions observed* even in cases when the new functional skill had been presented to them only on a single occasion and it's re-enactment takes place a week (or, even months) later (as demonstrated e.g. by Meltzoff's, 1988, 1995).

Second, it is crucial to account for (let alone predict) the adaptive ability of human infants (and more widely, of 'human cultural novices') to *flexibly but appropriately generalize and selectively reproduce* the newly acquired motor skill *across a variety of functionally relevant novel contexts.* The proposed inference-based account can provide solutions for the above two problems, since 1) infants exhibit in our study as well that they encoded a novel behavior after only several demontrations; further more 2) a recent study using a similar head- touch props provided evidence on functionally appropriate generalization of the 'head-touch' action across different person-contexts as well as to across featurally clearly different new token items belonging to the artifact kind (see Chen et al., (2012). These problems, however, represent challenges for motor resonance accounts.

In sum, our results argue for the claim that re-enactment in the observer is achieved not by direct matching but by emulative action reconstruction process. It is important to emphasize, that though we refer to 'imitation' (underlining the fact that infants tend to follow a new means-action), this term is used in a broad sense: a closer look at the concrete form of re-enacted target actions uncovered that the means were used in a flexible manner. Infants did not always bend forward and contacted the lamp with their forehead, rather they either lifted the lamp up or bend forward to approach the lamp with the head. Moreover, in lots of cases (30 % in Study 1 and 58 % in Study 2, respectively) infants did not bring about the outcome but they 'reinstated' the observed goal. Hence, the main findings of the presented studies supports the view that action understanding and goal inferences rather precede, than follow from, action mirroring processes (Csibra, 2007).

Acknowledgement:

This research was supported by an OTKA K 109352 research grant to Ildikó Király. The European Union and the European Social Fund have provided financial support to the project under the Grant agreement TÁMOP 4.2.1/B-09/1/KMR-2010-0003. The authors thank the infants and their parents for their participation; and Erna Halász, Eszter Somogyi and Leila Preisz for running the experiments.

References

Bekkering, H.,Wohlschlager, A. & Gattis, M. (2000). Imitation of gestures in children is goal-directed. *The Quarterly Journal f Experimental Psychology*, 53A (1), 153-164.

Buttelmann, D. & Zmyj, N. (2012). Evaluating the Empirical Evidence for the Two-Stage-Model of Infant Imitation. A Commentary on Paulus, Hunnius, Vissers and Bekkering (2011). *Frontiers in Psychology*, 3, 512. doi: 10.3389/fpsyg.2012.00512

Brugger, A., Lariviere, L.A., Mumme, D.L. & Bushnell, E.W. (2007). Doing the right thing: Infants' selection of actions to imitate from observed event sequences. *Child Development*, 78 (3), 806-824.

Carpenter, M., Akhtar, N., & Tomasello, M. (1998). Fourteen- to 18-month-old infants differentially imitate intentional and accidental actions. *Infant Behavior and Development, 21,* 315-330. Carpenter, M., Call, J., & Tomasello, M. (2005). Twelve- and 18-month-olds copy actions in terms of goals. *Developmental Science*, 8, F13-F20.

Call, J., Carpenter, M., Tomasello, M. (2005). Copying results and copying actions in the process of social learning: chimpanzees (Pan troglodytes) and human children (Homo sapiens). *Animal Cognition,* 8 (3), 151–163

Csibra, G.(2007). Action mirroring and action understanding: an alternative account. In: P. Haggard, Y. Rosetti, & M. Kawato (Eds.), Sensorimotor Foundations of Higher Cognition. Attention and Performance XXII. Oxford: Oxford University Press. pp435-458.

Csibra, G. & Gergely, G. (2006). Social learning and social cogniton: The case for pedagogy. In Y. Munakata & M. H. Johnson (Eds.), *Processes of Change in Brain and Cognitive Development. Attention and Performance XXI* (pp. 249-274). Oxford: Oxford University Press.

Csibra, G. & Gergely, G. (2009). Natural Pedagogy. *Trends in Cognitive Sciences* 13, 4, 148-153.

Gergely, G. and Csibra, G. (2003). Teleological reasoning in infancy: The naive theory of rational action. *Trends in Cognitive Sciences, 7* 287-292.

Gergely, G., & Csibra, G. (2006). Sylvia's recipe: The role of imitation and pedagogy in the transmission of human culture. In: N.J. Enfield & S.C. Levinson (Eds.), Roots of

Human Sociality: Culture, Cognition, and Human Interaction (pp. 229-255).Oxford: Berg Publishers.

Gergely, G., Bekkering, H., & Király, I. (2002). Rational imitation in preverbal infants. *Nature, Vol. 415*, p. 755.

Iacoboni, M., (2005). Neural mechanisms of imitation", *Current Opinion in Neurobiology, 15*:632-637 (2005)

Meltzoff, A. N. (1988). Infant imitation after a one week delay: Long term memory for Novel acts and multiple stimuli. *Developmental Psychology, 24*, 470-476.

Meltzoff, A. N. (1996). The human infant as imitative generalist: A 20-year progress report on infant imitation with implications for comparative psychology. (pp. 347-370) In: C. M. Heyes & B. G. Galef, (Eds). *Social learning in animals: The roots of culture*. NY: Academic Press.

Meltzoff, A. N. (2002). Imitation as a mechanism of social cognition: Origins of empathy, theory of mind, and the representation of action. (pp. 6-25), In U. Goshwami (Ed.) *Blackwell Handbook of Childhood Cognitive Development*, Oxford: Blackwell.

Meltzoff, A. N., and Moore, M. K. (1977). Imitation of facial and manual gestures by human neonates. *Science, 198*, 75-8.

Meltzoff, A. N., and Moore, M. K. (1989). Imitation in newborn infants: Exploring the range of gestures imitated and the underlying mechanisms. *Developmental Psychology, 25*, 954-62.

Meltzoff, A. N. and Moore, M. K. (1997). Explaining facial imitation: theoretical model. *Early Development and Parenting, 6*, 179-92.

Nielsen, M. (2006). Copying actions and copying outcomes: Social learning through the second year. *Developmental Psychology*, 42 (3): 555-565.

Paulus, M., Hunnius, S., Vissers, M., & Bekkering, H. (2011a). Imitation in infancy: Rational or motor resonance? *Child Development, 82,* 1047-1057.

Paulus, M., Hunnius, S., Vissers, M., & Bekkering, H. (2011b). Bridging the gap between the other and me: The functional role of motor resonance and action effects in infants' imitation. *Developmental Science, 14*, 901-910.

Paulus, M. (2012). Is it rational to assume that infants imitate rationally? A theoretical analysis and critique. *Human Development, 55,* 107-121.

Pinkham, A. M., & Jaswal, V. K. (2011). Watch and learn? Infants privilege efficiency over pedagogy during imitative learning. *Infancy, 16,* 535-544

Richerson, P. J., Boyd, R. (2005). *Not by Genes Alone: How Culture Transformed Human Evolution*, Chicago, IL: University of Chicago Press.

Rizolatti, *G., &* Craighero, *L.* (*2004*). The mirror-neuron system. *Annual Revue Neuroscience,* 27, 169-192.

Schwier, C., van Maanen, C., Carpenter, M., & Tomasello, M. (2006). Rational imitation in 12-month-old infants. *Infancy*, 10, 303-311.

Southgate, V., Chevallier, C., & Csibra, G. (2009). Sensitivity to communicative relevance tells young children what to imitate. *Developmental Science*, 12, 1013-1019.

Uzgiris, I. (1981). Two functions of imitation during infancy. *International Journal of Behavioral Development, 4,* 1-12.

Want, S. C., & Harris, P. L. (2002). How do children ape? Applying concepts from the study of non-human primates to the developmental study of 'imitation' in children. *Developmental Science, 5(1),* 1-13.

Tomasello, 1993

Tomasello, 1996

Tomasello, M. (1999). The Cultural Origins of Human Cognition. Cambridge, MA:

Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences, 28*, 675-735.

Wohlschläger, A., Bekkering, H., & Gattis, M. (2003). Action generation and action perception in imitation: An instantiation of the ideomotor principle. *Philosophical Transactions of the Royal Society of London: Biological Sciences*, 358, 501-515.

Zmyj, N., Daum, M. M., & Aschersleben, G. (2009). The development of rational imitation in 9- and 12-month-old infants. *Infancy*, 14, 131-141.

Table Captions.

Table 1. The proportion of imitators in each condition in Study 1

Table 2. The proportion of imitators in each condition in Study 2

Table 1

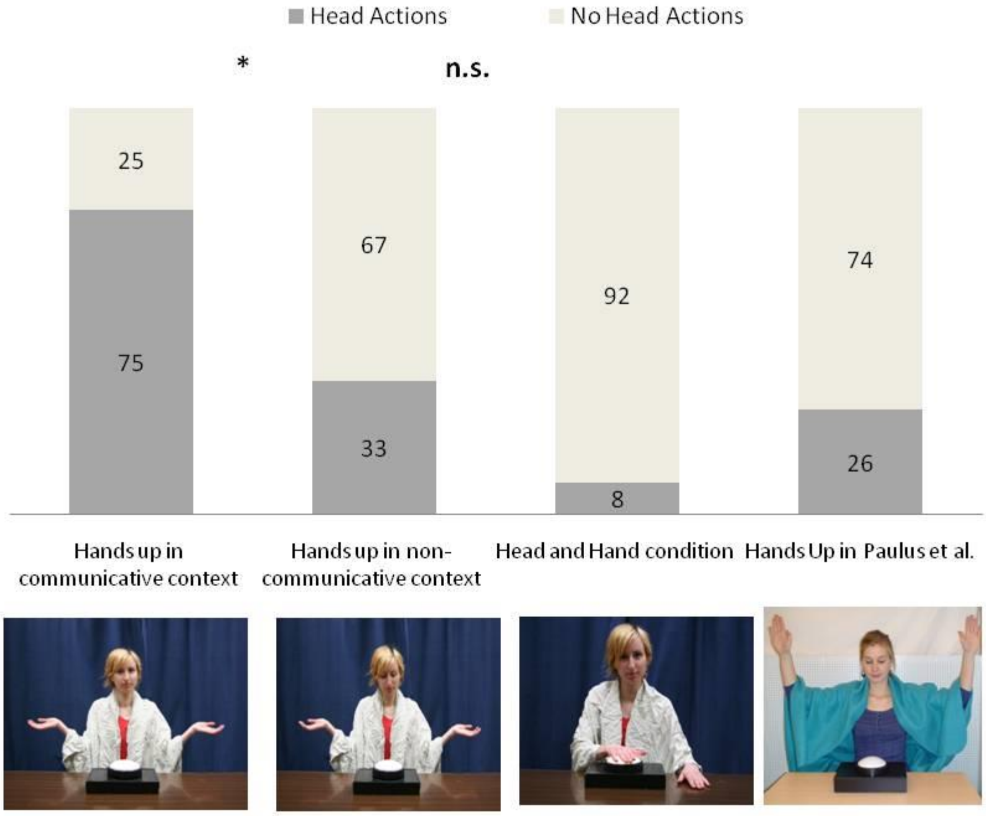
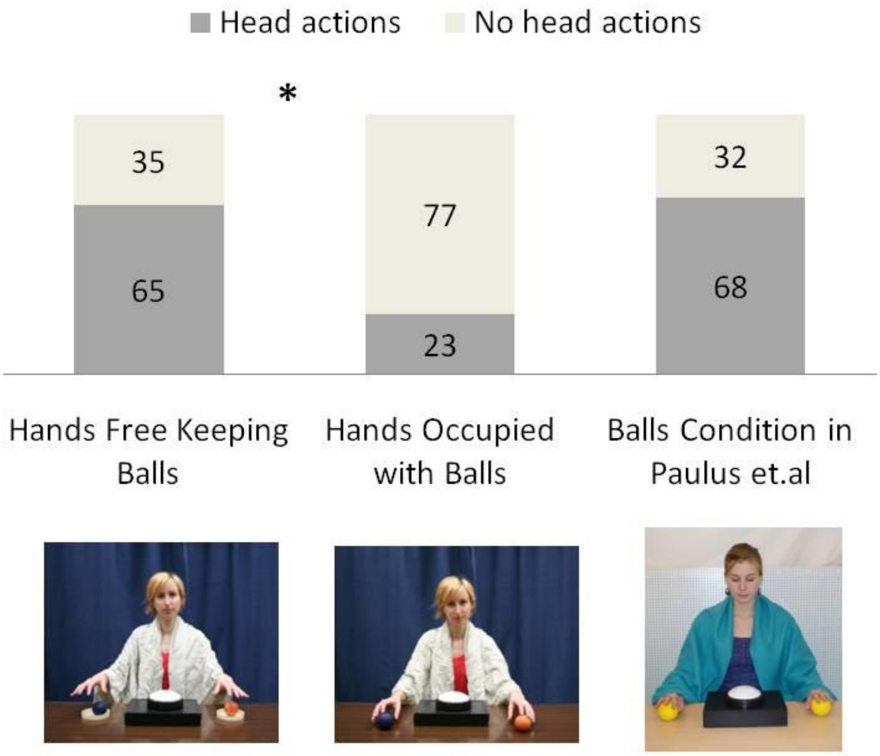

Head Actions
No Head Actions
*
n.s.
25
75
67
33
92
8
74
26
Hands up in communicative context
Hands up in non-communicative context
Head and Hand condition
Hands Up in Paulus et al.

Table 2


Head actions
No head actions
*
35
65
77
23
32
68
Hands Free Keeping Balls
Hands Occupied with Balls
Balls Condition in Paulus et.al

II. MEMORY

The main interpretative schemas of social learning not only help the interpretation of ongoing, real time behavior; additionally, they also shape the selection of valid information for longer periods of time.

2.1 Thesis 4. The interpretative schema for understanding goal directed actions plays a significant role in forming predictive, general memories

Király I. (2009): Memories for events in infants: goal relevant action coding. In: Striano, T., Reid, V. (eds.) *Social Cognition: Development, Neuroscience and Autism*, pp. 113-128. Wiley-Blackwell (ISBN 978 1 4051 6217 3)

2.2 Thesis 5. The inflexibility of early memory competences can be considered as the byproduct of being unable to encode the specifics of an event: in this sense episodic memory serves the function to reorganize and update knowledge based on mnemonic retrieval.

Kampis, D., **Király, I.,** Topál, J. (2014): Fidelity to cultural knowledge and the flexibility of memory in early childhood. IN: Pléh, Cs., Csibra, G., Richerson, P. (eds.) *Naturalistic Approaches to Culture*. pp.157-169. Budapest: Akadémiai

2.3 Thesis 6. The access to shared semantic (cultural) knowledge can be optimized through the memory organization of the individual. Consequently, different memory organization strategies are hypothesized that are dependent on the actual capacities and capacity constraint of the individual.

Király, I. Takács, Sz., Kaldy, Zs., Blaser, E. (2017): Preschoolers have better memory for text than adults, *Developmental Science* 20 (3). e12398; Doi:http://dx.doi.org/10.1111/desc.12398

Social Cognition
Development, Neuroscience, and Autism

Edited by
Tricia Striano
and Vincent Reid

WILEY-BLACKWELL
A John Wiley & Sons, Ltd., Publication

This edition first published 2009

Blackwell Publishing was acquired by John Wiley & Sons in February 2007. Blackwell's publishing program has been merged with Wiley's global Scientific, Technical, and Medical business to form Wiley-Blackwell.

Registered Office
John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, United Kingdom

Editorial Offices
350 Main Street, Maiden, MA 02148-5020, USA
9600 Garsington Road, Oxford, OX4 2DQ, UK
The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, UK

For details of our global editorial offices, for customer services, and for information about how to apply for permission to reuse the copyright material in this book please see our website at www.wiley.com/wiley-blackwell.

Library of Congress Cataloging-in-Publication Data

Social cognition : development, neuroscience, and autism / edited by Tricia Striano and Vincent Reid,
p. ; cm.
"This edited volume stems from questions and issues that were explored by contributors of the European Science Foundation Exploratory Workshop on Person Perception, held in Leipzig, Germany, in June, 2005. Developmental advances were also presented at the European Developmental Psychology Conference, Dual-Symposium on Person Perception During Infancy in Tenerife in August, 2OO5."-ECIP summary.
Includes bibliographical references and index.
ISBN 978-1-4051-6217-3 (hardcover : alk. paper) 1. Social perception in children. 2. Cognition in children. 3. Cognition in infants. 4. Autism-Pathophysiology. 5. Cognitive neuroscience. I. Striano, Tricia. II. Reid, Vincent.
[DNLM: 1. Cognition. 2. Adolescent. 3. Autistic Disorder—psychology. 4. Child. 5. Infant. WS 105.5.C7 S6775 2009]

BF723.S6S58 2009
155.4'13-dc22

2008009665

A catalogue record for this book is available from the British Library.

Set in 10.5 on 13 point Minion by SNP Best-set Typesetter Ltd., Hong Kong
Printed in Singapore by Fabulous Printers Pte Ltd

1 2009

Chapter 8

Memories for Events in Infants: Goal-Relevant Action Coding

Ildiko Király

Introduction: The Role of Goal Understanding in Memory Research

Research concerning adult event representation has revealed a remarkable body of evidence regarding the organizing role of goals. Investigations of different representational forms of events, such as narratives (Mandler & Johnson, 1977), observed action sequences (Lichtenstein & Brewer, 1980), and abstract knowledge of routine behaviours (Schank and Abelson, 1977), all share the basic assumption that adults appear to impose the interpretative framework of goal-directed action on the human behavior they encounter. The main knowledge structures that are based on goals support segmenting of the continuous flow of action sequences into actions with boundaries (like episodes in text reading - Black & Bower, 1979), identifying relevant knowledge (like scripts - Schank and Abelson, 1977), and establishing valuable hierarchical organization in memory (like plans - Lichtenstein & Brewer, 1980). Thus goal-based organization seems to play a primary role in adult event representations.

The problem of developmental continuity in the organization of event representations has initiated numerous studies investigating older children's event representations. As early as the age of 4, children display event representations that incorporate information about temporal order, causal relations, and goals: their verbal reports on earlier events are skeletal but include only such elements that adults would recount (Nelson & Gruendel, 1986). Thus, children's event memory resembles adult event memory with respect to goal-based organization, even if they use less complex forms of representations (Fivush, Kuebli, & Clubb, 1992; Nelson & Fivush, 2000). Nevertheless, much less evidence is available for younger children, such as non-verbal toddlers.

Support for early organized event representations comes from an experiment in which irrelevant elements were inserted into novel action sequences. The elicited imitation study of Bauer & Mandler (1989), revealed the assumption that causal-enabling relations between event components facilitate organization to entail better retrieval of action sequences. Infants (of 16 and 20 months) could retrieve novel events imitating their components while omitting or displacing their irrelevant components if there were causal relations between their event steps. Causal-enabling relations in an action determine the only meaningful temporal order of the event components that lead to the targeted outcome, and as such can support the memory organization of it. Events and actions in the world, however, do not always possess such inherent, enabling temporal organization; there are events with unbound temporal relations among their goal-relevant components. For instance, when someone would like to make a cup of cocoa, it is up to the actor's habit whether he or she puts milk or cocoa powder into the cup first, before mixing them to attain the very same tasty drink. According to Bauer and Mandler (1989), the improvement in retrieval of events containing enabling relations in comparison to events lacking such inherent structure was evidenced by superior ordered recall. This suggests that toddlers are sensitive to temporal irreversibility. Furthermore, causal structure as a source of information on temporal organization enhances memory tracing. An alternative explanation for the better "recall" of events containing enabling relations is that planning on the basis of goal-state configuration is enough for the reconstruction of these types of events, since goal information in itself can guide the threading of related event components.

In a further study, to test whether the improved recall performance of events containing enabling relations was due to problem solving rather then retrieval from memory, Bauer Schwade, Wewerka, and Delaney (1992) presented infants (of only 20 months) with the goal-state configuration both of enabling event sequences and of event sequences lacking such structure (the action sequences necessary to reach the goals were not demonstrated). The performance of infants after they had been encouraged to produce the entire event was poor; they rarely demonstrated the target sequences, and they were no more successful with enabling sequences than they were with arbitrary sequences. Somewhat in contradiction to this result, Bauer et al. (1999) showed that 20- and 27-month-old infants were able to use goal-state information to support their planning attempts in the case of novel enabling events, providing evidence on the assumption that goal-state configuration has a central organizational role, though in a problem-solving context.

A possible solution to the above puzzle is raised by a study by Carpenter, Call, and Tomasello (2002). The authors have shown that prior exposure to the end-state or outcome configuration of an action sequence, followed by full modelling of the target action sequence, results in superior performance in imitation of event components as compared to exposure just to the full modelling of the event. This result confirms that goal information plays a central role in the interpretation and encoding of events, and that it is a major factor in the organization of events for later retrieval, as prior information on the end-state, of the event facilitates the monitoring of event components.

In a task that required infants to represent relations between temporally separated actions and their converging structure, in which multiple actions served to enable a single outcome, Travis (1997) was able to show that 24-month-olds were capable of representing and imitating elements of an event in relation to its goal-based hierarchical structure. In particular, in the case of an event with embedded goal-irrelevant steps (in which two otherwise independent actions enabled a third action), infants grouped actions related to a common goal temporally, and reproduced goal-relevant action more than goal-irrelevant actions. The results clearly prove that 2-year-olds are able to represent converging causal structure, which is a characteristic of goal-directed action organization.

The importance of the early availability of goal-based organization can be apprehended in the fast parsing and identification of event types, and probably in forming general event representations. Despite the fact that there is empirical evidence that goal information plays a central role in the organization of event representations in the first years too, in the domain of memory development the notion of teleological stance is not widely appreciated. Teleological stance - an early interpretative schema for action coding proposed by Gergely and Csibra (1998) - is a convenient frame for guiding the perception and encoding of relevant, adequate ("real") components of events, even after only one brief exposure to them. The significance of this model lies in its power to clarify the central role of goal information in action representations through describing the inference structure and basic mechanism of interpretation as mediated and triggered by the rationality principle.

The aim of the present chapter is to introduce the consequences and experimental implementation of the model of teleological stance in the domain of imitative learning and memory development.

Early Interpretative Scheme for Action Understanding: The Teleological Stance

The theory of the teleological stance is based on the results of a series of habituation studies (Csibra, Biro, Koós, & Gergely, 2003; Csibra, Gergely, Bíró, Koós, & Brockbank 1999; Gergely, Nádasdy, Csibra, & Biro, 1995) that demonstrated goal attribution in 9- and 12-month-olds. These studies pointed out that by at least 9 months of age infants can (a) attribute goals to observed actions; (b) do so even if the agents are unfamiliar abstract entities that lack human features; (c) evaluate the relative efficiency of the goal approach in relation to the situational constraints on actions; and (d), if the relevant environmental constraints change, expect the agent to modify or change its means action adaptively to achieve efficient goal attainment in the new situation (Csibra et al., 1999, Csibra et al., 2003; Gergely et al, 1995).

Findings since 2000 confirmed that even 6-month-olds are able to interpret an ongoing action within the frame of the teleological stance: at this age infants are willing to attribute goals to humans and human-like robots (Kamewari, Kato, Kanda, Ishiguro, & Hiraki, 2005) and to any kind of inanimate object if it appears to be able to vary its goal approach (Csibra, forthcoming).

To account for these findings, Csibra and Gergely proposed that infants are equipped with an abstract and domain-specific action interpretation system, the teleological stance (Csibra & Gergely, 1998; Gergely & Csibra, 1998, 2003). Briefly, the teleological stance is a representational system that relates three kinds of elements in a specific type of (teleological) explanatory structure: (a) action: the observed behaviour, (b) goal: the consequent change of state in the world, and (c) situational constraints: the relevant aspects of the situation that constrain actions leading to the goal. An essential component of the teleological stance is the "principle of rational action". This principle is responsible for (a) driving inferences about goal-directed actions and, at the same time, for (b) providing criteria of well-formedness for teleological action interpretations. The importance of the rationality principle is rooted in the piece of evidence that it can guide the selection of goal-related (in contrast to unrelated), or goal-relevant (in contrast to goal-irrelevant) acts, as it can guide the online assessment of the ongoing action sequence. The mechanism of continuous evaluation by the rationality principle allows us to predict the outcome of an ongoing action just by assuming that it is a "direct way toward" an end-state or outcome or (in this case obviously) a goal.

Teleological Stance and Imitation: The Selective Interpretative Nature of Imitative Learning in Human Infants

In the domain of memory development, Meltzoff (1988) has demonstrated that infants are able to re-enact — that is, retrieve novel actions - after a one-week delay; in other words, infants are able very early on imitatively to learn novel means actions by way of observing others. In the most impressive task of the above mentioned study, 14-month-olds watched as a human model leaned forward from the waist and touched the top panel of a light box with her forehead, thereby illuminating it. A week later, 67 per cent of the infants re-enacted the novel "head action," while none performed it in a base-line control group for whom the action was not demonstrated. This result was an obvious indicator of long-term memory retention in infants for a specific event. Alternatively, this result seemed unexpected from the point of view of the 1-year-old's teleological stance (Csibra and Gergely, 1998), since, on the grounds of this model, one would have expected that in this task infants, as rational agents, should have performed the most efficient goal-directed action available to them (using their hand to contact the light box), instead of imitating the unique, but less efficient, "head action."

To clarify this situation, Gergely, Bekkering, and Király (2002) performed a modified version of Meltzoff' s task (1988). They hypothesized that "if infants noticed that the demonstrator declined to use her hands despite the fact that they were free, they may have inferred that the head action must offer some advantage in turning on the light. They therefore used the same action themselves in the same situation" (Gergely et al., 2002, p. 755). To test this idea, Gergely and colleagues tested two groups of 14-month-olds varying the situational constraints of the model. In the "hands-occupied" condition, the model's hands were visibly occupied: she pretended to be chilly and wrapped a blanket around her shoulders, holding it with both hands while performing the "head action." In the "hands-free" condition, however, after wrapping the blanket around her shoulders, the model placed her hands onto the table, so that they were visibly free, before demonstrating the "head action."

When the model's hands were occupied, 14-month-olds were less likely to imitate the "head action" (21 percent). Instead, they illuminated the box by touching it with their hand, performing the simpler, and equally effective, emulative response available to them, but not to the model. In contrast, when the model's hands were free, but she still used her head to illuminate the box, 69 per cent of 14-month-olds imitated her "head action" ($p < 0.02$) (this result is a replication of Meltzoff's results of 1988). So, differential imitation in the two conditions suggests that imitative learning is not an automatic "copying" process invoked by identification with the human actor, nor is it due to automatic behavioral "copying" of the modeled action. Rather, imitative learning is a selective interpretative process that involves the evaluation of the rationality of the means in relation to the situational constraints of the actor. Thus, re-enactment of the novel means takes place only if (a) the action is judged as rational given the situational constraints of the model, and (b) the action is judged as more rational than other available alternatives given the situational constraints of the infant him or herself.

With their recent model of human pedagogy, Gergely and Csibra (2005) shed new light on the constraints of teleological action interpretation. Their theory's main argument is

> that Mother Nature's "trick" to make fast and efficient learning of complex - and, for the learner, cognitively "opaque" - cultural knowledge possible was to have humans evolve specialized cognitive resources that form a dedicated interpersonal system of mutual design in which one is predisposed to "teach" and to "learn" new and relevant cultural information to (and from) conspecifics, (pp. 471-472)

A fundamental statement of their argument is that expert humans who possess cultural knowledge are disposed not only to use, but also ostensively to manifest, their knowledge to inexpert conspecifics, and inexpert conspecifics are specially receptive to ostensive communicative manifestations of others.

Briefly, the model of Human pedagogy outlines three major constituents that serve pedagogical knowledge transfer. First, there is a design specification that an expert conspecific (a "teacher") *ostensively communicates* her cultural knowledge by *manifesting* it for the novice (the "learner") with the help of *referential cues* (such as eye contact, turn-taking contingency). Secondly, because of her *special kind of receptivity,* the learner is predisposed to interpret the teacher's ostensive-communicative cues that accompany her knowledge manifestation as evidence that the manifestation will convey *new* and *relevant* cultural information for her. As a consequence, this allows fast learning of the communicated content without any further need to test its relevance independently. Thirdly, the built-in *presumption of relevance* of pedagogically communicated knowledge manifestations also enables the acquisition of knowledge contents that are *arbitrary, conventional,* and causally/functionally *non-transparent,* which stand for many forms of cultural knowledge.

The selective imitational findings of the Gergely et al. (2002) study is a nice example of how pedagogy operates: how infants infer differentially in two conditions what is new and relevant information for them (see also the argument of Gergely & Csibra, 2005). In the "hands-occupied" condition of the head-on-box study, the novel outcome, including the presented property of the object (illuminability-upon-contact), is the *new* information, so it is going to be retained in memory and reproduced through action. Taking the teleological stance in this case, infants can infer that, given the physical constraints of the actor (hands occupied), touching the box with her forehead is justified as a sensible and efficient means to the goal, as the physical-causal efficiency of the "head action" is cognitively "transparent" here.

In the "hands-free" condition the situation is different. The goal state involving the newly experienced affordance of the box is *new* information here, too, so it will be reproduced. In contrast, when setting up a teleological interpretation as to what particular action would constitute the most rational/efficient means to the goal under these situational constraints, given the fact that the actor's hands were free, the infant must have identified the available "hand action" as the most efficient means to perform. Unexpectedly, however, the demonstrator chose not to use her free hands, but performed the unusual "head action" instead. This contrastive choice marked the "head-action" as *new* and *relevant* information that the ostensive-communicative manifestation conveyed. As a result, both the new goal *and the new means* were retained and imitated.

The head-on-box study (Gergely et al., 2002) from this perspective confirms that pedagogical cues are necessary factors for imitative learning in human infants, although (1) there is interpretative selectivity guiding what aspect of the modeled behavior will be imitatively learned, and (2) this is directed by the implicit assumptions of the infant's "pedagogical stance" - namely, that the observed individual is about to manifest "for" them some significant aspect of cultural knowledge that will be *new* and *relevant.*

Relevance-Guided Selective Imitation: Verbal Labels Serving Human Pedagogy

Regarding the domain of memory development, it is a normal and common feature of imitational paradigms that the target event is presented in a rich ostensive context, comprising *communicative-referential speech acts* and overt verbal instructions (for example, "Look, I'll show you something!") before the target action is demonstrated. Such speech acts or short verbal instructions (labels) could enrich any situation in a Gricean sense, letting the observer perceive the actor's/speaker's intent of presenting something new and relevant. These assumptions are directly analogous, if not identical, to the Gricean pragmatic assumptions of ostensive communication, as made obvious by Sperber and Wilson (1986). From a slightly different perspective, however, pedagogy is a primary adaptation for cultural learning that is not necessarily conscious but a cue-dependent fast-learning attitude, and not a specialized module dedicated to the recovery of the speaker's intent in linguistic communication, an assumption that has evolved later as a sub-module of human theory of mind (Sperber and Wilson, 2002). Mentalistic terms are not necessary for conceiving the understanding of relevance. Speech acts are ready to convey the intention of the speaker. At the same time, however, these verbal acts appear as part of the external situation: they convey intents or goals through setting up in advance a possible end-state and thus highlighting a possible goal as part of the external situation. This alternative goal can either correspond to or mismatch with the outcome/goal of the ongoing action sequence. Thus verbal labels can enhance the understanding of the ongoing action in terms of its goal: if the verbal act is in line with a specific end-state achieved in the following action (consistent with its "intention-in-action"), it helps the encoding of the relevant steps of the event sequence; if the uttered goal or intent and the goal of the action sequence do not overlap completely, it can alter the encoding of what is *relevant* in the situation by setting up a goal hierarchy.

Verbal instructions, as part of the communicative referential context (the external situation), can serve as overt articulation *{manifesto)* of what to learn (what is relevant) in the situation. For instance, a closer look at our example "Look, I'll show you something!" reveals that this verbal act can imply that manifestation of *new* and *relevant* information in general is a *goal* of the agent. The main purpose of a further study was to examine this supposition. To do this, we manipulated and controlled the verbal labels used in the experimental situation. Our aim was to investigate the role of verbal labels in highlighting the relevant features of the ongoing actions that are "to be encoded." Our hypothesis was that verbal labels pragmatically referring to the manifestation of new and relevant information bring about imitative learning of cognitively opaque subevents, while verbal labels that accord with the presented event sequence and end-state (that refer to a script) result in the omission of cognitively opaque subevents, even though they are part of a communicative context. To test this specific hypothesis and to reinforce the assumption that imitation is not "blind." we introduced a novel-enabling and a novel-arbitrary event, both with embedded irrelevant components, following the method of Bauer (1992; see also, Bauer & Mandler, 1989; Király, 2003;Travis, 1997). By introducing irrelevant components, the authors aimed to test the relevance-guided selective mechanism involved in imitation, since the omission or re-enactment of goal-irrelevant components of the modeled actions can inform us about the underlying mechanisms of coding and organizing processes. We would like to confirm the primary role of goal information in the encoding of actions, as well as the leading role of the teleological stance in the organization and recall of event memories in the form of imitation.

Method

Participants

The final sample consisted of seventy-seven 29-month-old infants (M = 28.9, SD = 2. 67 months, range = 24 months to 34 months), who were visited in local day-care centers and playing centers. Four additional infants were excluded from the sample, because of maternal help/interference (two) and shyness (two) during the test.

Test materials

Each infant in the experimental conditions was exposed to either a novel-enabling (Figure 8.1) or a novel-arbitrary event sequence. In the baseline-control groups infants could see only the props themselves.

Figure 8.1. The event of "planting a flower"

Note: The presentation of the sequence was as follows: (1) taking a pot; (2) putting soil into it; (3) blowing on the flower (irrelevant step); (4) planting the flower.

Source: Courtesy of Ildikó Király (previously unpublished).

Both sequences were designed so that in each event sequence one of the event steps was irrelevant or unnecessary to the outcome of the event: one of their intermediary components was considered unimportant or irrelevant to the outcome of the event by adult raters.

Procedure

Infants were tested individually in the presence of their mothers in one of the kindergarten's playrooms. After a short warm-up period, they were seated on their mother's lap in front of a table, about 1 meter away, so that they could not reach the toy objects. The mothers were asked not to assist their children during the experimental sessions. The experimenter (an adult model) sat at the other side of the table. Elicited imitation procedure was used in which the experimenter modeled the target actions twice with the aid of the props, while commenting on these actions verbally in one of two ways, also making sure that the infant was paying attention. The sessions were video-recorded.

Two experimental conditions were introduced. In the "script or goal specification" condition, the presentation of the action sequence started with a reference to the event (by naming its purpose: "Look, I'm planting a flower" or "Look, I'm making a turtle"), while in the "pragmatically implied new information" condition the experimenter did not announce the aim of the event, instead referring to the pragmatic aim of the action itself as demonstrating new information ("Look, I'll show you something"). Her action sequence was the same in both conditions; thus she expressed the same "intention-in-action" too. In these experimental conditions the different types of events were presented to independent groups of infants. (There were 16 infants in the "pragmatic condition" with a novel-enabling event, 14 in the "script condition" with a novel-enabling event, 13 in the "pragmatic condition" with a novel-arbitrary event, and 10 in the "script condition" with a novel-arbitrary event.) During the test session the experimenter encouraged imitation with instructions such as "It is your turn now". Each of the infants was tested immediately after the modeling.

To assess the spontaneous production of target actions in the absence of adult demonstration, for each event a baseline-control group of 12 infants was exposed to the props.

Data analysis and scoring

The video-recordings of the test sessions were scored by two independent observers who were uninformed about which of the two conditions the children belonged to. The observers scored the presence or absence of each target action. Inter-observer reliability was calculated using kappa statistics (K = 0.912).

An imitation score was calculated for each infant by summing the number of target actions produced during the test (the range for each event type was 0-4). The observers also registered the order. Using this source of data, a quantitative and also a qualitative variable were introduced to assess ordering errors. For the quantitative variable the correctly ordered pairs of actions were summed. Here, the acceptance criterion was the replication of the modeled step sequence with or without the omission of the irrelevant step. For the qualitative variable (borrowing the method from Bauer and Mandler, 1989), the subjects' performance was classified into one of three categories: (a) exact reproduction of the sequence (with the irrelevant step), (b) displacement or omission of the irrelevant component, but preserving the rest of the modeled order; (c) other ordering "errors" - for example, displacement or omission of any other target action(s).

Results

The mean imitation scores of infants in the experimental and control conditions were as follows:

novel-enabling event

baseline control condition: Mean = 1.75 (SD = 0.62)
script condition: Mean = 2.78, (SD = 0.57)
pragmatic condition: Mean = 3.47, (SD = 0.5);

novel-arbitrary event

baseline control condition: Mean = 0.53 (SD = 0.79)
script condition: Mean = 3.1, (SD = 1.3)
pragmatic condition: Mean = 3.5, (SD = 1.08).

An overall 2 (event type: novel enabling; novel arbitrary) x 3 (condition: pragmatic, script, baseline) analysis of variance was conducted to assess the possible quantitative differences in imitative performance in the different conditions for the novel-enabling compared to the novel-arbitrary sequence. First, there was no main effect for event type ($F(1,71) = 3.32$, n.s.); thus there was no significant difference between the performance on the novel-enabling compared to the novel-arbitrary sequence. However, there was a significant main effect for condition, $F(1,71) = 40.29$, $p < 0.001$, and there was also a significant event-type x condition interaction, $F(1, 71) = 8.07$, $p < 0.001$. Post-hoc analysis (Tukey test of significant differences at the 0.05 level) yielded that infants' performance was greater in each experimental (pragmatic and script) condition compared to the baseline-control condition. Thus, regarding both event types, infants in the experimental conditions produced more target actions than children in the baseline-control condition. The significant interaction described above can result from the variance in the difference between imitation scores in the experimental conditions, dependant on event type. To test this possibility, for each event type, imitation scores were subjected to a one-way analysis of variance.

On the novel-enabling event sequence, there was a significant effect of condition: $F(2,39) = 17,19$, $p < 0.001$. Post-hoc analysis (Tukey test of significant differences at the 0.05 level) revealed that experimental conditions differ significantly from each other. In the pragmatic condition infants produced more target actions than in the script condition. This analysis confirmed that the performance in both experimental conditions differs significantly from baseline.

In the case of the novel-arbitrary event sequence there was a quantitative difference in infants' performance between conditions: $F(2,32) = 24.16$, $p < 0.001$. Post-hoc analysis (Tukey test of significant differences at the 0.05 level) yielded that, besides the significant difference between each experimental condition and baseline, there was no quantitative difference in the imitation scores in the two experimental conditions.

The difference between the mean of imitation scores in the experimental conditions do not clarify, however, whether this result is caused by a variation in the imitated pattern of components or not. It is likely that components are not equally produced or left out during imitation in the two experimental conditions. To test this possibility, we used non-parametric analysis to compare the number of imitators in the different experimental conditions (pragmatic versus script) for each event component, separately for event types.

In the case of the novel-enabling event, there was a significant difference only for the third component, the irrelevant step - Fisher's exact test: $p < 0.01$. In the case of the novel-arbitrary event, there was a significant difference only for the second component, again the irrelevant step - Fisher's exact test: $p < 0.05$. Table 8.1 shows the percentage of imitators for each event component in the experimental conditions.

Infants imitated the irrelevant component to a different extent in the two experimental conditions for both event types: in the pragmatic condition the imitation of the irrelevant component was more frequent than in the script condition, and this pattern of performance was the same for the novel-enabling and for the novel-arbitrary event.

In order to evaluate the performance in temporal ordering of event components, first a 2 (experimental condition: pragmatic versus script) x 2 (event type: novel arbitrary versus novel enabling) ANOVA was performed with the score for correctly ordered pairs. The descriptive statistics for the mean number of pair of actions produced in the modeled order are presented in Table 8.2. This analysis yielded a significant main effect for condition, $F(3,49) = 13.18$, $p < 0.001$, as well as for event type, $F(3, 49) = 19.0$, $p < 0.001$.

These results reflect that infants imitated more components in target order in the case of novel-enabling as compared to novel-arbitrary events. It also seems that the performance of retrieving event components in correct order was superior for the pragmatic condition; however, this effect can appear to be due to the higher imitation scores (which results in higher ordering scores). With the overall qualitative categories for ordering errors, this problem can be eliminated. Therefore, infants' performance was classified by whether they (a) reproduced the sequence exactly as modeled, (b) omitted (or displaced) the irrelevant component, but preserved the rest of the modeled order, (c) made other ordering errors. The percentage of infants in each category is shown in Table 8.2. To test whether there is a different pattern of ordering errors with respect to event type, a non-parametric test was conducted with combined data from the two experimental conditions. This analysis

Table 8.1. The percentage of imitators for each event component in the experimental conditions

	Condition	
Event sequence	*Pragmatic condition*	*Script condition*
Novel enabling: "Planting a Flower		
1 Taking a pot	100	86
2 Putting soil in it	100	100
3 *Blowing the flower*	**50**	**0**
4 Planting the flower	100	100
Novel arbitrary: "Making a Turtle"		
1 Fixing a leg to the "body"	80	85
2 *Throwing a little cube into it*	**100**	**61**
3 Fixing another leg to it	90	85
4 Putting the back on the turtle	80	85

Table 8.2. The mean number of pairs of actions produced in the modeled order and the percentage of infants in each category of ordering errors

	Condition	
Event sequence	*Pragmatic condition*	*Script condition*
Novel enabling: "Planting a Flower"		
Mean	2.45	1.71
SD	0.51	0.73
Correct ordering (%)	44	0
Omission (or displacement) of irrelevant component (%)	50	86
Other ordering errors (%)	6	14
Novel arbitrary: "Making a Turtle"		
Mean	1.54	0.61
SD	1.03	1.43
Correct ordering (%)	10	7.5
Omission (or displacement) of irrelevant component (%)	20	7.5
Other ordering errors (%)	70	85

yielded a significant difference between the pattern of ordering errors on the novel-enabling compared to the novel-arbitrary event, $X^2 = 25.58$ (df = 2), $p < 0.001$. A separate analysis for each event type revealed that ordering errors were more likely in the script conditions; however, in the case of the enabling sequence $X^2 = 8{,}04$, df = 2, $p < 0.05$), as is clearly visible in Table 8.2, infants in the pragmatic condition were more likely to produce the components in the correct order with the irrelevant component, while, in the case of the novel-arbitrary sequence ($X^2 = 0{,}85$, df = 2, n.s.), there was no significant difference in the pattern of ordering errors between the two experimental conditions.

Discussion

Since one of the main purposes of the study was to emphasize the role of teleological stance in the encoding and organization of actions, two different event sequences were used, a novel-enabling and a novel-arbitrary one. Our studies reinforced that children of 28 months re-enacted the elements of novel-enabling and novel-arbitrary events equally well.

Children performed better in the recall of the novel-enabling event sequence as compared to the novel-arbitrary one in only one respect: the recollection of its original temporal order, which can be interpreted as a consequence of the difference in the teleological organization of the event. In the case of enabling events there is only one specific, optimal way of attaining the goal, and this coincides with the original order of presentation in the modeling situation. Meanwhile, in the case of arbitrary events, there are several possible arrangements that are equally efficient in arriving at the same outcome, one of which is presented in the modeling situation. With regard to "teleology," this means that in the case of novel-enabling events the presented order of the subevents is the most effective way for attaining the outcome, while in the case of novel-arbitrary events the presented order of the event steps is only one of several equally efficient possibilities. Thus, an ordering error in the case of an enabling event impairs its efficiency in goal attainment, while variations in temporal order in the case of an arbitrary event bring about the same end result. The differential degrees of re-enactment in the two conditions indicate that goal information is actively used by 28-month-olds to evaluate the causal relevance of event components and as a consequence confirmed our claim that imitation is not a "blind" process.

Our studies demonstrate the continued centrality of the teleological stance in encoding action sequences in terms of goals and suggest that goal information is a primary organizing factor in the formation of event representations. From the perspective of investigating the memory performance of infants, our study reinforced (1) that infants encode the components of novel enabling as well as novel arbitrary events, though (2) they do not recall (code) the ordering information in the case of arbitrary events.

An important claim of ours was that verbal labels (we systematically varied them in our conditions) guide what to encode as goal-relevant information in the situation. Even though the demonstrated events were identical, infants in the "pragmatically implied new information" condition produced more target actions than in the "script or goal specification" condition for both event types. Moreover, in the pragmatic condition the imitation of the irrelevant component was more frequent than in the script condition, and this pattern of performance was the same for the novel-enabling and the novel-arbitrary event. With the help of a speech act that referred to a forthcoming manifestation of something *new* and *relevant,* infants selectively imitated an irrelevant, thus new and cognitively opaque, subevent. Within the framework of human pedagogy, we can assume that the model's ostensive communicative cues led infants to attend to the modeling, ready to apply their explanatory schemes to infer which aspects of the manifested behavior convey *new* and *relevant* information. In the "script or goal specification condition" the verbal label underlines the goal of the ongoing action, by reinforcing its specific end-state. Applying the teleological stance towards actions, infants can infer that the utterance of the actor calls their attention to the most efficient, optimal attainment of the expressed goal. Accordingly, they encode (and recall in imitation) only the adequate event steps leading to the specific outcome - the most efficient way of achieving the goal - as new and relevant information. At the same time, they omit the irrelevant step that is incomprehensible for them in the given situation.

In the "pragmatically implied new information" condition the verbal label refers implicitly to the manifestation of something new and relevant *in general.* Our data reinforced that both the goal-relevant components and the unexpected, new-event component (that was irrelevant with respect to the outcome of the situation) were imitated. Accordingly, our results reinforce our claim that, though pedagogical context highlights what is new and relevant, and is thus worth learning in a situation, at the same time teleological action interpretation guides the selection of what is to be imitated. So we would argue that pragmatic verbal instruction alters the analysis and set-up of the goal hierarchy of the event. With the general aim being that of learning something new and relevant, the overall goal of the situation can be an "interesting" attainment of the same outcome. Therefore, infants carefully monitor the whole event sequence. They definitely apprehend the intention-in-action of the event; they encode the adequate event components leading to a specific outcome. Simultaneously, they focus on the cognitively opaque subevent highlighted by the pragmatic verbal label. In this case, the teleological interpretation is launched on a more sophisticated overall goal: to fulfill the requirements of effective attainment of this, infants encode and later imitate the interesting cognitively opaque subevent of it as well.

Both of the verbal labels we used appeared in an ostensive communicative context: the verbal labels served to point to what is relevant and new in the situation. It is a question of whether ostensive-communicative cues are in themselves sufficient to trigger imitation or not; however, we suppose that the behavior should receive an at least partially completed interpretation, an understanding highlighted by a "mode of construal" available to them. One of the prosperous candidates here is teleology, as it is able to account for selective imitation: teleological action interpretation can in itself be a scheme for what aspects of the observed behavior should be copied and what aspects should be omitted. Meanwhile, verbal labels as external sources of information take part in setting up the goal hierarchy of the situation. In the "pragmatically implied new information" condition the strong prediction of the pedagogical model in itself would be that only the new, irrelevant component emerges in the imitative performance of infants. Disputing with this claim, we would argue that, in the case of our action sequences, the relation of the means actions and the goal (thus their "intention-in-action") is quite transparent, so that the observer can conceive the means-end sequence with the help of teleological action interpretation. Our results confirmed that infants can encode and later retrieve not just the new and relevant event step but also the adequate means sequence in the pragmatic condition. Implicit, pragmatic reference to a forthcoming new and interesting information leaves a set of possible goals open, prevalent enough to help the encoding of the cognitively opaque subevent that is irrelevant with respect to the end-state achieved in the action. In the case of the script, verbal labeling of the goal expressed by the verbal act and the intention-in-action are in accordance, which deepens the encoding of the goal of the ongoing action as the most relevant information in the situation. Using the teleological interpretation, infants omit the cognitively opaque subevent, and selectively imitate the relevant event steps of the action.

In sum, we can conclude that two major interacting factors guiding action interpretation and encoding in infants are the teleological principle and the overall aim of the demonstration situation as expressed in the verbal label communicated in the pedagogical context by the demonstrator. While the former serves as a selective frame that determines the automatic selection of goal-relevant event components, the latter may reinforce this selection or extend it with novel, goal-irrelevant elements. Our study revealed the importance of the teleological stance in encoding action sequences in terms of goals and reinforced that goal information is a primary organizing factor influencing the mnemonic performance of infants through guiding the formation of general event representations.

Note

The research outlined in this paper was supported by a grant from OTKA (T 047071) and by a Bolyai Research Grant to Ildikó Király.

References

Bauer, P. J. (1992). Holding it all together: How enabling relations facilitate young children's event recall. *Cognitive Development, 7,* 1-28.

Bauer, P. J., & Mandler, J. M. (1989). One thing follows another: Effects of temporal structure on 1- to 2-year-olds' recall of events. *Developmental Psychology,* 25, 197—206.

Bauer, P. J., Schwade, J. A., Wewerka, S. S., & Delaney, K. (1999). Planning ahead: Goal directed problem solving by 2-year-olds. *Developmental Psychology,* 35(5), 1320-1337.

Black, J. B., & Bower, G. H. (1979). Episodes as chunks in narrative memory. *Journal of Verbal Learning and Verbal Behaviour,* 18, 309-318.

Carpenter, M., Call, J., & Tomasello, M. (2002). Understanding "prior intentions" enables 2-year-olds to imitatively learn a complex task. *Child Development,* 73, 1431-1441.

Csibra, G. (forthcoming). Goal attribution to inanimate agents by 6.5-month-old infants. *Cognition.*

Csibra, G., Biró, S., Koós, O., & Gergely, G. (2003). One-year-old infants use teleological representations of actions productively. *Cognitive Science,* 27(1), 111-133.

Csibra, G., & Gergely, G. (1998). The teleological origins of mentalistic action explanations: A developmental hypothesis. *Developmental Science,* 1(2), 255-259.

Csibra, G., Gergely, G., Biro, S., Koós, O., & Brockbank, M. (1999). Goal attribution without agency cues: The perception of "pure reason" in infancy. *Cognition,* 72, 237-267.

Fivush, R., Kuebli, J., & Clubb, P. (1992). The structure of events and event representations: A developmental analysis. *Child Development,* 63, 188-201.

Gergely, G., Bekkering, H., & Király, I. (2002). Rational imitation in preverbal infants. *Nature,* 415, 755.

Gergely, G., & Csibra, G. (2003). Teleological reasoning about actions: The naive theory of rational action. *Trends in Cognitive Sciences, 7,* 287-292.

Gergely, G., & Csibra, G. (2005). The social construction of the cultural mind: Imitative learning as a mechanism of human pedagogy. *Interaction Studies,* 6, 463-481.

Gergely, G., Nádasdy, Z., Csibra, G., & Biro, S. (1995). Taking the intentional stance at 12 months of age. *Cognition,* 56(2), 165-193.

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, 303-320.

Király, I. (2003). Guidance by goals: The possible causes of irrelevant act omission in infants' imitative behaviour. Poster presented at Society for Research in Child Development. Tampa, USA.

Lichtenstein, E. H., & Brewer, W. F. (1980). Memory for goal-directed events. *Cognitive Psychology,* 12, 412-445.

Mandler, J., & lohnson, N. S. (1977). Remembrance of things parsed: Story structure and recall. *Cognitive Psychology,* 9, 111-151.

Meltzoff, A. N. (1988). Infant imitation after a one week delay: Long term memory for novel acts and multiple stimuli. *Developmental Psychology,* 24, 470-476.

Nelson, K., & Fivush, R. (2000). Socialization of memory. In E. Tulving & F. I. M Craik, (Eds.), *The Oxford Handbook of Memory* (pp. 283-295). Oxford: Oxford University Press.

Nelson, K., & Gruendel, I. (1986). Children's scripts. In K. Nelson (Ed.), *Event Knowledge: Structure and Function in Development* (pp. 47-69). Hillsdale, NJ: Lawrence Erlbaum.

Schank, R. C, & Abelson, R. (1977). *Scripts, Plans, Goals and Understanding.* Hillsdale, NJ: Lawrence Erlbaum.

Sperber, D., & Wilson, D. (1986). *Relevance: Communication and Cognition.* Oxford: Blackwell.

Sperber, D., & Wilson, D. (2002). Pragmatics, modularity and mind-reading. *Mind & Language,* 17(1), 3-23.

Travis, L. L. (1997). Gol-based organisation of event-memory in toddlers. In P. van den Broek, P. I. Bauer, & T. Bourg (Eds.), *Developmental Spans in Event Comprehension and Representation: Bridging Fictional and Actual Events* (pp. 111-138). Hillsdale, NJ: Lawrence Erlbaum.

FIDELITY TO CULTURAL KNOWLEDGE AND THE FLEXIBILITY OF MEMORY IN EARLY CHILDHOOD[99]

Dóra Kampis, Ildikó Király, József Topál

The role of imitation is of prevailing significance as a tool for knowledge transmission in relation to both the individual development of an infant when becoming a competent, knowledgeable individual, and the establishment of a cumulative culture spanning over generations (Csibra and Gergely 2009; Tomasello 1999; Boyd et al. 2011). While imitation enables individual learning based on environmental cues (and so the formation of new and inventive ideas), it also lets us pass knowledge from generation to generation and thus accumulate improvements and establish culture (Boyd et al. 2011).

Exhaustive research on children's imitation highlighted that young children show flexibility in their choice of social learning strategies, thus, blind imitation is not the only form they can use (for a comprehensive review, see Want and Harris 2002). It is well documented that children in various situations re-enact selectively (some examples are the following: on copying intentional actions but not mistakes or failed attempts, see Meltzoff 1995, and Carpenter et al. 1998; on imitating only those actions that are considered relevant in the situation, see Gergely, Bekkering and Király 2002; on copying only those intentional actions that seem causally related to the goal of the actions, see Brugger et al. 2007 and Király 2009). In other situations, however, children are ready to copy surprisingly faithfully (Whiten et al. 2009). Recently, it has been proposed that there is a dominant form of imitation, namely *overimitation,* a tendency to reproduce even the causally irrelevant actions of a modeled behavior (Lyons et al. 2007; Nielsen and Tomaselli 2010).

The main challenge for a developmental perspective is to explain the underlying mechanisms responsible for the choice between the above-mentioned, seemingly contradictory tendencies of selectivity and fidelity in imitation. Such an explanation could help us to understand why imitation is the most successful means for the propagation of cultural knowledge (Richerson and Boyd 2005). Moreover, it could highlight the possibilities of the integration of individual learning and social learning strategies.

[99] Acknowledgements: The European Union and the European Social Fund have provided financial support to the project under the grant agreement TÁMOP 4.2.1/B-09/1/KMR-2010-0003, and from the Hungarian Scientific Research Fund under K76043. The authors thank the infants and their parents for their participation.

Explanations of flexibility in imitative performance

Interestingly, the phenomenon of selective imitation has been explained consensually in terms of children's understanding of others' goals and intentional actions, suggesting that the major function of selective imitation is learning (e.g., Bekkering et al. 2000; Gergely et al. 2002; Tomasello et al. 1993; Over and Carpenter 2012). At the same time, the most influential explanations of overimitation assume that the phenomenon is a sign of the need to learn about causally opaque but culturally significant artifacts as well (Gergely and Csibra 2006; Lyons et al. 2007; Whiten et al. 2009; Flynn and Whiten 2010).

Regarding the availability and relation of selectivity and fidelity in imitation, the various explanatory theories are different in their view despite the fact that they share the assumption that imitation is a learning strategy. On the one hand, a theory of Flynn and her colleagues (Flynn 2008) proposes that children initially imitate faithfully because they accept the information as culturally relevant and contributing to the maintenance of traditions. However, during their own practice, children reconsider the role of each action element with respect to the success of goal attainment: in their performance, irrational, inefficient elements tend to fade away. In that sense, Flynn and her colleagues (Flynn 2008) allow sensitivity to off-line factors, such as time and practice, thus, their model assumes a flexible learning process.

On the other hand, Lyons et al. (2011) in their account imply that information is automatically encoded in a causal manner. Children attribute causal importance to each and every action-element presented by a model – overwriting even their experience-based expectations if necessary. The automatic causal encoding unavoidably leads to high-fidelity imitation, as a result of an inflexible process.

There is a recent approach with the objective of providing an explanatory model for the dominance of overimitation and the existence of selective imitation in a single frame. According to Over and Carpenter (2012), the complexity of children's imitative performance can only be fully understood if the social context of behavior and the potentially emerging social motives are considered, too. In their social psychological model, they claim that the goal of learning in itself, which is usually claimed to be the main function of imitation (see above), is only one factor that influences imitative behavior. There are other critical factors in determining what is copied that can be called social goals or social motives, namely, children's identification with the model and with the social group in general, and the social pressures which children experience within the imitative situation. In their view, selective imitation (emulation) and overimitation (high-fidelity imitation) are not independent processes but can be intertwined, and they might even interact with each other. They distinguish three cases of social learning situations, where the type of re-enactment depends on the particular constellation of the above-introduced specific learning or social goals on behalf of the child.

In cases when learning goals predominate, the goal of the copying is to acquire a new skill and reach the goal (hence, goal emulation). Therefore, in these cases children concentrate on the necessary steps to reach the action goal (that is, on the steps that are in causal relation with it). In this case, children pay attention more to the nature and details of the task – to the function and efficiency of the objects and the steps –, and less to their relationship with the model and their interaction.

Over and Carpenter (2012) show that even when the learning goal is important, overimitation might occur. Williamson and colleagues' (Williamson et al. 2008) results show that if children's previous experience suggests that they cannot solve the task alone, then they are more likely to faithfully copy the action of the model than in the condition when they had the experience that the task is easy to solve (this being the opposite case to when they selectively imitate because they 'know a better way' to reach the goal). In this way, however, we should say that faithful imitation emerges from an understanding of the situation, and it is not completely blind.

Another type of case is when learning goals and social goals cannot be separated from each other: they are either present in parallel, or they are strongly intertwined: such as when children learn cultural norms. The mixture of social goals and learning goals represents a special function: learning about the normative aspects of culture, about social rules that cannot just be learnt via individual learning. We will later return to the discussion of these types of situations.

The third type of copying situation is when social goals dominate. In these cases, identification with the model is of top priority for children, without necessarily aiming to learn a new skill at all. Rather, children wish to convey the message: "I am like you." The content of the social goals can vary with age. A related finding is that being imitated makes us like the imitator better because we like more those who are similar to us. In Meltzoff's (1990) study, he tested 14-month-old infants' reaction to a social partner. Results showed that infants preferred partners who imitated them to another partner who was reacting equally contingently but did not imitate them.

The advantage of this approach is that, with the help of taking into account social motives and different social factors of the situation, it shows how different functions (both epistemic functions and social functions per se) of a copying process can be bridged and used in a dynamic way. They argue for flexibility on a level of a hierarchy that exceeds the epistemic function of cultural transmission. They describe the combination of the social and learning functions of imitation as a deeply social phenomenon, though without explaining the proposed dynamic relation of the two types of goals. An implicit assumption of this model is that there is an initial choice of the overall (learning vs. social) goal by infants triggered by situational factors. The problem arises, though, how to define what kind of factors result in the dominance of learning goals, or in the dominance of social goals, or in their combination.

An interesting subfield of imitation research can help us to specify the question more precisely. There are findings where children seem to either emulate or imitate based on the model's features, the social partner's characteristics. Difference in physical features (like gender or age), or behavioral cues (like success or competence) that imply reliability can influence the extent of learning new information (Henrich and Gil-White 2001). Indeed, unconfidence by itself is a factor that entails selectivity: if 2-3-year-old children have the opportunity to choose which model to learn from, they prefer the confident one (Birch, Akmal and Frampton 2010). A recent study by Elekes and Király (2012) revealed that infants react sensitively to the features of the model and the situation: they integrate both sources to learn the most information possible. Faithful imitation was only evoked when the model seemed to be knowledgeable and the situation was pedagogical. Whenever one of these conditions was not met, infants turned to emulative strategies.

These results highlight an interesting problem: the question of how the 'decision' is made about the goal the child has in a situation. The above findings lead us to the issue that in certain cases (as in the above-mentioned examples) it is even problematic to decide whether the goal was of a learning or of a social nature (or maybe both). Children's relation to the model (their understanding of the model's characteristics) causes the goal of learning to be more important, or rather, the fact that they know "a better way" to reach the goal makes them revise their relationship with the model. Hence, they do not feel a motive to identify with her because the model's knowledgeability turns out to be questionable. The social psychological framework of copying, therefore, offers flexibility in deciding the function of imitation in different situations. By this assumption, the model solves the apparent confusion of children's choice between selectivity and fidelity in their imitative performance with respect to their learning goals. However, this still cannot identify the causal factors that are responsible for guiding the flexible 'decision' process: it is unclear what factors trigger learning goals, or social goals, or both of them simultaneously.

Alternatively, if we step back to the theoretical perspective that (1) imitation is a successful means for cultural transmission, and thus (2) the main function of imitation is learning, then it can be proposed that even the deeply social aspects of culture need to be learnt first. Selective imitation is often regarded as a heartless, cool-headed act, serving individual learning, whereas high-fidelity imitation (overimitation) is often seen as a warm, social action, serving cultural knowledge transmission.

A theoretical angle that poses an overall learning function on imitation that facilitates both the acquisition of instrumental, functional information and the acquisition of social rules and socially constituted knowledge would suggest that fidelity and selectivity are guided by the content and type of knowledge in propagation. Indeed, in many ways selective imitation might subserve culture better since if we would simply copy each other, new knowledge would never arise. On the other hand, it would be really inefficient if we had to invent everything over and over again. A good selective imitator can produce an optimal combination of innovation and knowledge transmission (Király, Szalai and Gergely 2003; Richerson and Boyd 2005).

Natural pedagogy theory (Csibra and Gergely 2009) represents this perspective and argues that the guiding function of imitation is learning – not only about instrumental knowledge, but also about the socially constituted knowledge of culture as well. More specifically, the authors claim that imitation itself is only one form of how knowledge acquisition takes place. However, this model does not deny the role of social motives in triggering imitation in several cases, though it highlights that a pedagogical setting is in itself sufficient to activate a stance in children that the situation is for knowledge transmission. Hence, this model defines the factors that help children to choose whether the situation invites their learning goals or, rather, their social motives.

Indeed, this model expects high-fidelity imitation, but only in pedagogical situations. This model claims that ostensive communication triggers in infants a stance to accept the demonstrated behavior as a relevant and generalizable piece of cultural information, even when the action is cognitively opaque (i.e., it is seemingly not the most efficient way to achieve the goal state; they are unable to comprehend it by their instrumental, functional knowledge). This approach emphasizes that imitative behavior is guided by cognitive and informational adaptivity, and since high-fidelity imitation is triggered only in ostensive communicative situations (see Király, Csibra and Gergely 2013), this model presumes an on-line selection of what is to be learnt, guided by pedagogical cues. When the model produces her actions deliberately while engaging in ostensive communication with the infant, despite the fact that her action is cognitively opaque, her intentional choice guides infants to encode the socially relevant objective of the situation (i.e., learning about social rules).

Thus, pedagogy theory predicts that ostensive demonstration highlights aspects of situations that are worth learning, so the demonstration itself guides the process of selection: elements that are manifested as relevant and new pieces of cultural information are learnt, and thus they appear in re-enactment since they are encoded as generic information in the social domain. This means at the same time that the fidelity predicted by this theory in the case of social rules and knowledge presented in pedagogical setting does not appear as a result of choosing 'imitation' as a dominant form of social learning for this type of knowledge, it rather reflects that in the case of social rules the information content cannot be filtered by instrumental, functional reasoning, and by the means of pedagogical demonstration all of the elements are labeled as relevant.

In other words, children do not tend to choose between imitation and emulation depending on the situational requirements of the setting, but they always try to find the essentially irrelevant features of the situation that could be filtered out. On the one hand, when there are obvious physical, causal affordances that help their reasoning process, the result of this filtering reflects emulation as a form of re-enactment. On the other hand, when pedagogical settings induce relevance for an otherwise opaque behavior, the selection process results in richer content and appears in the form of re-enactment, which is similar to high-fidelity imitation. In essence, it is still the emulation of the subgoals of the situation that are labeled as relevant by the pedagogical demonstration. From this perspective, re-enactment by itself is the retrieval of the information that was encoded as relevant content in the modeling situation. Seemingly high fidelity imitation emerges as a result of an acquisition process that is evolved to encode the social rules and the socially constituted knowledge that cannot be interpreted by other interpretive schemas, like instrumental, functional, or efficiency rules.

Learning and memory processes in imitation

Overall, if we accept that imitation has dominantly epistemic functions, we need to take into consideration the competencies and processes constraining learning and memory on the individual level. As Richerson and Boyd (2005) emphasize, the nature of the behavior that is available to imitate is itself strongly affected by the psychology that shapes the way we learn from others. The way learning and memory factors influence and contribute to the form of copying was studied by Simpson and Riggs (2011). They tested whether 3- and 4-year-olds' imitative behavior depends on whether they are forming short-term or long-term memories of events. They predicted that when tested immediately after demonstration, children would find it easier to remember all the steps (including the irrelevant ones), whereas after a delay the memory of the irrelevant action would fade and result in selective imitation. Indeed, results showed that during immediate re-enactment, with fresh memory traces, children copied the demonstrated action faithfully, but after about a week's delay the semantically processed long-term memories were activated, and this led children to emulate the action sequence. This suggests that they rather use their semantic knowledge for problem-solving, and they do not recall the details of original actions.

In an earlier study, Williamson et al. (2008) showed that 3-year-old children varied their copying strategies according to their experience with the task: they did not follow the new, alternative strategy if they experienced the task to be easy to solve, but if the situation changed and the task turned out to be harder than expected, children recalled the strategy of the model to solve the task.

Together, these results suggest that prior knowledge and children's memories related to the demonstrated event influence their copying strategies. Furthermore, the experience and the inferences drawn from the demonstrated behavior that is reflected in the social learning strategy depends on when the re-enactment takes place, and on the type of memory (i.e., short- vs. long-term) children have to rely on. Thus, it seems that in the case of long-term recollection children show a strong tendency to recall the 'essence,' the more semantic elements of memories. Finally, it seems that children are able to flexibly reach back to those elements of the observed action that were previously labeled irrelevant and selected out, in cases when the relevance of these steps changes and the new context reveals the efficiency or necessity of these particular elements.

The above examples of flexible selective imitations seem to suggest that the behavior of children at retrieval can also be influenced by an adaptive, efficient strategy (e.g., in the Williamson et al. 2008 study). Furthermore, it seems that the selecting mechanisms in memory processes can influence behavior: in the case of Simpson and Riggs (2011), while online, when encoding all of the observed elements were re-enacted, and after a delay the memory's pressure led to selection. This suggests a picture of imitation where different strategies can be used flexibly, and which is sensitive to social factors, as well as to efficiency analyses of situations, and which can flexibly generalize and activate memory traces.

Nevertheless, the above results are not clearly conclusive on the question during *which* memory process information selection takes place or plays a significant role. In a broader framework of cultural learning, however, the prediction in this case is biased – since cultural information in essence does not need to be transparent for the individual learning system, selection processes should be guided online, during the phase of acquisition with the help of the expert teacher. As natural pedagogy theory proposes, the culturally expert teacher manifests the relevance of the demonstrated event steps and the explicit manifestation lets the novice learner filter out the culturally shared and relevant content of the demonstration. From the above focal question, it would mean that selection mostly occurs during the encoding phase.

If selection takes place during encoding, then it has to be decided immediately (1) which steps and elements are necessary (relevant), and which ones are unnecessary (irrelevant), and (2) what should be done with the irrelevant steps. The second point is especially important, since if the irrelevant steps are kept at encoding, then they have to be left out during retrieval. This leads us to the question of how the irrelevant elements are stored: is the whole action sequence stored in one memory, and the irrelevant parts are left out at retrieval? Or are they stored separately, and reached back to in case they are needed?

In a preliminary study, we investigated whether the memories and memory processes of children, which influence encoding as well as storing and retrieving processes, are flexible enough to adapt to a situation with changed contextual parameters. That is, what is the nature of selection observed in imitative behavior? Does it happen on-line, during the encoding phase (as predicted by a cultural learning perspective), or is it a flexible process and selection occurs in retrieval?

In our experiment, we demonstrated 2-year-old children events with changing contexts, where the different contexts either verified or disproved the use of a novel tool as opposed to a familiar action to reach a certain goal. We tested that regarding the tool use with semi-opaque relevance, (1) what is the typical copying mechanism during immediate re-enactment? Do children selectively imitate the tool use in the condition where the context requires it, and leave it out when it is does not? (Not only does tool use in the 'relevant' condition make reaching the goal easier, but it is also a necessary step without which it is not possible to reach the goal. In the 'irrelevant' condition, it is not necessary at all, and it is rather inconvenient and requires plus effort which is not justified by the context). And (2) whether in deferred imitation, if the context changes (i.e., from being irrelevant it becomes relevant, therefore, its function has to be recalled and the step inserted into the causal chain of action steps; or the other way around, from relevant it becomes irrelevant and, therefore, has to be eliminated from the action steps), children flexibly adapt their retrieval processes. More specifically, whether they leave out the step after the delay if it becomes unnecessary and integrate it into their actions when it becomes necessary.

We used two boxes to create the two separate contexts. The two boxes were almost identical; they differed in only one key aspect that determined the relevance of the tool use. The tool was a small finger glove with a Velcro-like ending that stuck to the little plush toys that were inside the box and whose retrieval was the goal of the action. One of the boxes had a small hole; therefore, the finger gloves were necessary to retrieve the toys (the fingers did not fit into the hole well enough to grab the toys). The other box had a large hole; therefore, the gloves were unnecessary for the retrieval of the toys.

We also checked whether a step that is always irrelevant (rolling the finger with the glove on the top of the box) and never necessary to achieve the goal because it is not in the causal chain of events remains in the action sequence of the children on the first and second testing occasion. We call this step the opaque irrelevant step since it is unnecessary in both contexts but its function is opaque to the children.

Demonstration always included both steps in question: the step with the changing relevance and the always-irrelevant step. Half of the children saw the demonstration and performed the immediate re-enactment with the box with the big hole. A week later (without demonstration), they received the box with the small hole. The other half of the children had everything the other way around. To see whether it is necessary to highlight the function of the tool, in order for children to selectively use it when necessary, we also varied whether we highlighted the situational constraint at the beginning of the demonstration (showing whether her hand fits into the hole), or not.

On one hand, the contexts on the two occasions were very similar since we had almost the same object and same situation (occurring at the same place, with the same experimenter). Nonetheless, they differed in one key element which might challenge children's memory processes. This is similar to the Williamson et al. (2008) study in the sense that the situation provides children with experience about the efficiency of the prepotent response, but, compared to the original situation, in deferred imitation we changed the contextual frames in both cases. One of the situations, therefore, is very similar to Williamson et al.'s (2008) study since the situational information changes in a way that a strategy, a means of action that was efficient before becomes inefficient; therefore, children have to implement another, previously observed but not yet used means of action as a strategy. Here, the decision of using one of the available (not demonstrated) means of actions is online, and later the retrieval of the other (demonstrated) one is made based on the efficiency criteria of action. In contrast, the other condition aims to test whether, if something has been learnt to be relevant and hence to be followed, children can update this social knowledge in accordance with the changing situational constraint with regard to efficiency indicators. In this case, if we assume that a means of action is manifested as a culturally or socially accepted way of attaining a specific goal, it is encoded as a subgoal of the situation. Hence, the choice criteria are more social than instrumental, the efficiency criteria do not apply, therefore, they cannot guide the retrieval later. This would test whether we unconditionally accept something that we regard as relevant during encoding. Still, it is worth seeing whether, despite of this, we can flexibly update the information formerly labeled as social affordance, and whether its (irrelevant) meaning becomes transparent. Hence, this makes it necessary to use immediate and deferred re-enactment (as opposed to Simpson and Riggs 2011), to see whether retrieval strategies change and adapt in a flexible way.

Results so far suggest that at immediate re-enactment children imitate the use of the tool selectively if the model previously directed their attention to the situational constraints determining the relevance of tool use. Far more children used the glove when it was necessary to achieve the goal, and almost none of them used it when it was not. Interestingly, at the second time (when they received the other box) they did not change their strategy: whatever they did the first time, they repeated the second time. With regard to the always-irrelevant step the picture looks slightly different: roughly one third of the children imitated this step the first time, equally in the two conditions. However, the second time, children who experienced the demonstration and the immediate re-enactment in the context where the tool use was irrelevant kept this step even in the second testing, but those for whom the tool use was relevant the first time, left out this step at the second timing.

It seems that children do not change their strategies even though their own action is proven to be unsuccessful the second time (in the condition where the tool use becomes relevant and, therefore, there is pressure towards it), or it is clearly unnecessary (where the gloves become unnecessary to use because their hands fit perfectly into the whole). This might suggest that selection happens during encoding, and this guides the appearance of the matching social learning mechanism. Furthermore, it implies that their memory processes indeed seem to be inflexible: it is not just that they 'forget' the element in question and, therefore, they cannot use it in the changed situation (in the condition where the successful goal attainment requires the recall of that specific element) since in the other condition (where the step in question becomes irrelevant) all they would have to do is to leave out that step.

This is in accordance with the model of natural pedagogy (note that the demonstration included ostensive communicative cues): the information that receives the label of generic social knowledge is kept stable and is considered relatively trustworthy. Knowledge transmission happens in a cultural situation, and even though at the first time there is immediate selection, later on that information is kept. In other words, information selection occurs during encoding as a result of cues that highlight the relevance of event elements for the infants. The selection process, in this sense, could be conceived as a result of social guidance in the demonstration situation, and not as filtering by the individual learning system. The significant role of the social partner in this process thus endorses the stability of information that was transmitted – a crucial aspect of cultural contents.

This supports the notion that the process of imitation and emulation are not two separate entities since we propose that learning processes reflect the level of the goal hierarchy where emulation takes place. This stance can provide the basis for the seemingly faithful imitation of knowledge labeled as social information that produces less variability and result in more stable cultural knowledge transmissions. This can be more fruitful than parallel emulation or imitation processes – there is some amount of selection that occurs, but then they keep the acquired information. As Richerson and Boyd (2005: 12) write, "[i]n many kinds of environments, the best strategy is to rely mostly on imitation, not your own individual learning ... then the lucky or clever of the next generation can add other tricks." There are, of course, other cases when the information is not kept. For instance, in the condition where the tool use became irrelevant there was further selection towards the always irrelevant step, causing children to leave out that step during second re-enactment. Nonetheless, situations might occur where there is imitation without selection. The question is, however, if there really is such a thing as 'blind' imitation.

An alternative possibility opposed to our above-argued emulation hypothesis is that children include a particular step 'just to be on the safe side' if the relevance of something is opaque, or if the imitation itself becomes the goal (e.g., in the case of social norms) – not a 'social goal,' but, because of the social situation, it becomes a relevant step. This, then, results in cases where the observer might just have this in mind: "I trust you. There has to be a reason why we do it."

Then again, regarding the always irrelevant step, it is possible that if the tool use itself is irrelevant during demonstration, then infants label the whole event as 'irrelevant' and code everything in this frame. Remarkably, they leave out tool use during immediate re-enactment if it is irrelevant, but they still keep the other irrelevant step. These two steps differ in the opacity of their function. In the case of tool use, it might be easier to detect its necessity for reaching the goal, but there is no explicit information about the other irrelevant step, hence its relevance is opaque – this might explain why more infants keep this step than the tool use. But it is still a question why those infants leave out this step in the second testing, when the tool use was relevant during demonstration and immediate re-enactment (first testing). It is possible that, when they try out the action for themselves, the causal relations become more obvious, and it becomes clear that the irrelevant step is not necessary, hence a second selection occurs between the first and the second testing. However, it is interesting that these children, despite the possible 'enlightenment' about the causal structure of the scene, do not leave out the tool use itself at the second re-enactment, even though all they would need to do is use their hands instead of a slightly complicated procedure of the tool use.

This again means that memory processes play a role in the type of behavioral answer that occurs: the overall goal manifested by communication alters what is marked as generalizable, semantic knowledge, and, therefore, kept in the child's memory. The informational content influences the perceived form of re-enactment, whereas the more detailed semantic content seems to reflect high-fidelity imitation (Király 2008). An everyday example is that, even though certain things would be much easier to eat with our hands, it depends on the cultural circumstances (for instance, if we are having a Hungarian meal at home, or enjoying an Indian dinner at our colleague's place) whether we stick to our socially accepted way of eating and using cutlery. In any case, we rely on our social scripts, thus on our socially constituted generic knowledge.

Potential effects of memory processes on cultural transmission

In Simpson and Riggs's (2011) experiments, children showed a rather different pattern: during immediate reenactment, they imitated the actions faithfully, and, after a delay, they imitated the necessary action steps selectively and left out the causally irrelevant step. The authors refer to Heyes and Bird (2007), who suggested that these two forms of copying could be explained by two separate types of mechanisms. While overimitation can be explained by sensorimotor models, where sensory input representations are directly linked to motor representations and hence enable the copying of meaningless actions, selective imitation involves a process where, between sensory and motor representations, there are higher level, conceptual representations (such as representations of the action goal) that enable goal emulation. Simpson and Riggs (2011) argue that the most plausible explanation is that, when children observe the demonstrated behavior, they form two kinds of representation: one sensorimotor and one conceptual. While the conceptual one remains strong, the sensorimotor representation fades away, and this leads to emulation in the long term. We sense a contradiction here: if the sensorimotor representations arise because the conceptual, higher-level representations cannot be formed, then how is it possible to form both types at once? Another question is what decides during immediate re-enactment which 'form' will win? According to Simpson and Riggs's (2011) data, sensorimotor representations 'win,' and this leads to the imitation of the irrelevant steps. But this suggests that the simple, sensorimotor representation is more dominant than the conceptual one, and that the conceptual representation could only exert influence if the sensorimotor has already faded.

In contrast with this, our data suggest that, already at immediate re-enactment, the conceptual representation would be in charge. The semantic content of encoding depends on both the cognitive opacity of the situation and the communicative signal present at modeling. These two factors induce in children the identification of the information that is worth learning. Inherently, when infants are guided to be able to reason according to their physical and functional knowledge, they select the semantic content by the help of their instrumental knowledge base. On the other hand, when children are guided by pedagogical settings to encode the cognitively opaque aspects of the situation as relevant, they select more elements (or all of the elements) as generalizable social knowledge, that is, semantic, generic elements as well. In the first case – since instrumental, efficiency guided reasoning was responsible for the encoding of the content – the semantic information is associated with the functional, instrumental domain. Yet, in the case of cognitively opaque contents labeled as relevant by pedagogical settings, the learnt semantic information could be associated with the social domain.

Based on this differentiation, it can be further supposed that the instrumental domain can be used not only for encoding, but children can also apply their efficiency based interpretative schemas to update the information. However, in the social domain, these efficiency based interpretative schemas do not apply, so that is why they do not update the information within this domain. The information which was introduced in a social learning situation, and which was signaled as 'culturally' relevant seems to be sensitive only to social refinement cues, and not to efficiency cues. This could be the reason for their inflexibility in performance. This inflexibility represents the guiding role of high-fidelity imitation and a selection bias during the learning phase (during encoding) in the case of contents that are signaled as relevant and also generalizable pieces of cultural information. It could also be a developmental issue, since Simpson and Riggs (2011) tested older children, hence there might be a fundamental difference between 2-year-olds on one hand, and 3- and 4-year-olds on the other. The question, however, is: what is it that changes? It can be supposed that already at the age of 2 children form conceptual representations even during immediate re-enactment. These representations can be part of 2 domains: instrumental and social. Regarding the social domain, we do not expect flexibility since there is no clear-cut inferential principle that could guide the updating. In the instrumental domain, it seems that 3-year-olds are ready to flexibly retrieve additional information according to the needs of the situation, resulting in a different strategy in re-enactment. However, at the age of 2 they are not yet able to do so.

Another factor in our results could be that the immediate re-enactment after demonstration would have an effect on children's strategy after the delay. But even if motor reinforcement is so strong that it leads children to stick to their original strategies (that is, if they would switch flexibly if there was only a delayed test and no immediate re-enactment), it leaves us with a puzzle. Is motor reinforcement, having performed a task once in a certain way, stronger than infants' conceptual and causal understanding of events? Once we do something in a particular way, will we not be able to change our way of doing it? One could argue that a second demonstration (before the delayed re-enactment) could help infants to correct their strategies – but they do not even always follow the demonstrator in the first place (i.e., they overwrite the demonstrator's strategy in the tool use irrelevant condition).

Our claim is that cultural transmission is a mixture of propagating knowledge from the instrumental and social domains. When a situation 'delivers' knowledge in the instrumental domain, even children and infants use their individual learning strategies and reasoning skills to optimize the content, and they learn an efficient way to attain a functional, instrumental goal. The learning is flexible, selective online, and probably also flexible during memory retrieval, later in development. However, when the pedagogical setting induces that the delivered knowledge is social in essence, for instance, it is a socially accepted way of attaining a goal, a social rule or a social norm, infants encode it as an important subgoal. The result of this encoding process is inflexibility since social norms do not necessarily follow efficiency criteria. Since the social content cannot be inferred with the help of reasoning principles, this inflexibility holds on.

We agree with Richerson and Boyd (2005: 8) on the notion that "individual psychologies determine which ideas are likely to be easy to learn and remember and which kinds of people are likely to be imitated" in the sense that the domain of the conceptual knowledge in question (whether it is social or instrumental) also influences the dynamic way of re-enactment in imitative situations. We claim that cognitive processes strongly influence cultural transmission; moreover, there is a mutual correspondence between cognitive processes acting on the potential knowledge domains and the cultural contents transmitted. This dynamic relationship determines where fidelity remains, promoting stability of culture in the form of the matrices of social norms and rules, and where we can give space to invention and creativity. And finally, this is how imitation can be mixed with the influence of individual learning, resulting in the population adaptation outreaching any individual achievement possible.

References

Bekkering, H., Wohlschlager, A. and Gattis, M. (2000) Imitation of gestures in children is goal-directed. *Quarterly Journal of Experimental Psychology* 53, 153–164.

Birch, S.A.J., Akmal, N., and Frampton, K.L. (2010) Two-year-olds are vigilant of others' non-verbal cues to credibility. *Developmental Science* 13, 363–369.

Boyd, R., Richerson, P.J. and Henrich, J. (2011) Rapid cultural adaptation can facilitate the evolution of large-scale cooperation. *Behavioral Ecology and Sociobiology* 65, 431–444.

Brugger A., Lariviere, L.A., Mumme, D.L. and Bushnell, E.W. (2007) *Child Development* 78, 806–824.

Carpenter, M., Akhtar, N. and Tomasello, M. (1998) 14- through 18-month-old infants differentially imitate intentional and accidental actions. *Infant Behavior and Development* 21, 315–330.

Csibra, G. and Gergely, G. (2009) Natural pedagogy. *Trends in Cognitive Sciences* 13 (4), 148–153.

Elekes, F. and Király, I. (2012) Szelektív utánzás a modell megbízhatóságának függvényében: A modell életkorának szerepe. *Magyar Pszichológiai Szemle* 67 (3), 449–466.

Flynn, E. (2008) Investigating children as cultural magnets: Do young children transmit redundant information along diffusion chains? *Philosophical Transactions of the Royal Society B* 363, 3541–3551.

Flynn, E.G. and Whiten, A. (2010) Studying children's social learning experimentally "in the wild." *Learning and Behavior* 38 (3), 284–296.

Gergely, G. and Csibra, G. (2006) Sylvia's recipe: The role of imitation and pedagogy in the transmission of cultural knowledge. In: N.J. Enfield, S.C. Levenson (eds.) *Roots of Human Sociality: Culture, Cognition, and Human Interaction.* New York, NY: Berg.

Gergely, G., Bekkering, H. and Király, I. (2002) Rational imitation in preverbal infants. *Nature* 415, 755.

Henrich, J. and Gil-White, F. (2001) The evolution of prestige: Freely conferred deference as a mechanism for enhancing the benefits of cultural transmission. *Evolution and Human Behavior* 22 (3), 165–196.

Heyes, C. and Bird, H. (2007) Mirroring, association, and the correspondence problem. In: P. Haggard, Y. Rossetti and M. Kawato (eds.) *Sensorimotor Foundations of Higher Cognition: Attention and Performance* XXII. New York, NY: Oxford University Press, 461–479.

Király, I. (2008) Memories for events in infants: Goal relevant action coding. In: T. Striano, V. Reid (eds.) *Social Cognition: Development, Neuroscience and Autism.* New York: Wiley-Blackwell, 113–128.

Király, I. (2009) The effect of the model's presence and of negative evidence on infants' selective imitation. *Journal of Experimental Child Psychology,* 102, 14–25.

Király, I., Csibra, G. and Gergely G. (2013) Beyond rational imitation: Learning arbitrary means actions from communicative demonstrations. *Journal of Experimental Child Psychology* 116, 471–486, http://dx.doi.org/10.1016/j.jecp.2012.12.003.

Király, I., Szalay, Á. and Gergely, G. (2003) Mit utánzunk és miért? A vak mimikritől a belátásos utánzásig. *Magyar Tudomány* 8.

Lyons, D., Damrosch, D., Lin, J., Simeone, D. and Keil, F. (2011) The scope and limits of overimitation in the transmission of artifact culture. *Proceedings of the Royal Society B* (in press).

Lyons, D.E., Young, A.G. and Keil, F.C. (2007) The hidden structure of overimitation. *Proceedings of the National Academy of Sciences of the USA* 104 (19), 751–756.

Meltzoff, A.N. (1990) Foundations for developing a concept of self: The role of imitation in relating self to other and the value of social mirroring, social modeling, and self practice in infancy. In: D. Cicchetti and M. Beeghly (eds.) *The Self in Transition: Infancy to Childhood.* Chicago: University of Chicago Press, 139–164.

Meltzoff, A.N. (1995) Understanding the intentions of others: Re-enactment of intended acts by 18-month-old children. *Developmental Psychology* 31, 838–850.

Nielsen, M. and Tomaselli, K. (2010) Overimitation in Kalahari Bushman children and the origins of human cultural cognition. *Psychological Science* 21, 730–736.

Over, H. and Carpenter, M. (2012) Putting the social into social learning: Explaining both selectivity and fidelity in children's copying behavior. *Journal of Comparative Psychology* 126, 182–192.

Richerson, P.J. and Boyd, R. (2005) *Not by Genes Alone: How Culture Transformed Human Evolution,* Chicago, IL: University of Chicago Press.

Simpson, A. and Riggs, K.J. (2011) 3- and 4-Year-Olds Encode Modeled Actions in Two Ways Leading to Immediate Imitation and Delayed Emulation. *Developmental Psychology* 47 (3), 834–840.

Tomasello, M. (1999) *The Cultural Origins of Human Cognition.* Cambridge, MA: Harvard University Press.

Tomasello, M., Kruger, A.C. and Ratner, H.H. (1993) Cultural learning. *Behavioral and Brain Sciences* 16, 495.

Want, S.C. and Harris, P.L. (2002) How do children ape? Applying concepts from the study of non-human primates to the developmental study of 'imitation' in children. *Developmental Science* 5 (1), 1–14.

Whiten, A., McGuigan, N., Marshall-Pescini, S. and Hopper, L.M. (2009) Emulation, imitation, over-imitation and the scope of culture for child and chimpanzee. *Philosophical Transactions of the Royal Society B,* 364, 2417–2428.

Williamson, R.A., Meltzoff, A.N. and Markman, E.M. (2008) Prior experiences and perceived efficacy influence 3-year-olds' imitation. *Developmental Psychology* 44, 275–285.

Developmental Science

Developmental Science (2016), pp 1–8 DOI: 10.1111/desc.12398

SHORT REPORT

Preschoolers have better long-term memory for rhyming text than adults

Ildikó Király,[1,2] Szilvia Takács,[1,2] Zsuzsa Kaldy[3] and Erik Blaser[3]

1. *Department of Cognitive Psychology, Eötvös Loránd University, Hungary*
2. *Institute for Advanced Study, Central European University, Hungary*
3. *Department of Psychology, University of Massachusetts Boston, USA*

Abstract

The dominant view of children's memory is that it is slow to develop and is inferior to adults'. Here we pitted 4-year-old children against adults in a test of verbatim recall of verbal material. Parents read a novel rhyming verse (and an integrated word list) as their child's bedtime story on ten consecutive days. A group of young adults listened to the verse, matching the exposure of children. All participants subsequently performed a free-recall of the verse, verbatim. (Parents and young adults knew they would be tested; children did not.) Four-year-olds significantly outperformed both their parents and the young adults. There were no significant differences in the ability to recall the gist of the verse, nor the integrated word list, allaying concerns about differences in engagement or motivation. Verbatim recall of verse is a skill amenable to practice, and children, we argue, by virtue of the prominence of verse in their culture and their reliance on oral transmission, have honed this skill to exceed adults'.

Research highlights

- Long-term, verbatim memory for a novel, rhyming verse was tested in three groups: 4-year-olds and their parents (who read them the verse) and in a group of young adults.
- Four-year-olds outperformed both groups of adults, with free-recall of nearly twice as many correct words of the verse, and far fewer errors.
- Children's memory for *verbatim* recall is excellent as they cultivate a skill for retaining verse.
- Children form a preliterate society reliant on memory for rhythm and rhyme for the oral transmission of their culture.

> But when they came to letters, this, said Theuth, will make the Egyptians wiser and give them better memories … Thamus replied: … this discovery of yours will create forgetfulness in the learners' souls, because they will not use their memories; they will trust to the external written characters and not remember of themselves . . . . an aid not to memory, but to reminiscence . . . . (Plato, *The Phaedrus*, quoting Socrates, approx. 370 BC)

Introduction

Preliterate societies have relied on verbal memory and recall to transmit culture for thousands of years. While the memory abilities of members of these societies may not live up to the myth (Goody, 1998), the skill nonetheless finds continuous, obligatory exercise as there is no external storage. Here we describe a group of preliterate individuals that similarly exercise and rely on verbal memory – particularly for verse – for the transmission of their culture: young children.

In 1975 Ann Brown noted that 'rhymes, accompanied by music, are readily acquired and can be reproduced exactly even by quite young children . . . . The efficiency of using musical rhymes as information sources, while extensively used by media advertising aimed at children,

Address for correspondence: Ildikó Király, Department of Cognitive Psychology, Eötvös Loránd University, Izabella u. 46, Budapest, Hungary, 1064; e-mail: kiralyi@caesar.elte.hu

and programs like Sesame Street, has not been studied by developmental psychologists' (Brown, 1975, p. 112). Forty years later, this statement is still largely true. This leaves unevaluated the suspicion of many parents: that their children – characterized as having weaker memory in every other domain (STM, Gathercole, 1998; LTM and declarative memory, Bauer, Wenner, Dropik & Wewerka, 2000; Hayne, Boniface & Barr, 2000) – remember their nursery rhymes better than they do. Here we compare long-term memory for verse in a group of 4-year-olds to that of their parents (who read them the verse) and to a group of young adults (who passively listened to the verse). Four-year-olds outperformed both groups.

Children's verbal memory

Children's verbal memory shows protracted development. Development brings increases in representational flexibility (Bauer & Dow, 1994; Barnat, Klein & Meltzoff, 1996; Gergely, Bekkering & Király, 2002; Király, 2009) and in the capacity of declarative memory (Hayne *et al*., 2000; Bauer *et al*., 2000), as well as gradual increases in working memory performance for verbal material into adolescence (Gathercole, 1998), that is likely related to emerging language abilities (Simcock & Hayne, 2003). Protracted development is also evident in the Paired Associate Learning (PAL) test of the Wechsler Memory Scale (e.g. Halperin, Healey, Zeitchik & Ludman, 1989; Beardsworth & Bishop, 1994; see also Heil & Jansen, 2008) and in contrasts between verbatim and gist memory; e.g. Reyna and Kiernan (1994) found that verbatim memory for prose decayed faster than gist, in 6- and 9-year-olds (see also Brainerd & Gordon, 1994).

Children's verbatim memory for verse specifically (rhyming stories, songs, and poems) has not been well studied. Instead, the focus has been on whether verse helps children remember *content* (do rhymes help children learn?). Results have been mixed. For example, Sheingold and Foundas (1978) found that while rhyming helped 6-year-olds remember the sequence of story events, content memory was the same as for prose. Hayes and his colleagues (Hayes, Chemelski & Palmer, 1982; Hayes, 1999) found that memory for content presented in verse was actually worse than when presented in prose, unless children were tested specifically on content carried by the rhyming words themselves. Similarly, it was found that children show better memory for content when presented in prose form rather than an educational televised song (Calvert, 2001; Calvert & Billingsley, 1998). These and similar results have led to skepticism about the utility of verse as an educational aid.

How can we reconcile these results then with the conventional wisdom (Read, 1976; Brown, 1975) that children's memory for verse is somehow better than for prose? The critical factor is that the studies showing poorer memory for verse were based on tests of content, while anecdotal evidence that children better remember verse is typically based on their ability to *recall the verse itself*. For content retention, verbatim coding is not necessary. Indeed, Hayes *et al*. (1982; Hayes, 1999) supposed that since verse is better liked by children, and intrinsically encourages attention to phonological characteristics (i.e. the rhythms and rhymes themselves), children are biased toward the retention of this information, at the expense of content. This makes a concrete prediction: children's memory for the phonological characteristics of a verse – i.e. *verbatim* recall – should be excellent (see Calvert & Tart, 1993; Read, Macauley & Furay, 2014).

Memory for verse is a skill

Sachs's (1967) classic study in adults demonstrated that verbatim information decays quickly, even in short-term memory. With verse, memory is more robust. Tillmann and Dowling (2007) had adult participants attempt to discriminate between a phrase drawn from a memorized text versus a paraphrased lure. For prose, verbatim memory declined over time, but for verse, it did not. This is consistent with Rubin's (1995) observation that verse aids memory by providing constraints during recall (e.g. the position of a to-be-recalled word in the rhyme scheme of the verse may influence retrieval: a memory that the next to-be-recalled word is not just an animal, say, but also a one-syllable rhyme for 'rat', trims the possibilities; see Rubin, 1995; Rubin & Wallace, 1989; see also Bower & Bolton, 1969).

Importantly for our study, the exploitation of such constraints is a skill amenable to practice – analogously to experts' famously skilled memory for chess positions (Chase & Simon, 1973; for a review see Ericsson & Kintsch, 1995; a phenomenon that appears even with young, 10-year-old chess experts, whose memory can exceed that of naïve adults (Chi, 1978; Schneider, Gruber, Gold & Opwis, 1993)). With verse, Rubin, Wallace and Houston (1993) found that when novice adults memorized and recalled a set of ballads, they were better at memorizing and recalling a subsequent, novel ballad than untrained adults; various useful constraints, for example 'surface-level' cues (rhyme and rhythm) and content-level cues (stereotyped event structure), had been acquired. We argue that children – by virtue of the prominence and ubiquity of verse in their life, their preliterate inclination to memorize it, and their dependence on oral transmission – practice this skill more than adults.

Method

Participants

Thirteen *Parents* (mean age = 35;6 years, *SD* = 3;4 years, age range: 29–41 years, all females) and their 14 *Children* (one twin pair; mean age = 4;8 years, *SD* = 2 months, age range = 4;6–4;10 years, 10 females) from Budapest, Hungary participated in the study (one additional pair was excluded due to the child's refusal to participate during testing). Four-year-olds were chosen for this study because they were not yet reading (as confirmed by their parents) but had experience with verse. Thirteen university students from Eötvös Loránd University in Budapest participated in the *Young Adult* group (mean age = 25;10 years, SD = 4;2 years, age range: 21–33 years, 7 females), and received class credit. All participants were native Hungarian speakers. Sample sizes were determined by an a priori power analysis and were similar to Rubin *et al*. (1993). All experiments were conducted in accordance with the relevant ethical regulations, and the approval of the Ethical Committee of the Faculty of Education and Psychology, Eötvös Loránd University. Adults gave informed consent prior to participation; children, assent.

Materials

Our verse was a short, 167-word, rhyming (AABB rhyme scheme) poem, '*The Radish-nosed King*' by Aliz Mosonyi (see Supplementary Materials). We chose this poem because, while suitable for 3–5-year-old children, it has a varied, interesting vocabulary, verse structure, and content that makes it engaging for adults as well. The verse was novel to all participants.

We also tested participants with a 'word list' of eight unrelated words. This list was included as a measure of general attention and engagement. Four of these words were nonsense words that conformed to the phonological rules of Hungarian (*irim*, *tentusz*, *kavu*, *bólum*), and four were meaningful words selected from the 400 and 800 most frequent words in the Essex Children's Printed Database (*kalap [hat]*, *ruha [dress]*, *csónak [boat]*, *tenger [sea]*). We introduced this distinction to probe the contrast that Calvert and Billingsley (1998) reported in the verbatim recall of 'Frere Jacques' where English-only speaking children better recalled the 'nonsense' (French) version than the meaningful English 'Brother John' version. Critically, the word list was integrated with the verse but, by design, did not share in its rhythm and rhyme. Integration was achieved by adding a short introduction to indicate that the main character (the Radish-nosed King) spoke the words on the list. The word list either preceded or followed the verse (counterbalanced within groups). When it appeared before, the verse started with, 'I will tell you a story about the Radish-nosed King who is a very peculiar fellow. When the King is angry, he shouts like this: {*Kalap!*, *Irim!*, *Ruha!*, *Tentusz!*, *Csónak!*, *Kavu!*, *Tenger!*, *Bólum!*}'; after the verse it read, 'I've told you a story. . . .'

Procedure

Parent–child protocol

Parents were asked to read *The Radish-nosed King* (from a picturebook) as their 4-year-old's bedtime story for ten consecutive nights. We chose this procedure because Reyna and Brainerd (1995) suggested that a greater opportunity to practice enhances recall of verbatim and content information even in young children. Parents were instructed to avoid discussing or reading the verse outside these readings. Parents were asked whether their child had interrupted the reading (e.g. with comments or questions). Only one mother reported that her child asked her to explain the word 'cigánykerekeztek' [*do cartwheels*]. (This is consistent with a pilot study using the same procedures, where videotaped recordings of the reading sessions showed no substantive interaction.) Importantly, parents were told that they would be tested at the end of the series of sessions, while children were not.

Young adult protocol

Young Adults received the instructions and test materials as an audio recording (recorded by a female reader, who used child-directed speech to mimic the recitation style of parents, and signaled when to turn the pages of the book). Young adults were asked to listen, and only listen, to the verse at bedtime for ten consecutive days, while looking at the pictures in the book (the text was excised), thus mimicking the experience of the child group. Young adults were given a written schedule, periodic reminders, and were asked to report any lapses in protocol (none were reported). Young Adults were told that they would be tested on their recall.

Free-recall and gist tests

On the day following the last session, a battery of tests was administered. First, all participants attempted a free-recall of the verse, verbatim, using just the original storybook's illustrations as cues (please see Supplementary Materials for an example recording from a child participant). We used free-recall as this places greater demands on verbatim memory than recognition tests, and more efficiently assesses knowledge of surface structure, i.e. the full set of words, in sequence. These procedures are similar to those used in Rubin *et al.*'s (1993) work on memory for ballads, and parallel those of oral traditions in preliterate societies (Goody, 1998). During recall, if a participant paused for more than 3 seconds, or asked for help, they were prompted with the next word in the verse. (We introduced prompts since Beardsworth and Bishop (1994) found that children who were unable to recall a verse after a 45-minute delay often showed dramatic improvements when given a single prompt.) Following verse recall, participants attempted to recall the word list. Prompts were not given during word list recall. Next, participants were asked about the gist (e.g. 'What was the story about?'). If a participant failed to list the main characters and the three central events in the verse spontaneously, additional, open-ended questions were asked (e.g. 'Who else was there?', 'What happened next?').

Afterwards, we administered the Peabody Picture Vocabulary Test (Dunn, 1959, Hungarian adaptation: Csányi, 1976) to measure children's verbal competence. Also, the socioeconomic status of parent–child pairs was assessed using a standard SES questionnaire. Results from these tests are reported in the Supplementary Materials.

Scoring

Verse recall reflects the number of correct words, in proper sequence, produced during free-recall, including articles. Correct word stems, but with the wrong case, were considered correctly recalled. We also calculated *verse error*, a sum of *intrusion* errors (erroneous words produced during recall) and *confusion errors* (paranyms, synonyms, and word or line order transpositions). *List recall* was analyzed separately, counting each correctly recalled word and nonsense word (maximum: 8). *Gist recall* was coded by determining the number of recalled main characters (out of 3: the Radish-nosed King, the radish children, and the mouse) and the number of correctly recalled main events (out of 3: the anger of the Radish-nosed King, the actions of the mouse, and the King's forgiveness), as scored by six independent raters.

Results

Free-recall for the verse

Children correctly recalled significantly more words, with significantly fewer errors, than both adult groups. First, the dependent variable of the mean number of correctly recalled words (*verse recall*) was analyzed, using an ANCOVA with between-subject grouping variables for group (Children, Parents, or Young Adult) and 'list placement' (list before, or after, verse). The number of *prompts* each participant received was used as a covariate. The analysis revealed no significant effect of list placement ($F(1, 39) = 1.118$, $p = .298$) so it was dropped from further analyses. The number of *prompts* did not play a significant role in recall performance ($F(1, 39) = 0.513$, $p = .479$) either. However, we found a significant main effect of group ($F(2, 37) = 6.230$, $p = .005$; $\eta^2 = 0.277$). Post-hoc analyses revealed that Children recalled more words on average (mean *verse recall*: 117.4 words, $SD = 30.7$, out of the 167 total words in the verse) than Parents (mean *verse recall*: 87.2 words, $SD = 38.6$; $t(25) = 2.284$, $p = .081$, effect size $r = 0.397$) and Young Adults (mean *verse recall*: 70.3 words, $SD = 34.5$; $t(25) = 3.831$, $p = .003$, effect size $r = 0.584$ (see Figure 1). The difference in performance between the two adult groups was not significant ($t(24) = 1.220$, $p = .606$), and there was no significant interaction between the factors ($F(2, 37) = 0.990$, $p = .381$). All t-tests were two-tailed and Bonferroni corrected.

Verse errors

The number of inaccurately recalled words was compared using a univariate ANOVA (with group serving as a between-subject variable). Since Levene's test of homogeneity of variance was significant ($F(5, 34) = 7.113$; $p < .001$), we used Welch's ANOVA. This analysis yielded a significant effect of group (Welch d (2, 16) = 17.160, $p = .0002$; $\eta^2 = 0.360$). Post-hoc analyses confirmed that Children made fewer errors (mean *verse error*: 7.6 words, $SD = 7.94$) during recall than Parents (mean *verse error*: 41.6 words, $SD = 33.88$; $t(25) = -3.740$, $p = .009$, effect size $r = .599$) and Young Adults (mean *verse error*: 54.9 words, $SD = 35.23$; $t(25) = -5.00$, $p = .0002$, effect size $r = 0.707$). There was no significant difference between parents and Young Adults ($t(24) = -0.980$, $p = .697$; see Figure 1).

Potential differences in the pattern of error categories were analyzed using a repeated measures mixed-type ANOVA with the number of errors by error type (*intrusion* versus *confusion*) as within-subject variables and group as a between-subject variable. The main effect of group was significant ($F(2, 37) = 10.330$; $p < .0001$, $\eta^2 = 0.358$), as was error type ($F(1, 39) = 25.570$; $p < .0001$, $\eta^2 = 0.409$). In addition, there was a significant interaction between the two factors ($F(2, 37) = 5.230$, $p = .01$, $\eta^2 = 0.220$). Post-hoc tests showed that

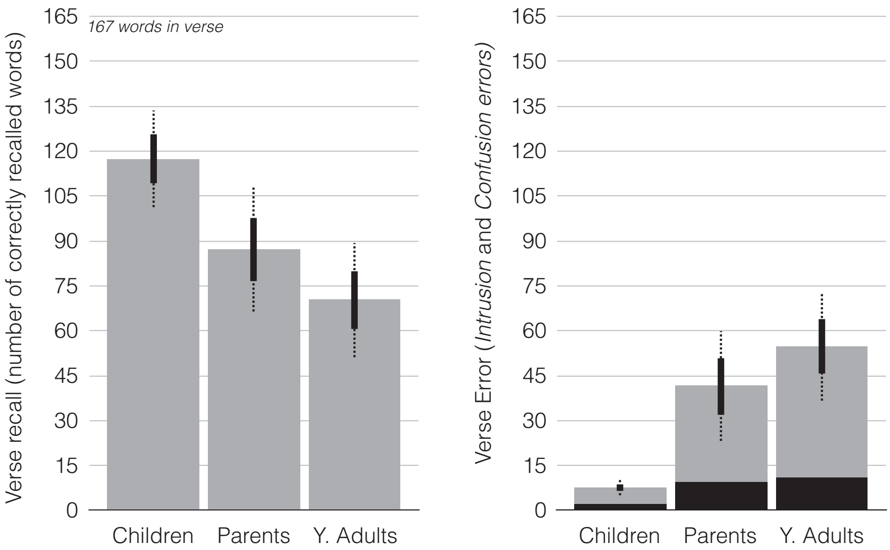


Figure 1 *Verbatim, free-recall performance of rhyming verse; 4-year-olds versus adults. Mean verse recall (left panel) and verse error (right panel) is presented for each group. Verse error consists of intrusion errors (top, gray area of stacked bar) and confusion errors (black area). Solid error bars indicate standard errors; dotted, 95% confidence intervals.*

Children's pattern of errors significantly differed from that of Parents ($F(1, 25) = 13.960$, $p = .009$, $\eta^2 = 0.358$) and from the Young Adults' ($F(1, 24) = 23.315$, $p < .0001$, $\eta^2 = 0.493$): Children made relatively fewer *intrusion* errors (mean *intrusion* errors: 5.6, $SD = 3.36$; mean *confusion* errors = 2.0, $SD = 1.8$), in comparison to Parents (mean *intrusion* errors = 32, $SD = 28.0$; mean *confusion* errors = 9.6, $SD = 10.6$) and Young Adults (mean *intrusion* errors = 44, $SD = 33.0$; mean *confusion* errors = 10.9, $SD = 8.8$). A repeated measures ANOVA showed no significant difference in error patterns between the two adult groups ($F(1, 24) < 1$).

In addition, we investigated sequencing errors alone. A one-way ANOVA showed no significant difference between the mean number of sequencing errors made by each group: $F(2, 37) = 3.128$, $p = .056$ (Children: mean = 1.28, $SD = 1.2$; Parents: mean = 3.23, $SD = 3.05$; Young Adults: mean = 1.76, $SD = 1.58$).

Word list recall

The three groups' performance on the word list recall was statistically indistinguishable (see Figure 2). Data were analyzed using a univariate ANOVA, with between-subject variables of group and list placement. There was no significant difference among the groups or list placement, and there was no significant interaction between factors (all Fs < 1; Children: mean = 5.71, $SD = 1.64$; Parents: mean = 6.00, $SD = 1.73$; Young Adults: mean = 5.85, $SD = 1.63$).

We next examined the effect of meaningfulness on *list recall*. A repeated measures mixed-type ANOVA was conducted with meaningfulness (*meaningful* vs. *nonsense* words) as within-subject variables and group as a between-subject variable. (Since there had been no effect of the list placement, this factor was not used.) The main effect of group was not significant ($F(2, 37) < 1$; $p = .906$), nor was there a significant effect of meaningfulness on performance ($F(1, 39) < 1$; $p = .396$), but there was a significant interaction between the factors ($F(2, 37) = 7.760$, $p = .002$; $\eta^2 = 0.296$). Post-hoc tests, however, did not find significant differences for word list types (meaningful vs. nonsense) between groups (Bonferroni-corrected t-tests, $p = 1.00$, for each comparison).

Gist recall

All three groups performed similarly for gist recall, with mean scores (out of 6) of 5.57 ($SD = 0.852$), 5.53 ($SD = 0.77$), and 5.85 ($SD = 0.376$) for Children, Parents, and Young Adults, respectively. The effect of group was not significant ($F(2) < 1$). Gist recall scores were not normally distributed (Kolgomorov-Smirnov test: 0.467, $df = 40$, $p = .0001$) and were effectively at ceiling, and so primarily confirm that there were no gross lapses in effort, retention, or adherence to our protocol.

Effect of word position on verse recall

We also looked at the relationship between *verse recall* and the position of a word within a line of the verse

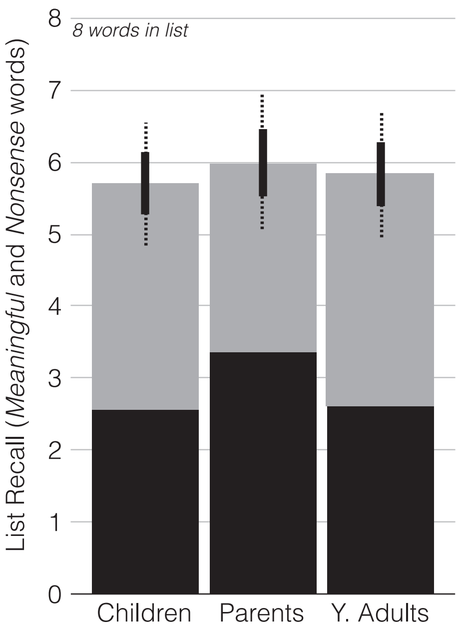


Figure 2 *Average number of words correctly recalled in the word lists. Word list recall for the four meaningful (top, gray area of stacked bar) and four nonsense words (black area) for each group. Solid error bars indicate standard error; dotted, 95% confidence intervals.*

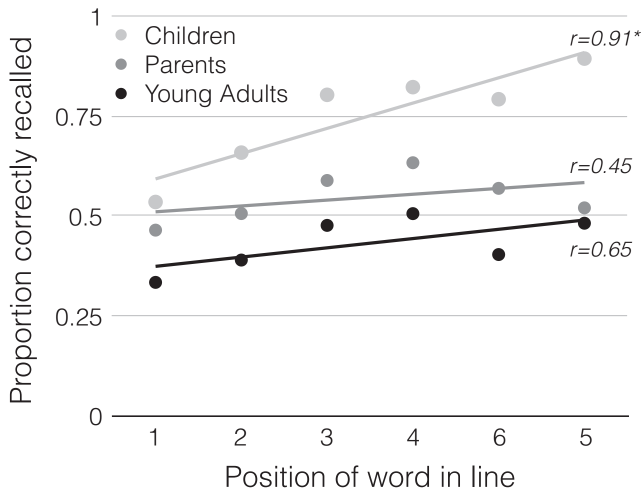


Figure 3 *Proportion of times a word was correctly recalled as a function of its position in the line, for all groups. Linear fits are shown with corresponding r values. Data are collapsed over the lines of the verse (e.g. the data points shown in position 1 represent the average recall rate over the first-position words from all 40 lines of the verse). Constraints on rhythm and rhyme increase with line position.*

(Figure 3). All groups showed a trend for better recall of words that appear later in a line; a correlation that was especially pronounced, and significant, in Children (r = 0.91, p = .013; r = 0.45, p = .375; and r = 0.65, p = .161, for Children, Parents, and Young Adults, respectively).

Discussion

In this study, parents read a novel rhyming children's poem as their 4-year-old's bedtime story for ten consecutive days. A group of young adults passively listened to the same verse for ten consecutive nights, simulating children's exposure. Following this, we measured participants' verbatim free-recall. In contrast to results in other memory domains, children significantly outperformed both adult groups. When memory was tested on a random word list embedded in the verse, and for the gist of the verse, children and adults performed similarly; children's verbatim memory advantage was for verse per se.

Engagement during reading

Given our own lapses in mindfulness, it is tempting to suppose that adult participants may have drifted into a distracted state while reading (parents) or listening (young adults), resulting in poorer encoding than the assumedly more engaged children. This is unlikely for a few reasons. To begin with, we specifically chose a verse that, while suitable for children, was short (~1.5 minutes to read out loud), complex, and entertaining enough to be of sustained interest to adults. It is worth remembering too that participants in both adult groups knew they were taking part in an experiment, and that they would be tested on their recall of the verse (children had no such knowledge); a salient motivator, especially for the young adult group of university students. During debriefing, we asked about lapses in engagement and none were reported.

More directly though, we can look to the magnitude and ubiquity of the children's advantage. This is no small effect: children remembered nearly twice as many words as adults did, with far fewer errors. This main effect finds confirmation both in a pilot study we ran that pitted 4-year-olds against their parents, and a recent follow-up with another (N = 10) Young Adult group (which only achieved a mean *verse recall* of 56 words; for further discussion, please see Supplementary Materials). To assess the ubiquity of this pattern, we ranked 50 participants (14 children and 26 adults from our main study plus the 10 from the follow-up) by *verse recall*, and found that the top three performers were exclusively 4-year-olds (with nearly all, 12 of 14, having above median performance) and the bottom 22 were exclusively adults.

But most directly, the integrated word list (see *Materials*) was included as a control to gauge the effect of general factors – like engagement – on performance. Since the word list was embedded in the verse, then general factors should affect memory for the word list as well. Simply put, if adults were simply more distracted than children, they should have done more poorly than children on the verse *and* the word list. However, adults did not show poorer memory on these control measures; their relatively poor performance was confined to the verse. The explanation for these results then cannot appeal to such general factors, but must be verse specific.

Implications of the verbatim verse advantage in children

The current results reconcile the apparent conflict between the anecdotal evidence that children have excellent memory for verse with the research that shows that less information is retained when presented in verse: children have excellent *verbatim* memory for verse (that may come at the expense of content). To evaluate this, we looked for evidence in the pattern of performance within the verse. Specifically, children's advantage should be especially evident for words that appear in later positions in a particular line, since these later words are more constrained by the rhythm, and especially the rhyme, of preceding words (e.g. later words are more likely to be the final, rhyming words). To examine this, we looked at the relationship between *verse recall* and the position of the word in the line. While all groups tended to have better recall for words that appear later in a line, the trend was much stronger, and significant, in children. Put together, this helps restore the promise of verse as an educational tool: if to-be-learned material is coded *verbatim* in a verse, with the help of rhyme as a constraining literary device, as in the alphabet song or when, say, introducing new vocabulary for animal names (see the findings of Read *et al*., 2014) children should readily retain it, perhaps better than their teachers.

In conclusion, an oral tradition in children

Testing how well children remember a novel children's verse presents a fortuitous coincidence, like testing memory for novel ballads in balladeers. As has been recognized by cultural anthropologists for many years, children form a dispersed, but interconnected community with a distinct oral culture (Opie & Opie, 1959). We argue that children are better than adults at recalling verse because they exercise the skill more in order to participate in the transmission of their culture through songs and stories, poems and taunts (consider that '*Eenie, meenie, miney, mo*', the most ubiquitous counting-out rhyme in the English-speaking world, has been transmitted faithfully for over a century (Rubin, Ciobanu & Langston, 1997)). Over development, practice diminishes due to shifts in culture, and the availability of external, written memory and memory processes introduced by and related to literacy. (Further research is required to determine whether the difference between adults and children is even deeper: perhaps children's advantage is not just a matter of better-practiced verbatim memory for verse, but reflects the deployment of specialized mechanisms, e.g. tuned to prosody (Nelson, Hirsh-Pasek, Jusczyk & Cassidy, 1989), that subserve early language learning (Rubin, 1995)). Echoing Socrates, Merlin Donald (2010) suggested that, 'teaching children to read and write, or training them in the use of any exographic system, including those employed in music and mathematics, proved to change the mnemonic strategies they use' (p. 73). Children then, we argue, form a preliterate society, hidden in plain sight, similarly reliant on a well-practiced memory for rhythm and rhyme for the oral transmission of their culture.

Acknowledgements

We would like to thank Krisztina Peres and other members of the Baby Lab at Eötvös Loránd University for their help in data collection. This research was supported by a Hungarian Eötvös Scholarship awarded to KI and by National Institutes of Health Grant 2R15EY017985 awarded to ZK and EB. The European Union and the European Social Fund have provided financial support to the project under Grant agreement TÁMOP 4.2.1/B-09/1/KMR-2010-0003.

References

Barnat, S.B., Klein, P.J., & Meltzoff, A.N. (1996). Deferred imitation across changes in context and object: memory and generalization in 14-month-old infants. *Infant Behavior and Development*, **19**, 241–251.

Bauer, P.J., & Dow, G.A. (1994). Episodic memory in 16- and 20-month-old children: specifics are generalized but not forgotten. *Developmental Psychology*, **30**, 403–417.

Bauer, P.J., Wenner, J.A., Dropik, P.L., & Wewerka, S.S. (2000). Parameters of remembering and forgetting in the transition from infancy to early childhood. *Monographs of the Society for Research in Child Development*, **65**, Serial No. 263.

Beardsworth, E., & Bishop, D. (1994). Assessment of long-term verbal memory in children. *Memory*, **2**, 129–148.

Bower, G.H., & Bolton, L.S. (1969). Why are rhymes easy to learn? *Journal of Experimental Psychology*, **82** (3), 453–461.

Brainerd, C.J., & Gordon, L.L. (1994). Development of verbatim and gist memory for numbers. *Developmental Psychology*, **30**, 163–177.

Brown, A.L. (1975). The development of memory: knowing, knowing about knowing, and knowing how to know. In H.W. Reese (Ed.), *Advances in child development and behavior* (Vol. 10, pp. 103–152). San Diego, CA: Academic Press.
Calvert, S.L. (2001). Impact of televised songs on children's and young adults' memory of educational content. *Media Psychology*, **3**, 325–342.
Calvert, S.L., & Billingsley, R.L. (1998). Young children's recitation and comprehension of information presented by songs. *Journal of Applied Developmental Psychology*, **19**, 97–108.
Calvert, S.L., & Tart, M. (1993). Song versus prose forms for students' very long-term, long-term, and short-term verbatim recall. *Journal of Applied Developmental Psychology*, **14**, 245–260.
Chase, W.G., & Simon, H.A. (1973). Perception in chess. *Cognitive Psychology*, **4**, 55–61.
Chi, M.T.H. (1978). Knowledge structures and memory development. In R. Siegler (Ed.), *Children's thinking: What develops* (pp. 73–96). Hillsdale, NJ: Erlbaum.
Csányi, I. (1976). A Peabody-Szókincsvizsgálat hazai alkalmazásának elsõ tapasztalatai siket és nagyothalló gyermekeknél. *Magyar Pszichológiai Szemle [Hungarian Psychological Review]*, **3**, 242–260.
Donald, M. (2010). The exographic revolution: neuropsychological sequelae. In L. Malafouris & C. Renfrew (Eds.), *The cognitive life of things: Recasting the boundaries of the mind* (pp. 71–79). Cambridge: McDonald Institute Monographs.
Ericsson, K.A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review*, **102**, 211–245.
Gathercole, S.E. (1998). The development of memory. *Journal of Child Psychology and Psychiatry*, **39**, 3–27.
Gergely, G., Bekkering, H., & Király, I. (2002). Rational imitation in preverbal infants. *Nature*, **415**, 755.
Goody, J. (1998). Memory in oral and literate traditions. In P. Fara & K. Patterson (Eds.), *Memory* (pp. 73–94). Cambridge: Cambridge University Press.
Halperin, J.M., Healey, J.M., Zeitchik, E., & Ludman, W.L. (1989). Developmental aspects of linguistic and mnestic abilities in normal children. *Journal of Clinical and Experimental Neuropsychology*, **11** (4), 518–528.
Hayes, D.S. (1999). Young children's exposure to rhyming and non-rhyming stories: a structural analysis of recall. *Journal of Genetic Psychology*, **160**, 280–293.
Hayes, D.S., Chemelski, B.E., & Palmer, M. (1982). Nursery rhymes and prose passages: preschoolers' liking and short-term retention of story events. *Developmental Psychology*, **18**, 49–56.
Hayne, H., Boniface, J., & Barr, R. (2000). The development of declarative memory in human infants: age-related changes in deferred imitation. *Behavioral Neuroscience*, **114**, 77–83.
Heil, M., & Jansen, P. (2008). Aspects of code-specific memory development. *Current Psychology*, **27**, 162–168.
Király, I. (2009). Memories for events in infants: goal relevant action coding. In T. Striano & V. Reid (Eds.), *Social cognition: Development, neuroscience and autism* (pp. 113–128). London: Wiley-Blackwell.
Nelson, D.G., Hirsh-Pasek, K., Jusczyk, P.W., & Cassidy, K.W. (1989). How the prosodic cues in motherese might assist language learning. *Journal of Child Language*, **16**, 55–68.
Opie, I., & Opie, P. (1959). *The lore and language of schoolchildren*. Oxford: Oxford University Press.
Plato (1997). *Complete works* (Ed. by J.M. Cooper, translated by A. Nehamas & P. Woodruff). Indianapolis, IN: Hackett Publishing Co.
Read, K. (1976). *The nursery school: Human relationships and learning*. Philadelphia, PA: Saunders.
Read, K., Macauley, M., & Furay, E. (2014). The Seuss boost: Rhyme helps children retain words from shared storybook reading. *First Language*, **34**, 354–371.
Reyna, V.F., & Brainerd, C.J. (1995). Fuzzy-trace theory: an interim synthesis. *Learning and Individual Differences*, **7**, 1–75.
Reyna, V.F., & Kiernan, B. (1994). Development of gist versus verbatim memory in sentence recognition: effects of lexical familiarity, semantic content, encoding instructions, and retention interval. *Developmental Psychology*, **30**, 178–191.
Rubin, D.C. (1995). *Memory in oral traditions: The cognitive psychology of epic, ballads and counting-out rhymes*. Oxford: Oxford University Press.
Rubin, D.C., Ciobanu, V., & Langston, W. (1997). Children's memory for counting-out rhymes: a cross-language comparison. *Psychonomic Bulletin & Review*, **4**, 421–424.
Rubin, D.C., & Wallace, W.T. (1989). Rhyme and reason: analyses of dual retrieval cues. *Journal of Experimental Psychology: Learning, Memory and Cognition*, **15**, 698–709.
Rubin, D.C., Wallace, W.T., & Houston, B.C. (1993). The beginnings of expertise for ballads. *Cognitive Science*, **17**, 435–462.
Sachs, J.S. (1967). Recognition memory for syntactic and semantic aspects of connected discourse. *Perception & Psychophysics*, **2**, 437–443.
Schneider, W., Gruber, H., Gold, A., & Opwis, K. (1993). Chess expertise and memory for chess positions in children and adults. *Journal of Experimental Child Psychology*, **56**, 328–349.
Sheingold, K., & Foundas, A. (1978). Rhymes for some reasons: effect of rhyme on children's memory for detail and sequence in simple narratives. *Psychological Reports*, **43**, 1231–1234.
Simcock, G., & Hayne, H. (2003). Age-related changes in verbal and non-verbal memory during early childhood. *Developmental Psychology*, **39**, 805–814.
Tillmann, B., & Dowling, W.J. (2007). Memory decreases for prose, but not for poetry. *Memory & Cognition*, **35**, 628–639.

Received: 28 June 2015
Accepted: 4 November 2015

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Data S1. Additional supplemental data, analyses and stimuli.
Video S1. Verse recall by a child: Example video.

III. NAVIGATING IN THE SOCIAL WORLD: NAÏVE PSYCHOLOGY

Sociality is rooted in the capability that humans can *learn from and about a partner* as well in the course of communication. In the process of learning, the communicative partner can play two separate roles: could be the main source of information, and also the target of observation and learning. We would like to highlight that from very early on, children are able to exploit both roles of the partner in order to enrich their knowledge on different fields of the environment in an integrated format.

3.1 Thesis 7. In their early months, children primarily learn from others and encode the content learnt as predictive generic information.

Gergely, G., Egyed, K., **Király, I.** (2007): On pedagogy. *Developmental Science* 10 (1), 139-146.

Egyed, K., **Király, I**., Gergely, Gy (2013): Communicating Shared Knowledge in Infancy, *Psychological Science*, Vol.24, No 7. pp. 1348-1353. DOI: 10.1177/0956797612471952

3.2 Thesis 8. The availability of both the object centered and the person centered interpretative schemas when interpreting another person's behavior invites the reconsideration of findings on theory-of mind competencies in young infants: we propose the primacy of the object-centered approach In addition, we postulate that the emergence of flexible use of interpretative models might contribute to a more and more elaborate performance.

Kampis D, Somogyi E, Itakura S, **Király I.** (2013): Do infants bind mental states to agents?, *Cognition 129*: (2) pp. 232-240.

Király, I. Oláh, K., Csibra, G., Kovács, Á. (in preparation): Retrospective attribution of false beliefs in 3-year-old children.

3.3 Thesis 9. The primary function of theory of mind is to enable the observer to monitor in real time, say spontaneously and prospectively, the knowledge state of the partner. This capability facilitates both learning about the partner in the here and know and also evaluating the possibility whether learning from the partner would contribute to a valuable shared representational space for the long term or not

Elekes, F., Varga, M. **Király, I.** (2016). Evidence for spontaneous level-2 perspective taking in adults. *Consciousness and Cognition 41*. 93–103.

Elekes, F., Varga, M**., Király, I**. (2017). Level-2 perspective is computed quickly and spontaneously: Evidence from eight to 9.5-year-old children. *British Journal of Developmental Psychology*. DOI: 10.1111/bjdp.12201

Developmental Science 10:1 (2007), pp 139–146 DOI: 10.1111/j.1467-7687.2007.00576.x

On pedagogy

György Gergely, Katalin Egyed and Ildikó Király

Institute for Psychological Research, Hungarian Academy of Sciences, Hungary

Abstract

Humans are adapted to spontaneously transfer relevant cultural knowledge to conspecifics and to fast-learn the contents of such teaching through a human-specific social learning system called 'pedagogy' (Csibra & Gergely, 2006). Pedagogical knowledge transfer is triggered by specific communicative cues (such as eye-contact, contingent reactivity, the prosodic pattern of 'motherese', and being addressed by one's own name). Infants show special sensitivity to such 'ostensive' cues that signal the teacher's communicative intention to manifest new and relevant knowledge about a referent object. Pedagogy offers a novel functional perspective to interpret a variety of early emerging triadic communicative interactions between adults and infants about novel objects they are jointly attending to. The currently dominant interpretation of such triadic communications (mindreading) holds that infants interpret others' object-directed manifestations in terms of subjective mental states (such as emotions, dispositions, or intentions) that they attribute to the other person's mind. We contrast the pedagogical versus the mindreading account in a new study testing 14-month-olds' interpretation of others' object-directed emotion expressions observed in a communicative cueing context. We end by discussing the far-reaching implications of the pedagogical perspective for a wide range of early social-cognitive competences, and for providing new directions for future research on child development.

Introduction: Early mindreading versus pedagogical knowledge transfer

More than 20 years have passed since the by now classical demonstrations (Wimmer & Perner, 1983) that around 4 years of age children start to exhibit explicit 'mindreading' skills. This is evidenced by their verbal ability to predict and justify others' false belief based actions by attributing causal intentional mental states to them (such as desires, intentions, and beliefs). Since then research on early social cognitive development has been preoccupied with searching for the ontogenetic origins and earliest forms of attributing mental states to others (Leslie, 1987; Perner, 1991; Repacholi & Gopnik, 1997; Trevarthen & Aitken, 2001; Carpenter, Nagell & Tomasello, 1998; Tomasello, Carpenter, Call, Behne & Moll, 2005; Moses, Baldwin, Rosicky & Tidball, 2001). This approach has led to important discoveries through the use of non-verbal violation-of-expectation looking time methodologies suggesting that already 15-month-olds may possess an implicit capacity to infer others' intentional mental states even when these represent a counterfactual state of affairs (false beliefs; Onishi & Baillargeon, 2005) or fictional (mentally stipulated imaginary) representations of reality (as in understanding pretense actions; Onishi, Baillargeon & Leslie, in press).

Recent research on social referencing[1] (Moses *et al.*, 2001; Mumme & Fernald, 2003), 'proto-declarative' pointing[2] (Carpenter *et al.*, 1998; Liszkowski, 2006) or predicting others' object-directed actions (Phillips, Wellman & Spelke, 2002; Sodian & Thoermer, 2004) has converged on the mentalistic view that by 12 to 14 months of age, based on others' emotion expressions directed at objects identified by referential cues (such as direction of gaze, or pointing), infants attribute intentional mental states to others such as emotional attitudes, desires, or dispositions about referent objects. It has been argued that infants can 'recognize the central role that such internal states play in others' behavior' (Moses *et al.*, 2001, p. 733) and rely on them to predict others' object-directed actions (Phillips *et al.*, 2002), to 'share' their own mental attitudes towards the referent with those of others (Tomasello *et al.*, 2005), or to modulate their own object-directed behaviours based on the other's emotional attitude toward the object (Mumme & Fernald, 2003).

[1] One-year-olds' ability to seek out and rely on others' object-directed emotion expressions to modulate their own behaviour towards novel objects.

[2] Infants' pointing to direct an adult's attention to novel objects and elicit commentary and joint communicative interactions from them about it.

Address for correspondence: György Gergely, Institute for Psychological Research, Hungarian Academy of Sciences, 18–22 Victor Hugo Street, H-1132. Budapest, Hungary; e-mail: gergelyg@mtapi.hu

Arguably, however, the enthusiastic search for early forms of intersubjective understanding of minds had an additional undesirable effect of sometimes too hastily embracing mentalistic interpretations for early social cognitive phenomena (including social referencing, imitative learning, facial and vocal interactions that have a turn-taking 'proto-conversational' structural organization, proto-declarative pointing, or predicting others' object-directed actions) at the expense of exploring alternative functional explanations that do not necessarily involve or rely on infants' capacity to attribute mental states (Gergely, 2002; Csibra & Gergely, 2006). Elsewhere we have proposed such alternatives – namely, the infant's 'teleological stance'[3] (Gergely & Csibra, 2003) or the system of human 'pedagogy' (Csibra & Gergely, 2006) – that, in our view, represent a novel perspective on the functional nature and underlying mechanisms of several early social cognitive capacities that are currently standardly considered as involving early forms of mindreading.

In particular, the theory of *human pedagogy* (Csibra & Gergely, 2006; Gergely & Csibra, 2005, 2006) proposes that many types of early emerging triadic communications about referent objects are best conceived as serving the primarily *epistemic* function of actively seeking out and cooperatively providing reliable, new and relevant information by knowledgeable adults to ignorant infants about the generalizable properties of referent objects and their kinds that constitute universally shared cultural knowledge to be fast-learned by infants (such as the object's name, proper function, manner of use, or valence qualities). There is evidence (see Csibra & Gergely, 2006, for a review) that human infants actively seek out and show early sensitivity, orientation, and preference for specific types of communicative cues (such as eye-contact, eyebrow raising, turn-taking contingent reactivity, motherese, or being addressed by their name) that typically accompany triadic interactions about referents. According to pedagogy theory, infants are adapted to automatically interpret such cues as 'ostensive' signals (cf. Sperber & Wilson, 1986) indicating the other's overt communicative intention to manifest new and relevant information 'for' them to acquire about the object that is identified by non-verbal referential cues (such as gaze-direction or pointing). We hypothesize that ostensive cues constrain and direct infants' interpretation of adults' object-directed behavioural manifestations (such as their object-referential emotion expressions, verbal labelling, demonstrations of the functional properties of objects, or specific manners of artifact use) as conveying to them new and relevant knowledge about the referent that they need to extract and bind to its representation as its essential property.

Furthermore, such pedagogical manifestations are interpreted to convey information that is generalizable to the object class that the referent belongs to and is assumed to be part of universally shared cultural knowledge about the object kind. Therefore, pedagogy theory predicts that when another person's object-directed behavioural manifestations are observed in an ostensive cuing context, infants will *not* interpret the content of such manifestations as expressing the specific subjective mental states that the other holds about the referent, but rather they will use such communicative displays as the basis to infer the new information about the relevant properties of the referent object that they are being taught about.

Learning 'about' versus learning 'from' other minds: the role of ostensive cuing in triggering pedagogical information transfer

Below we shall directly contrast the alternative explanatory perspectives of the pedagogical versus the standard mindreading account by comparing their differential predictions about how young infants interpret others' object-directed emotion displays in *social referencing* situations, i.e. when infants seek out and use others' emotion expressions to modulate their own behaviour towards novel and ambiguous objects. The mindreading account assumes three crucial steps. First, from the other's object-directed emotion display (say, fear or interest/joy) infants infer what specific emotional or dispositional mental state the other holds towards the object (being afraid vs. liking). Second, the infant predicts from this attributed mental state the type of action the other could be expected to perform toward the referent (approaching or avoiding it). Third, infants rely on the mental attitude attributed to the other and/or the action prediction derived from it to modulate their *own* behaviour towards the object.

A potential problem for this interpretation stems from its under-determination by the observable evidence, as others' object-referential emotion expressions can be compatible with qualitatively different interpretations. One possibility is, of course, the '*person-centred' interpretation* assumed by the mindreading account, namely, that infants indeed interpret the other's emotion display

[3] Young infants' capacity to represent actions as goal-directed and agents as rational who are expected to pursue their goals in the most efficient manner available under the physical constraints of the situation. The existence of this teleological system specialized for action interpretation in terms of goals and efficiency is based on non-verbal violation-of-expectation looking time studies (e.g. Csibra *et al.*, 2003; Gergely *et al.*, 1995) showing that 1-year-olds can productively infer and attribute goals to agents and interpret their behaviour as goal-directed actions on condition that their goal approach satisfies the assumption of efficiency.

as expressing the individual's person-specific subjective mental attitude toward the referent (e.g. that 'Alison *likes* broccoli'). Note, however, that the same emotion expression could be equally compatible with an *'object-centred' interpretation* as it could convey new information about some relevant property of the referenced object (e.g. that 'broccoli is *good*'). If infants adopted such an object-centred interpretation by construing the other's emotion manifestation as communicating relevant information about the valence qualities of the object, they could directly encode this new information by binding it to their representation of the referent. They could then access the relevant contents of their newly formed representation (that would now include information about the object's positive valence) to modulate their behaviour towards it (e.g. to approach rather than avoid it). Clearly, this way infants could succeed in social referencing without necessarily attributing or relying on the other's subjective mental attitude towards the referent.

This raises two questions: 1. Do infants interpret others' object-directed emotion displays by setting up 'person-centred' or 'object-centred' representations? 2. Do infants predict others' object-directed behaviours from their representation of the individual's person-specific subjective mental attitude towards the referent, or do they base their action predictions on the objective valence qualities of the referent that they have come to represent through their 'object-centred' interpretation of others' referential emotion displays?

The pedagogical approach proposes that during social referencing infants set up and rely on 'object-centred' interpretations of others' referential emotion expressions. This should be so because social referencing interactions involve salient ostensive and referential cues of triadic communication (such as eye-contact, turn-taking looking back and forth between the object and the infant, using motherese to address the infant and while commenting about the object, etc.). As hypothesized above (Csibra & Gergely, 2006), such ostensive cues identify the situation as a case of pedagogical knowledge transfer for the infants, triggering the interpretation that the other exhibits a communicative intention addressed to them to manifest new and relevant information for them to fast-learn about the referent.

Pedagogy theory also assumes (Csibra & Gergely, 2006; Gergely & Csibra, 2006) that ostensive cues trigger in-built assumptions in infants about the *generalizability* and *universality* of the epistemic information that the other's communicative manifestations convey about the referent. For social referencing this predicts that infants assume that the other's object-directed emotion manifestations convey universally shared information about the referent that is available to *all* individuals. We hypothesize, therefore, that infants rely on their 'object-centred' interpretations to form generalized expectations that all others (and not only the specific person manifesting the emotion to them) will perform the same kind of object-directed actions that are appropriate and rational given the objective valence quality of the referent that the infant's newly formed object representation contains.

Below we report a study with 14-month-olds designed to test the contrastive predictions of the pedagogical vs. mindreading account using a violation-of-expectation looking time procedure. Two subject groups watched different series of familiarization events in which two demonstrators repeatedly presented ostensively cued object-directed emotion manifestations of different valence towards two novel objects. Both demonstrators were *consistent* in manifesting over trials the same (positive vs. negative) emotions towards the referents, but they always expressed the *opposite* emotion towards the same target than the other demonstrator.

The familiarization series presented to the two subject groups differed in the relative frequency with which the demonstrators appeared across trials, and in the overall number of events. In the *'Symmetric' condition* both demonstrators appeared with equal frequency. In the *'Asymmetric' condition* the 'frequent' demonstrator appeared three times more often than the 'infrequent' demonstrator.

These familiarization series were followed by four 'object-directed action' test trials in which each demonstrator appeared twice, choosing alternately one or the other target to act on. Thus, both demonstrators performed one *'attitude-consistent' object-choice* (acting on the object towards which they consistently expressed positive emotion during familiarization), and one *'attitude-inconsistent' object-choice* (choosing the object towards which they had expressed a negative emotion).

We can derive differential predictions for the two conditions from the mindreading vs. pedagogical account. According to the mindreading account, during familiarization infants attribute to each demonstrator different (in fact, opposite) person-specific mental attitudes of 'liking' vs. 'disliking' the two targets. From these, infants are assumed to generate opposite expectations as to which referent the demonstrators will choose during test trials: expecting both to make person-specific 'attitude-consistent' object-choices. Thus, longer looking times are predicted for the (unexpected) 'attitude-*in*consistent' object-choices for *both* conditions.

In contrast, the pedagogical account generates different predictions for the two conditions. In the Symmetric condition both targets are manifested to have positive vs. negative valence equally often by the two demonstrators. Therefore, binding the valence value of each emotion manifestation to the 'object-centred' representation of its referent should not change the perceived valence of the two objects (that were initially novel and equally neutral for the infants). Since according to the universality assumption, infants assume *all* individuals (including both demonstrators) to have access to the objective valence information manifested for them about the referents, they will expect all others' object-directed choices and actions to be similarly determined by the objects' valence qualities. Therefore, no differential looking times are predicted for alternative object-choices in this condition (irrespective of whether the object-choices are 'attitude-consistent' or 'attitude-inconsistent').

In the 'Asymmetric' condition, however, infants' valence representations of the referents should be differentially modified during familiarization trials due to the unequal frequency of the two demonstrators' opposite-valued valence manifestations. The referent manifested to have positive valence more frequently across trials will become represented as 'good or better' than the other object that was manifested more frequently to have negative valence. Therefore, the pedagogical account predicts a valence-based object-choice (of the 'better' object) for *both* demonstrators (irrespective of whether their object-choice is 'attitude-consistent' or 'attitude-inconsistent').

Method

Participants

Sixty-four 14-month-olds participated in the experiment. Thirty-two were assigned to the Symmetric (21 male, 11 female, mean age: 422 days, range: 406–440 days) and 32 to the Asymmetric condition (17 male, 15 female, mean age: 423 days, range: 409–437 days). An additional 41 infants were excluded due to technical problems (2), fussiness (26), or maternal interference (12).

Stimuli

All familiarization and test events were videotaped. In each event one (of two) female demonstrators appeared in the middle of the screen facing the infant. The demonstrator sat behind a table with two different objects placed in front of her on the left vs. right sides of the table. Demonstrator 1 was a long-haired brunette, Demonstrator 2 was short-haired and blond. The targets were two wooden objects of equal size (about 5 cm high) but they differed in shape and colour (Object A: red ball; Object B: yellow cube). Both objects were unfamiliar to the infants who showed no differential preference for either. In all events Object A was on the left, while object B was on the right side of the table.

Familiarization events

Each event contained the same action sequence: First, the demonstrator 'greeted' the infant manifesting ostensive-communicative cues (slightly tilting her head forward, looking and smiling at the baby while 'knowingly' raising her eyebrows). Then she turned to the left to gaze at Object A, displaying always the same (either positive: 'interest/joy' or negative: 'disgust') emotion towards it. She then turned back to the middle and looked at the infant again. Then she turned to the right to look at Object B, displaying always the other emotion of opposite valence than what she expressed towards Object A. Finally, she turned back to the middle looking at the infant again.

Across all familiarization trials Demonstrator 1 (the long-haired brunette) always displayed 'interest/joy' towards Object A and 'disgust' towards Object B, while Demonstrator 2 (the short-haired blond) always expressed 'disgust' towards Object A and 'interest/joy' towards Object B.

The familiarization series presented in the Symmetric vs. Asymmetric conditions differed in three respects: in the relative frequency and relative order of the two demonstrators' appearances across trials, and in the overall number of events. The 'Symmetric' condition consisted of six events in which the demonstrators appeared with equal frequency (three times each) across trials. These were presented in an ABABAB order for half of the subjects, while the other half saw the opposite BABABA sequence. The 'Asymmetric' condition consisted of 12 events across which Demonstrator 1 ('frequent person': FP) appeared nine times, while Demonstrator 2 ('infrequent person': IP) appeared only three times. Each of the 32 series started with FP. In half of them IP appeared in the 2nd, 7th, and 11th position. In the other 16 series IP appeared in the 2nd, 7th, and the final 12th position. Thus, across subjects FP and IP appeared equally often in the last position to control for possible recency effects.

Test phase

For both groups the familiarization phase was followed by four 'object-choice and object-directed action' test trials. Each demonstrator appeared twice across these events, always presenting first the same ostensive cues of 'greeting' as during familiarization. Then she turned to and chose either Object A or B to 'play with' fixating it with a neutral facial expression throughout. She grasped the chosen object, moved it to a new position (10 cm away), and then moved it back. This 'playing' action was repeated as long as the subject watched it. Across trials both demonstrators performed one *'attitude-consistent'* and one *'attitude-inconsistent' object-choice* (relative to their object-specific emotional attitude manifested during familiarization). The order of test events was counterbalanced across subjects.

Procedure

Infants sat on their parent's lap, 80 cm from a 21″ monitor. They were presented with the familiarization events in one block followed by the four test events. Their visual behaviour was recorded by a video-camera hidden above the monitor. An experimenter watched their looking behaviour on a monitor in an adjacent room and registered through a computer program the length of their visual fixations to each test event by pressing a key on a keyboard. Each test event lasted until the infant looked away for more than 2 seconds. A sound cue oriented the infant's attention back to the display before the next event started.

Results

We used the looking times during test trials as the sole dependent measure. A second experimenter re-coded off-line 25% of the video-records measuring subjects' looking times for the test events. The two coders' measurements showed significantly high correlation (Pearson: $r = .99$), indicating the reliability of the looking time data.

Figure 1 depicts the mean looking times to the different types of object-choices during test events for the Symmetric vs. Asymmetric conditions. To analyze the looking times, first we performed a repeated measures mixed ANOVA with Object (A vs. B) and Attitude-Consistency (Consistent vs. Inconsistent object-choice) as within-subject factors, and Condition (Symmetric vs. Asymmetric) as the between-subjects factor. This analysis yielded a significant main effect of Condition ($F(1, 62) = 7.243$, $p < .01$), which reflects the longer overall looking times for the test trials in the Asymmetric condition (Figure 1). We found a tendential main effect of Object ($F(1, 62) = 3.514$, $p < .07$) and a nearly significant Object × Condition interaction ($F(1, 62) = 3.685$, $p = .06$). The analysis also yielded a significant Attitude-Consistency × Condition interaction ($F(1, 62) = 4.30$, $p < .05$). Note that no main effect of Attitude-Consistency ($p = .555$) was present.

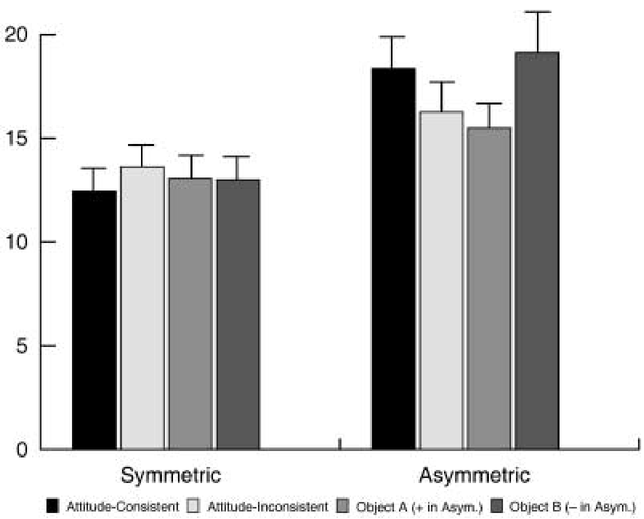


Figure 1 *Mean looking times to 'attitude-consistent' vs. 'attitude-inconsistent' object-choices and to choosing Object A vs. B in the Symmetric and Asymmetric familiarization conditions.*

To resolve the interactions, we ran separate repeated measures ANOVAs for the two conditions. In the Symmetric condition neither the effect of Object ($p = .971$) nor that of Attitude-Consistency ($p = .315$) approached significance and there was no Object × Attitude-Consistency interaction ($p = .935$) either. In contrast, a similar ANOVA for the Asymmetric condition yielded a significant main effect of Object ($F(1, 31) = 5.903$, $p = .021$). The main effect of Consistency did not reach significance ($F(1, 31) = 3.751$, $p < .07$), and the direction of difference was actually opposite (longer looking at the consistent events, Figure 1) to what the mindreading account would have predicted. We found no interaction between Object and Attitude-Consistency ($p = .521$). A non-parametric Sign test also confirmed the Object valence effect for the Asymmetric condition yielding a close to significant result ($z = -1.945$, $p = .052$).

Finally, we checked for any effect of (a) Person (Demonstrator 1 vs. 2) or (b) Order of test trials separately for the two conditions. While there was no Person effect in either the Symmetric ($p = .935$) or Asymmetric ($p = .521$) condition, the Symmetric condition yielded a significant Order effect ($F(3, 93) = 10.432$, $p < .001$), showing a continuous decrease in looking times from the first to the last test event (1st: 17.89 sec; 2nd: 12.14 sec; 3rd: 11.80 sec; 4th: 10.33 sec). No Order effect was present in the Asymmetric condition ($F(3, 93) = 1.978$, $p = .123$).

Discussion

First, as Figure 1 shows, the main prediction of the mindreading account that attitude-inconsistent object-choices, which violate infants' person-specific attitude-based expectations, should result in longer looking times in both conditions is clearly not supported by the results.

In accordance with the lack of an Attitude-Consistency main effect, there was no significant difference in looking times to attitude-consistent vs. attitude-inconsistent object-choices in either condition. In fact, the significant Attitude-Consistency × Condition interaction is, in part, dependent on a looking time difference in the Asymmetric condition that is in the opposite direction of that predicted by the mindreading perspective (Figure 1). In contrast, these findings are in line with the pedagogical account that predicted no difference for attitude-consistent vs. attitude-inconsistent object-choices for either condition.

The Object effects provide further support for the pedagogical approach. As shown in Figure 1, the tendential main effect of Object together with the nearly significant Object × Condition interaction stems entirely from the longer looking times for choosing Object B over A in the *Asymmetric condition only*.[4] This is exactly what was predicted by the pedagogical approach, according to which the 'object-centred' valence representation of Object B should have become more negative (while Object A was more positive) during familiarization in the Asymmetric but not in the Symmetric condition. Based on the universality assumption, the pedagogical approach predicted that infants in the Asymmetric condition will develop a generalized expectation that all others (including both demonstrators) should choose the more positive-valenced Object A: an expectation whose violation led to longer looking times when the more negative-valenced Object B was chosen in the Asymmetric condition.

The significant Object-Valence main effect (p = .021) in the Asymmetric condition provides further support for this interpretation showing that, as predicted, choosing the more negative-valenced Object B (M = 19.15 sec) led to significantly longer looking times than choosing the more positive-valenced Object A (M = 15.50 sec) (Figure 1). Note that this valence-based effect cannot be accounted for by the mindreading account that predicts longer looking times only for the attitude-inconsistent object-choices.

The significant Order effect present only in the Symmetric condition is also in line with the pedagogical interpretation that infants develop no specific expectations in this condition as to the type of object-choices either on the basis of object-valence (that remains equal for the two objects), or on the basis of person-specific referential attitude (which was either not attributed, or not used to predict the person's object-choice). Thus, following their initial recovery of attention upon seeing the novel action in the first test event, lacking any specific expectations concerning the type of object-choice, infants showed a steady decrease of interest during the repeated presentations of the same action type.

Some important questions, however, remain unanswered by these results. For example, how should one interpret the apparent lack of infants' attributing person-specific mental attitudes from observing others' object-directed emotion expressions? According to the 'theory-theory' view (Perner, 1991; Gopnik & Wellman, 1994), this may indicate that 14-month-olds have not yet acquired the representational concept of 'desire' or cannot yet attribute such a mental state to others from observing their referential emotion expressions (Repacholi & Gopnik, 1997). We find this view implausible in light of recent evidence showing implicit false-belief attribution (Onishi & Baillargeon, 2005) and understanding violations of mentally stipulated fictional pretence scenarios (Onishi *et al.*, in press) already at 15 months.

In this regard, we emphasize that our theory of pedagogical knowledge transfer (Csibra & Gergely, 2006) does not propose that 14-month-olds are unable to attribute mental states to others. Rather, our proposal is that whether they do so or not may be directed by the presence of ostensive cues that trigger object-centred interpretations for communicative manifestations. We hypothesize that ostensive cues can act as an 'interpretation switch' directing infants to construe others' referential knowledge manifestations as pedagogical 'teaching' events. Accordingly, we suggest that in the present study it was the ostensive cuing context that biased infants to encode the manifested emotions as conveying information about the objective valence qualities of the referents, and not about the other's person-specific subjective attitude towards them.

Clearly, our results are compatible with but, in themselves, not sufficient to prove our 'interpretation switching' hypothesis as we did not directly manipulate the presence of ostensive cues in our familiarization conditions. We are currently running experiments with 14-month-olds to see if the withdrawal of ostensive cues would result in 'switching' infants' interpretive stance to set up person-centred mentalistic representations for others' object-directed emotion expressions. We do have, however, new and promising preliminary results from a modified 'object-requesting' paradigm (cf. Repacholi & Gopnik, 1997), suggesting that the hypothesized 'interpretation switching' role of ostensive cuing is, indeed, present in 18-month-olds (Egyed, Király & Gergely, in preparation). Finally, we can refer to further evidence supporting the hypothesized interpretation-modulating role of ostensive cuing that comes from our recent studies on imitative

[4] The nearly identical mean looking times for Objects A vs. B in the Symmetric condition also indicate that the infants had no inherent preference for either of the novel objects on which they could have based egocentrically determined expectations about others' likely object-choices (cf. Repacholi & Gopnik, 1997).

learning. These show that selective relevance-guided imitation of novel means (Gergely, Bekkering & Király, 2002) is triggered in 14-month-olds *only* in the presence of an ostensive cuing context (Gergely & Csibra, 2005).

General implications of pedagogy theory for future directions of research on child development

The human social and cultural environment represents two major and species-unique challenges for infants. On the one hand, young children must develop an understanding of others' minds to be able to predict and interpret their actions in terms of causal mental states attributed to them. On the other hand, infants must adapt to and acquire an immense amount of cultural knowledge that is – at least partially – cognitively 'opaque' to them as they involve arbitrary, conventional, and often apparently nonadaptive features of social belief systems and representational devices, and highly complex cultural artifacts with often hidden mechanistic and functional design properties. Recent theories of social cognitive development mainly focused on the first, challenge advancing alternative models to explain the ontogenetic development of young children's explicit representational understanding of others' minds and their mindreading skills to infer the specific intentional mental states that drive others' actions.

We believe that the theory of human pedagogy provides a new perspective for approaching both of the major challenges of social-cultural development outlined above. The primary focus of the theory is on how humans meet the second type of challenge that the need to learn cognitively opaque social and technological knowledge represents for the young learner who, arguably, could not acquire such knowledge through classical mechanisms of observational learning such as associative learning or 'blind' imitative copying (see Gergely & Csibra, 2006, for arguments). Human pedagogy is a species-specific social learning system that has evolved as a specialized adaptation to solve the learnability problem that the cognitive opacity of human cultural forms represents for the individual observational learner (Csibra & Gergely, 2006). This social learning mechanism of mutual design allows infants to seek out and rely on the communicative manifestations of relevant knowledge that knowledgeable conspecifics are spontaneously inclined to provide and tailor to meet the receptive requirements that the juvenile learner's often inadequate existing knowledge base and limited cognitive interpretive skills represent.

There are two built-in design features of pedagogy that, in our view, also shed new light on how to approach the first challenge that the development of understanding other minds presents. Fast and efficient cultural knowledge transfer is ensured by two implicit assumptions about communicating agents that human infants must possess. On the one hand, the ostensive cues of others that infants are biased to interpret as indicating their communicative intention to transfer relevant knowledge for them to fast-learn must also trigger a built-in assumption of '*basic epistemic trust*' in the other as a benevolent, cooperative, and reliable source of cultural information. This activates a 'presumption of relevance' about the contents of manifested knowledge as a result of which the young apprentice can fast-learn it without the need to test or critically scrutinize its validity or relevance any further.

The second built-in functional assumption concerns the presumption that the pedagogically manifested knowledge is generalizable (semantic) information about referent kinds that is universally shared by other (equivalently knowledgeable) members of the cultural community. This default assumption about communicative agents as sources of universally shared cultural knowledge amounts to an implicit construal of *others as equally 'omniscient' sources of shared knowledge*. This assumption of communicating others as possessing functionally equivalent omniscient minds with identical and universally shared knowledge contents may, in fact, constitute the biological basis of the epistemic illusion (sometimes referred to as 'the curse of knowledge'; Birch & Bloom, 2004) whereby young children – and to a lesser extent adults – are biased to believe that their own knowledge about the world is universally shared by all others (leading to a difficulty in conceptualizing other minds as having different knowledge contents).

In short, pedagogy theory suggests that much of early social-cognitive development establishing a realistic understanding of other minds involves learning about the specific conditions under which the built-in default assumption of universal knowledge and omniscient other minds must be suspended or inhibited. This gradual and experience-driven learning process eventually leads to the ability to draw correct inferences to attribute and represent the differential knowledge contents of separate and individual minds of others. A further developmental task is to acquire the specific conditions under which the other default assumption of pedagogy, the generalized 'epistemic trust' in communicating others, needs to be suspended or inhibited. Young children have to learn the hard way (that is, through accumulating experience) to differentiate trustworthy, benevolent, and reliable communicative sources of information from communicators who are unreliable, uninformed, or downright bad-intentioned providers of useless or deceiving information. At the same time, children must develop differentiated knowledge about the social division of epistemic labour in their culture to identify domain-specific experts and trustworthy institutionalized information outlets (such as professional rather than intuitive educators). They also have to learn to modulate their belief fixation mechanisms as a function of their developing understanding of the relative reliability of different sources of information. In summary, while previous research on social-cognitive development considered as its central task the need to account for how children come to understand that other people *have* minds, the new theoretical perspective offered by pedagogy theory turns this question upside-down, identifying as the central task for children's early development the need to come to understand that others have *separate minds with different knowledge contents*.

Acknowledgements

This research was supported by an EU6 Framework Programme grant (NEUROCOM, #12738), and by a grant from the Hungarian National Science Foundation (OTKA, K63808). We thank Gergely Csibra and John S. Watson for commenting on the first draft of this paper

References

Birch, S.A.J., & Bloom, P. (2004). Understanding children's and adults' limitations in mental state reasoning. *Trends in Cognitive Sciences*, **8** (6), 255–260.

Carpenter, M., Nagell, K., & Tomasello, M. (1998). Social cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monographs of the Society for Research in Child Development*, **63** (4), Serial No. 255.

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.

Csibra, G., & Gergely, G. (2006). Social learning and social cognition: the case for pedagogy. In Y. Munakata & M.H. Johnson (Eds.), *Processes of change in brain and cognitive development. Attention and performance* XXI (pp. 249–274). Oxford: Oxford University Press.

Egyed, K., Király, I., & Gergely, G. (in preparation). The role of ostensive cues in triggering 'object-centred' vs. 'person-centred' interpretations of referential emotion expressions of others.

Gergely, G. (2002). The development of an understanding of self and agency. In U. Goswami (Ed.), *Blackwell's handbook of childhood cognitive development* (pp. 26–46). Oxford: Blackwell.

Gergely, G., & Csibra, G. (2003). Teleological reasoning about actions: the naïve theory of rational action. *Trends in Cognitive Sciences*, **7**, 287–292.

Gergely, G., & Csibra, G. (2005). The social construction of the cultural mind: imitative learning as a mechanism of human pedagogy. *Interaction Studies*, **6**, 463–481.

Gergely, G., & Csibra, G. (2006). Sylvia's recipe: the role of imitation and pedagogy in the transmission of human culture. In N.J. Enfield & S.C. Levinson (Eds.), *Roots of human sociality: Culture, cognition, and human interaction* (pp. 229–255). Oxford: Berg Publishers.

Gergely, G., Bekkering, H., & Király, I. (2002). Rational imitation in preverbal infants. *Nature*, **415**, 755.

Gergely, G., Nádasdy, Z., Csibra, G., & Bíró, S. (1995). Taking the intentional stance at 12 months of age. *Cognition*, **56** (2), 165–193.

Gopnik, A., & Wellman, H.M. (1994). The theory theory. In L.A. Hirschfeld & S.A. Gelman (Eds.), *Mapping the mind: Domain specificity in cognition and culture* (pp. 257–293). New York: Cambridge University Press.

Leslie, A. (1987). Pretense and representation: the origins of a theory of mind. *Psychological Review*, **94**, 417–426.

Liszkowski, U. (2006). Infant pointing at twelve months: communicative goals, motives, and social-cognitive abilities. In N. Enfield & S. Levinson (Eds.), *Roots of human sociality: Culture, cognition, and interaction.* Oxford: Berg.

Moses, L.J., Baldwin, D.A., Rosicky, J.G., & Tidball, G. (2001). Evidence for referential understanding in the emotions domain at twelve and eighteen months. *Child Development*, **72** (3), 718–735.

Mumme, D.L., & Fernald, A. (2003). The infant as onlooker: learning from emotional reactions observed in a television scenario. *Child Development*, **74** (1), 221–237.

Onishi, K.H., & Baillargeon, R. (2005). 15-month-old infants understand false beliefs. *Science*, **308** (5719), 255–258.

Onishi, K.H., Baillargeon, R., & Leslie, A.M. (in press). 15-month-old infants detect violations in pretend scenarios. *Acta Psychologica*.

Perner, J. (1991). *Understanding the representational mind.* Cambridge, MA: MIT Press.

Phillips, A.T., Wellman, H.M., & Spelke, E.S. (2002). Infants' ability to connect gaze and emotional expression to intentional action. *Cognition*, **85**, 53–78.

Repacholi, B., & Gopnik, A. (1997). Early reasoning about desires: evidence from 14- and 18-month-olds. *Developmental Psychology*, **33** (1), 12–21.

Sodian, B., & Thoermer, C. (2004). Infants' understanding of looking, pointing, and reaching as cues to goal-directed action. *Journal of Cognition and Development*, **5** (3), 289–316.

Sperber, D., & Wilson, D. (1986). *Relevance: Communication and cognition*. Oxford: Blackwell.

Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: the origins of cultural cognition. *Behavioral and Brain Sciences*, **28**, 675–691.

Trevarthen, C., & Aitken, K.J. (2001). Infant intersubjectivity: research, theory, and clinical applications. *Journal of Child Psychology and Psychiatry*, **42** (1), 3–48.

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, **13**, 103–128.

Psychological Science OnlineFirst, published on May 29, 2013 as doi:10.1177/0956797612471952

Research Report

Psychological Science
XX(X) 1–6

DOI: 10.1177/0956797612471952
pss.sagepub.com

Communicating Shared Knowledge in Infancy

Katalin Egyed[1], Ildikó Király[2], and György Gergely[3]
[1]Department of Developmental and Clinical Child Psychology, Eötvös Loránd University; [2]Department of Cognitive Psychology, Eötvös Loránd University; and [3]Cognitive Development Center, Department of Cognitive Science, Central European University

Abstract
Object-directed emotion expressions provide two types of information: They can convey the expressers' person-specific subjective disposition toward objects, or they can be used communicatively as referential symbolic devices to convey culturally shared valence-related knowledge about referents that can be generalized to other individuals. By presenting object-directed emotion expressions in communicative versus noncommunicative contexts, we demonstrated that 18-month-olds can flexibly assign either a person-centered interpretation or an object-centered interpretation to referential emotion displays. When addressed by ostensive signals of communication, infants generalized their object-centered interpretation of the emotion display to other individuals as well, whereas in the noncommunicative emotion-expression context, they attributed to the emoting agent a person-specific subjective dispositional attitude without generalizing this attribution as relevant to other individuals. The findings indicate that, as proposed by natural pedagogy theory, infants are prepared to learn shared cultural knowledge from nonverbal communicative demonstrations addressed to them at a remarkably early age.

Keywords
social cognition, cognitive development, infant development

Received 7/12/10; Revision accepted 11/26/12

Adults' object-directed emotion expressions provide an important source of social information that even preverbal infants can benefit from when learning about objects and persons in their environment. This benefit is evidenced by the emergence of *social referencing* by the end of the 1st year when infants in triadic object-referential communicative interactions encounter unfamiliar objects they cannot appraise. After establishing joint attention, infants rely on the valence information expressed by the adult's object-directed emotion display to modulate their approach/avoidance behavior toward the referent (Moses, Baldwin, Rosicky, & Tidball, 2001; Mumme & Fernald, 2003; Walden & Ogan, 1988).

Object-directed emotion expressions are inherently ambiguous, however, and it remains to be clarified how infants interpret them. Humans use emotion expressions in two fundamentally different ways. In humans, as in nonhuman primates, the automatic activation of evolved patterns of facial-vocal emotion expressions makes manifest the agent's current emotional reaction toward the referent. Therefore, such emotion displays license a person-centered interpretation that attributes to the agent a person-specific subjective emotional attitude toward the object (e.g., "Mom is frightened of the snake").

In contrast, humans are unique among primates in their ability to voluntarily display their emotion expressions, using them as symbolic referential devices to communicate to other individuals valence-related information about a referent. During social referencing, adults often present their emotion displays without actually being in the emotion state expressed (think of the calm snake charmer who, when noticing a toddler approach his venomous snake, displays a fearful expression toward it to warn the child about the danger). In such cases, an

Corresponding Author:
Katalin Egyed, Social Minds Research Group, Department of Developmental and Clinical Child Psychology, Eötvös Loránd University, Izabella St. 46, 1064 Budapest, Hungary
E-mail: egyedkatalinagnes@gmail.com

object-centered interpretation would be more appropriate to capture what the adult intends to communicate to the infant (e.g., "The snake is dangerous to approach").

In fact, when adults address infants by *ostensive* signals of communication (e.g., establishing eye contact and using infant-directed speech; see Csibra & Gergely, 2009; Gergely & Csibra, 2005; cf. Sperber & Wilson, 1995) that introduce or accompany their referential emotion display, they often intend to convey relevant knowledge about the referent that (a) is generalizable beyond the situation and (b) represents shared knowledge available to other people, forming part of the cultural common ground shared by one's social community (e.g., the knowledge that snakes are dangerous). If infants were sensitive to these implications of adults' ostensive communicative acts, they could assign an object-centered interpretation to individuals' object-directed emotion displays during social referencing that would allow them (a) to act in an emotion-congruent manner not only toward the particular referent in the here and now but also when encountering other referents of the same kind in future situations and (b) to expect that other people also share the same emotional disposition and will behave accordingly toward the same (kind of) referents.

These alternative interpretations of referential emotion displays—person centered versus object centered—license different kinds of inferences (cf. Gergely & Jacob, 2012). The person-centered interpretation supports predictions about the emoting individual's likely actions toward the referent in the current episodic situation. However, because in this case the attributed emotional attitude is person specific, it does not sanction the generalization of the inferred action to predictions about other people. In contrast, the object-centered interpretation would license the generalization of the manifested information as knowledge that is shared by and applicable to members of the community other than the expresser.

But can young infants differentiate between person-centered and object-centered interpretations of object-directed emotion expressions, and are they sensitive to the different kinds of inferences these two interpretations support? Infants' precocious ability to rely on social referencing to modulate their approach/avoidance behaviors is not sufficient in and of itself to answer these questions, because either a person-centered or an object-centered interpretation could provide sufficient information for the infants to modify their actions toward the object in an emotion-congruent manner. It may seem somewhat surprising, therefore, that most previous accounts of early emotion understanding have assumed that infants assign a person-centered interpretation to other people's emotion displays during social-referencing situations (Moses et al., 2001; Mumme & Fernald, 2003; Repacholi & Gopnik, 1997; Tomasello, 1999; Walden & Ogan, 1988), whereas the alternative assumption, that infants may rely on an object-centered interpretation, has largely been ignored (but see our previous study, Gergely, Egyed, & Király, 2007).

We proposed that human infants' special sensitivity to ostensive signals of communication, as characterized by natural pedagogy theory (Csibra, 2010; Csibra & Gergely, 2006, 2009; Gergely, 2010), may also play a crucial interpretation-modulating role in helping infants disambiguate person-centered interpretations of others' referential emotion displays from object-centered interpretations. We hypothesized that in a second-person, child-directed ostensive signaling context, infants would recognize the expresser's communicative intention (Sperber & Wilson, 1995) and symbolic use of the emotion display and would assign an object-centered interpretation to it. In particular, we hypothesized that during social referencing, adults' ostensive communicative signals would trigger the shared-knowledge assumption of natural pedagogy (Csibra & Gergely, 2009; Gergely, 2010), leading infants to expect that their object-directed interpretation of the demonstrated emotion display would be applicable to other individuals as well. In contrast, when the referential emotion expression is observed from a third-person perspective in a noncommunicative context, infants should assign a person-centered interpretation and should not generalize their agent-specific attributions as applying to other individuals.

Although this alternative account has not been directly tested in previous experiments on infants' social referencing, we previously provided evidence in line with the predictions of the natural pedagogy theory that when cued by ostensive communicative signals, 14-month-olds indeed assign an object-centered interpretation to others' referential emotion displays, leading to an across-person generalization effect (Gergely et al., 2007). Although suggestive, these results remain inconclusive, insofar as we tested for object-centered versus person-centered interpretation of emotion displays only in the context of a communicative demonstration. Therefore, our previous study provided no direct evidence that infants would be equally capable of switching to a person-centered interpretation of the same emotion expressions and attributing an agent-specific subjective mental attitude to the emoting agent when he or she was observed in a noncommunicative context from a third-person perspective (i.e., without the biasing effect of ostensive cues).

The design of the present study allowed us to test (a) whether 18-month-olds are equally capable of making person-centered, compared with object-centered, interpretations of others' referential emotion expressions; (b) whether object-centered interpretations for such displays are constructed only under the interpretation-biasing effect of ostensive communicative signals; and

(c) whether removing such ostensive cues in a noncommunicative context would lead infants to assign the same emotion displays to an agent-specific person-centered interpretation. Furthermore, unlike the paradigms of previous social-referencing studies, the present paradigm allowed us to directly test the shared-knowledge assumption of natural pedagogy theory (Csibra & Gergely, 2009), which posits that in an ostensive communicative context, infants will generalize their object-centered interpretation of the expresser's emotion display as applicable to other persons (who have not expressed an emotional attitude toward the referent) as well. No such generalization effect was predicted, however, for when infants observed emotion displays in a noncommunicative context.

Method

Participants

Forty-eight 18-month-olds (mean age = 18 months 5 days, *SD* = 8.89 days, range = 17 months 15 days–18 months 18 days; 23 females and 25 males) participated in the study. They were randomly assigned to three different conditions. A further 17 infants were excluded because of maternal interference (n = 2), distress (n = 1), failure to complete the task by walking away during the familiarization phase (n = 3), or failure to touch either of the stimulus objects during the test (n = 11).

Stimuli

Two unfamiliar objects of different shapes and colors were used. Both their position in relation to each other on the table and the valence (positive or negative) of the emotion the experimenter displayed toward them were counterbalanced across subjects in each condition.

Design and procedure

The procedure began with an initial familiarization phase in which object-directed emotion displays were presented in either a communicative or a noncommunicative context. Then an object-request test phase was administered in which we varied the identity of the person who made the request (the requester being either the same person who displayed the referential emotion expressions or a different person). We created three experimental conditions: (a) a communicative-context/different-person condition, (b) a noncommunicative-context/different-person condition, and (c) a noncommunicative-context/same-person condition (see Fig. 1).

Familiarization phase: display of object-directed emotion expressions. Infants in all three experimental groups first participated in an emotion-display familiarization phase. During familiarization, infants were seated in their mother's lap in front of a table on which the two unfamiliar objects were placed on the left and the right sides, just out of the infants' reach. Mothers were asked not to communicate or point and were instructed to look down in order not to see the familiarization and test.

Communicative condition. One group of infants observed the demonstrator's object-directed emotion in a communicative context. The demonstrator first ostensively addressed the infant by looking and smiling at the infant while greeting him or her by name in infant-directed speech, saying, "Hi, [baby's name], hi! Look!" The demonstrator next turned to look at one of the objects, displaying a positive facial-vocal emotion expression (joy/interest), and then turned toward the other object, presenting a negative emotion display (dislike/disgust). During these emotion displays, the demonstrator provided further ostensive and referential cues of communication by looking back and forth between the object and the baby. This sequence was repeated a second time. Then the demonstrator stood up and left the room.

Noncommunicative condition. Two further groups of infants observed the same demonstrator's object-directed emotion displays in a noncommunicative context. The familiarization procedure was identical to that of the communicative condition except that the demonstrator acted as if she were alone: She never looked at or talked to the infant either before or during her object-directed emotion expressions.

Test phase: object request in different- and same-person conditions. In the test phase, a female experimenter (the *requester*) came to the table and sat down at its opposite side, facing the baby. She communicatively addressed the infant using ostensive signals (looking and smiling at the baby while greeting him or her by name using infant-directed speech, saying, "Hi, [baby's name], hi!"), displayed a manual requesting gesture (placing her hand between the two objects with her open palm turned upward), and said, "Give me one of them!" Then the mother rolled her chair a bit forward so that the infant could reach the objects. Throughout the test phase, the requester looked only at the infant and never at the objects. All events were recorded with two video cameras.

Two separate subject groups were tested in the different-person condition, in which the requester in the test phase was not the same person as the demonstrator in the familiarization phase. One group was presented with the demonstrator's object-directed emotion displays in the communicative context during familiarization, Communicative-Context/ Different-Person Condition

Noncommunicative-Context/ Different-Person Condition

Noncommunicative-Context/ Same-Person Condition

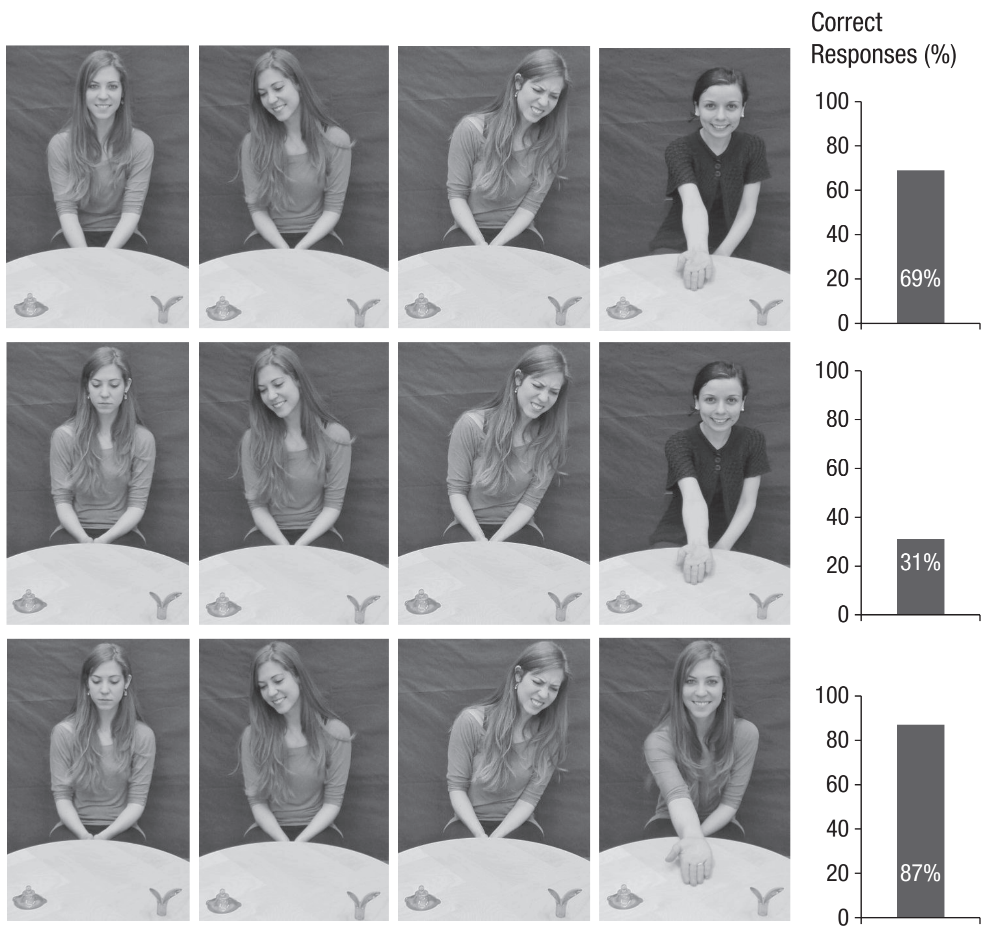


Fig. 1. Test-session sequence and results for the three conditions. The conditions were defined by whether the experimenter who presented displays of emotion addressed or did not address the infant and whether the same experimenter or a different experimenter requested an object. In all conditions, an experimenter made object-directed emotion displays, expressing a positive attitude toward one object and a negative attitude toward the other. The requester then asked the infant to give her one of the objects. The graphs show the percentage of responses in which infants chose (by giving to the requester or just touching) the object toward which they had seen a positive attitude displayed, separately for each condition.

whereas another group observed the same emotion expressions in the noncommunicative context.

Only one group participated in the same-person condition, in which the requester in the test phase was the same individual as the demonstrator in the familiarization phase. During familiarization, this group observed the demonstrator's emotion displays in a noncommunicative context. (At the end of familiarization phase, the demonstrator went to the door, stopped, and reapproached the table, playing the requester's role herself.)

Scoring. The video recordings were coded by two independent coders to identify which of the two objects the infant gave to the experimenter (or just touched) as their first response to the experimenter's request (coders were blind to the purpose of the study). Infants who gave no response were not included in the final sample. Interrater reliability between the two coders was perfect (100% agreement).

Results

We analyzed infants' preferential object choices during the object-request test phase in the different conditions. In the communicative-context/different-person condition, 11 infants chose the positively valenced object, and 5 infants chose the negatively valenced object. In the noncommunicative-context/different-person condition, 5 infants chose the positively valenced object, and 11 infants chose the negatively valenced object. In the noncommunicative-context/same-person condition, 14 infants chose the positively valenced object, and 2 infants chose the negatively valenced object (see Fig. 1). This analysis revealed a significant difference in the distribution of object choices among the three conditions, $\chi^2(2, N = 48) = 11.2$, $p = .004$. Planned pairwise comparisons showed that the pattern of choices in the communicative-context/different-person condition did not differ significantly from that in the noncommunicative-context/

same-person condition, $\chi^2(1, N = 32) = 1.646$, $p = .197$ (exact one-tailed). In contrast, the pattern of object choices in the communicative-context/different-person condition differed significantly from that in the noncommunicative-context/different-person condition, $\chi^2(1, N = 32) = 4.5$, $p = .038$ (exact one-tailed), odds ratio = 4.84. Finally, the pattern of object choices in the noncommunicative-context/same-person condition differed significantly from that in the noncommunicative-context/different-person condition, $\chi^2(1, N = 32) = 10.494$, $p = .002$ (exact one-tailed), odds ratio = 15.42.

We had tested the same hypotheses in an earlier pilot experiment in which we applied basically the same procedure we used in the study reported here; the main difference was that, to control for intrinsic object preferences, we ran a separate baseline object-choice condition (for details of this experiment, see Pilot-Experiment Method and Results, Table S1, and Fig. S1 in the Supplemental Material available online). The pattern of results in the three experimental conditions of our experiment replicated exactly the corresponding findings from our pilot study despite differences in the type of design, the stimulus objects used, and the experimenters. The results from our pilot study thus provide additional evidence supporting the reliability and generalizability of our findings.

Discussion

Previous approaches to infants' understanding of emotions during social referencing have assumed that when observing someone's object-directed emotion expression, infants attribute to that person the corresponding emotion as a person-specific emotional attitude state or a subjective disposition toward the object (Moses et al., 2001; Mumme & Fernald, 2003; Repacholi & Gopnik, 1997; Tomasello, 1999). The present results also demonstrate this capacity: After having observed a person's positive or negative emotion expressions toward two novel objects in a noncommunicative context, in response to the same person's subsequent object request, infants prosocially offered the object that the requester had earlier expressed a liking for. In other words, 18-month-olds can adjust their prosocial object-choice responses by consulting their memory representation of the requester's dispositional attitudes that was formed on the basis of the person's earlier object-directed emotion expressions. The fact that infants did not generalize this object choice when responding to another person's object request indicates that they indeed attributed person-specific dispositional attitudes to the demonstrator in the noncommunicative observation condition.

However, our findings go significantly beyond this by providing evidence, for the first time, that by 18 months of age (and probably even earlier; see our previous study, Gergely et al., 2007), infants can flexibly assign a qualitatively different interpretation to other people's referential emotion displays when these are presented to them in an ostensive communicative context. When communicatively addressed via ostensive signals, infants readily generalized their interpretation of the communicative agent's referential emotion manifestations as applicable to other individuals as well. This finding indicates that the infants interpreted these ostensive referential emotion displays as manifesting valence-related shared knowledge about the referent that they could rely on when making a prosocial object choice to satisfy the request addressed to them by a different individual (whose person-specific object preferences were unknown to them). This demonstration provides support for the shared-knowledge assumption of natural pedagogy theory (Csibra & Gergely, 2009, 2011), according to which, when communicatively addressed via ostensive signals, infants activate a default expectation that the manifested information about the referent will convey to them culturally shared knowledge that is available and applicable to other members of the community as well. This finding indicates that infants are prepared to learn about shared cultural knowledge through ostensive referential communicative acts at a remarkably early age.

Acknowledgments

The authors thank Gergely Csibra and John S. Watson for their valuable comments.

Declaration of Conflicting Interests

The authors declared that they had no conflicts of interest with respect to their authorship or the publication of this article.

Funding

This research was supported by the European Union's Sixth Framework Programme (Neural Origins of Language and Communication Grant 12738), the Economic and Social Research Council (Grant RES-062-23-0746), and the Hungarian Scientific Research Fund (Grants K63808 and PD78186).

Supplemental Material

Additional supporting information may be found at http://pss.sagepub.com/content/by/supplemental-data

References

Csibra, G. (2010). Recognizing communicative intentions in infancy. *Mind & Language*, *25*, 141–168.

Csibra, G., & Gergely, G. (2006). Social learning and social cognition: The case for pedagogy. In Y. Munakata & M. H. Johnson (Eds.), *Processes of change in brain and cognitive development: Attention and performance XXI* (pp. 249–274). Oxford, England: Oxford University Press.

Csibra, G., & Gergely, G. (2009). Natural pedagogy. *Trends in Cognitive Sciences*, *13*, 148–153.

Csibra, G., & Gergely, G. (2011). Natural pedagogy as evolutionary adaptation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *366*, 1149–1157.

Gergely, G. (2010). Kinds of agents: The origins of understanding instrumental and communicative agency. In U. Goshwami (Ed.), *Blackwell handbook of childhood cognitive development* (2nd ed., pp. 76–105). Oxford, England: Blackwell.

Gergely, G., & Csibra, G. (2005). The social construction of the cultural mind: Imitative learning as a mechanism of human pedagogy. *Interaction Studies*, *6*, 463–481.

Gergely, G., Egyed, K., & Király, I. (2007). On pedagogy. *Developmental Science*, *10*, 139–146.

Gergely, G., & Jacob, P. (2012). Reasoning about instrumental and communicative agency in human infancy. In F. Xu & T. Kushnir (Eds.), *Advances in child development and behavior: Rational constructivism in cognitive development* (Vol. 43, pp. 59–94). Waltham, MA: Academic Press.

Moses, L. J., Baldwin, D. A., Rosicky, J. G., & Tidball, G. (2001). Evidence for referential understanding in the emotions domain at twelve and eighteen months. *Child Development*, *72*, 718–735.

Mumme, D. L., & Fernald, A. (2003). The infant as onlooker: Learning from emotional reactions observed in a television scenario. *Child Development*, *74*, 221–237.

Repacholi, B., & Gopnik, A. (1997). Early reasoning about desires: Evidence from 14- and 18-month-olds. *Developmental Psychology*, *33*, 12–21.

Sperber, D., & Wilson, D. (1995). *Relevance: Communication and cognition* (2nd ed.). Oxford, England: Blackwell.

Tomasello, M. (1999). *The cultural origins of human cognition*. Cambridge, MA: Harvard University Press.

Walden, T. A., & Ogan, T. A. (1988). The development of social referencing. *Child Development*, *59*, 1230–1240.

Cognition 129 (2013) 232–240

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/COGNIT

Do infants bind mental states to agents?

Dora Kampis [a,b,*], Eszter Somogyi [b], Shoji Itakura [c], Ildikó Király [b]

[a] Institute of Cognitive Neuroscience and Psychology, Research Centre for Natural Sciences, Hungarian Academy of Sciences, POB 398, 1394 Budapest, Hungary
[b] Department of Cognitive Psychology, Eötvös Loránd University, Izabella utca 46, 1064 Budapest, Hungary
[c] Department of Psychology, Graduate School of Letters, Kyoto University, Yoshida-Honmachi, Sakyo-ku, Kyoto 606-8501, Japan

ARTICLE INFO

Article history:
Received 9 May 2012
Revised 6 July 2013
Accepted 8 July 2013
Available online 14 August 2013

Keywords:
Action perception
Infants
Belief attribution
Preferential choice

ABSTRACT

Recent findings suggest that infants understand others' preferential choice and can use the perspectives and beliefs of others to interpret their actions. The standard interpretation in the field is that infants understand preferential choice as a dispositional state of the agent. It is possible, however, that these social situations trigger the acquisition of more general, not person-specific knowledge. In a looking-time study we showed an Agent A demonstrating a choice, that only could have been interpreted as preferential based on the perspective (and thus the belief) of the agent, not the observer. Then we introduced a new agent (Agent B), who chose consistently or inconsistently with Agent A; also varying whether Agent B was an adult or a child. Results show that infants expected Agent B (both the adult and the child) to choose as Agent A, but only in the condition where according to Agent A's knowledge two objects were present in familiarization(confirming previous evidence on the importance of contrastive choice). We interpret these results in the following way: (1) infants do not encode the perspectives of other agents as person-specific sources of knowledge and (2) they learn about the object, rather than the agent's disposition towards that object. We propose that early theory of mind processes lack the binding of belief content to the belief holder. However, such limitation may in fact serve an important function, allowing infants to acquire information through the perspectives of others in the form of universal access to general information.

1. Introduction

In everyday life we constantly observe and interpret the actions of others – we make mental attributions and ascribe goals, intentions and dispositions to others. In developmental psychology it has become a question of interest as to when these abilities emerge. Contrary to previous claims that children start to understand the mental states of others around the age of 4 (e.g., Perner, 1991), recent findings suggest that the implicit roots of this ability, also called Theory of Mind (ToM), are already present in the first 2 years of life (Kovács, Téglás, & Endress, 2010; Luo, 2011; Onishi & Baillargeon, 2005; Surian, Caldi, & Sperber, 2007). However, it is not clear how these early competences are related to the full-blown ToM capacities (for a thorough conceptual review, see Rakoczy, 2012).

Indeed, infants are able to interpret the actions of agents as goal-directed from a very early age (e.g., Csibra, 2008; Gergely & Csibra, 2003; Gergely, Nádasdy, Csibra, & Bíró, 1995; Luo & Baillargeon, 2005; Southgate, Johnson, & Csibra, 2008; Woodward, 1998). In her seminal study, Woodward (1998) habituated infants to an event in which a hand grasped one of two toys. After habituation, the location of the objects was reversed and the hand then grasped either the old object in the new location or the new object in the old location. Results showed that the infants were surprised at the test events in which the hand grasped the new toy, suggesting that infants at 5–6 months of age encoded the goal of the action (the old object) and expected the hand to act accordingly.

* Corresponding author. Present address: Cognitive Development Center, Central European University, Hattyú u. 14, H-1015 Budapest, Hungary

E-mail address: Kampis_Dora@ceu-budapest.edu (D. Kampis).

0010-0277/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.cognition.2013.07.004

Realizing that actions are performed to reach a certain goal is the very basis of understanding the actions of agents. Since goals tend to be determined by the dispositions of the person acting, a second step is to understand that goal-directed actions are likely to be driven by certain dispositions. Recent findings (Luo & Beck, 2010; Song, Baillargeon, & Fisher, 2005) show that infants are capable of attributing dispositions to agents by around the first year of life. Using a modified Woodward-paradigm, Luo and Baillargeon (2005) habituated infants to a self-propelled box, which approached a cone in two conditions. In the control condition there was no other object present during familiarization. However, in the test condition, a second object, a cylinder was present, but was never approached by the box. Results showed that after the locationswitch in the test phase infants expected the box to approach the cone, but only when there was another object present during familiarization, suggesting that only in this case they attributed to the box the preference of the cone over the cylinder. Nevertheless, infants had no expectations about the action of the box in the one-object control condition, indicating the role of contrastive choice in attributing preferences. If, during the familiarization phase, only one object was present, the box could not make a real choice (since there was no other object that the box *did not* choose). Therefore, infants did not interpret the action of the box as an expression of preference; hence they could not predict the action of the box when a new object appeared.

In order to interpret an agent's actions, it is not always sufficient to determine the goals and dispositions of the agent; rather, under certain conditions, it is crucial to be aware of what the agent *knows* about the situation. Studies that create a knowledge gap between the infant and the agent about the situation show that infants are not only able to consider the agent's perceptual access when interpreting her actions, but they develop their expectations accordingly (Caron, 2009). Sodian and Thoermer (2004) found that 12-month-olds acted surprised when an agent incorrectly labeled an object after the agent had seen the object, but not if the agent had not had any perceptual access to it. Furthermore, in the experiment of Luo and Beck (2010), 16-month-old infants were firstfamiliarized to an agent choosing a red object over other objects. In the test phase, screens of different colors (red on one side, green on the other) were introduced between the agent and the infant, thus the perceived color of the screen was different for them. Results showed that infants developed an expectation of the agent's choice in the test phase according to the agent's perspective, suggesting that they could determine what the agent could see and used that information when developing their expectations.

In the above-mentioned experiments (e.g., Luo & Baillargeon, 2005 and Luo & Beck, 2010), infants interpreted the agent's actions as an expression of preference towards a certain object. However, it seems that this is only the case if the agent chooses a particular object over one or more other objects. Taking advantage of this feature of contrastive choice, Luo and Baillargeon (2007) designed an experiment in which they manipulated the agent's perceptual access and therefore created various possible interpretations of the scene from the side of the infants. In the familiarization phase, infants saw an agent reach repeatedly for Object A over Object B. Object B was either (1) visible to both the agent and the infant, (2) hidden from the agent behind a screen but visible to the infant, or (3) placedbehind the screen by the agent (so the agent knew it was there but no longer saw it). In the test phase, the objects' position was switched and both were visible to the agent and the infant. Results showed that the infants expected the agent to continue to reach for Object A in the test phase only if in the familiarization (a) both objects were present or (b) the agent placed the other object behind the screen. If the agent did not know about the presence of Object B, the infants did not interpret the action of the agent as the expression of his or her preference. This confirms the role of contrastive choice and suggests that infants were able to take the perspective of the agent as a basis for interpreting his or her actions. In a related study (Luo & Johnson, 2009), 6-month-olds showed a similar pattern, demonstrating that very young infants are able to understand preferential choice and can use the perspectives of others todetermine whether they have knowledge of certain objects to interpret their actions.

Moreover, building further on the results of Luo and Baillargeon (2007), in a subsequent study Luo (2011) introduced an experimental situation in which an agent falsely (or truly) *believed* that two objects were present in a setup. In the orientation trial, the agent herself positioned an object behind an opaque, or a transparent screen, and this object was subsequently removed. This removal was either visible (true belief one-object condition) or invisible (false belief two-object condition) for the agent. In familiarization trials, the agent chose the other object, in front of the second, always-transparent screen. The question was whether infants in test phase would attribute a preference to the agent, despite the fact that the infants themselves could only see one object during the choice (false belief two-object condition). Theresults show that 10-month-olds could figure out the basis of the agent's choice by inferring the beliefs of the agent in the situation. Based on this finding, the author suggested that even preverbal infants behave as though they can consider the *mental states* of others when making inferences about their actions.

From these results, one could conclude that infants encode the preferential choice of the agent as the expression of the agent's unique attitude towards that object. In this sense, the information acquired by the infants would be a highly specific, person-centered knowledge about the preference of that particular agent, which would only be useful in a limited number of situations. There is, however, a theoretical approach that allows a different interpretation. Egyed and her colleagues (Egyed, Király, & Gergely, 2013; Gergely, Egyed, & Király, 2007) argued that the interpretation of the expressions of referential attitudes is underdetermined. That is, a referential situation can allow more than one way of understanding the observed action. One view is the above-mentioned person-centered explanation, which leads to the acquisition of person-specific knowledge. Another interpretation isobject-centered, meaning that infants learn new information about the referent (about the particular object that was referenced). The former approach considers the content of the emotional expression as attached to the person (e.g., *this person chooses this object because she likes it better*), whereas the latter account considers it rather as a feature of the object (e.g., *this person chooses this object because this object is better*). Hence, the object-centered approach implies more general usability of the information in a wide variety of situations, and allows universal behavioral predictions regardless of the person involved. In line with these assumptions, Gergely et al. (2007) reported that in a violation-of-expectation study with 14-month-olds ostensive signals could induce an object-centered interpretation of the referential emotion displays of others in infants.

The possibility of encoding such a choice event in a person-independent way was raised by the results of Moore (1999, Experiment 2). He presented 12-month-old infants with habituation events, involving agents who attended to objects. Infants saw a person looking at and pointing to one of two toys. Following habituation to one event, they were shown new-object and new-side test events. Infants who saw the same agent throughout the procedure looked longer on new-object than new-side trials, replicating the results of Woodward (1998). Importantly, infants who saw one agent in habituation and a new agent in the test also showed exactly the same pattern of response. Moore argued that infants in this case did not represent the event in terms of intentional relation between the particular agent and the toy.

On the other hand, others (Buresh & Woodward, 2007; Henderson & Woodward, 2012) tested the same phenomenon with a paradigm that involved the original situation of reaching for objects. Their results suggested that infants track action goals over time by linking them to the individual person who performs them: even 9-month-olds were able to mark goals as attributes of individual people. However, the paradigms of Buresh and Woodward (2007), and also that of Henderson and Woodward (2012) share an important feature. After habituation, in the test phase where the new actor appears and the position of the objects is switched, the actor uttered the following questions: "Hi, where is it? Did they switch? Where did it go?". We assume that this utterance invites a pragmatic supposition that the objectthe agent is looking for is not present in the scenario, which makes the 'new person' situation ambiguous, and this ambiguity might account for null results in their switch-actor condition.

The aim of the present study was to test whether 10-month-old infants would encode knowledge conveyed in a social situation as person-specific, or if these situations trigger the acquisition of more general knowledge (in a non person-specific way). We wanted to see whether infants who see an agent expressing his or her attitude towards an object would expect a newly introduced agent to have the same attitude.

Following Luo (2011), we showed infants a scenario in which Agent A expressed a particular attitude (that is, a preferential choice) towards an object. We introduced the scenario with two objects present. In the next step we manipulated the perspectives of the infant and the agent such that the infant, but not the agent, saw that one of the two objects had been removed. After this the agent approached the remaining object. Therefore, in the eyes of the infants, the agent did not make a preferential choice (given the role of contrastive choice mentioned earlier), but according to the agent's knowledge two objects were present when the choice was made. In order for infants to interpret the situation as an expression of preference, they had to view the action from the agent's perspective; moreover, they had to take into account the agent's (false) belief based on what the agent previously perceived (namely, notwitnessing the removal of the object). In the test phase we introduced a new agent (Agent B), who once chose consistently with Agent A, once inconsistently. To test the range of agents that the acquired knowledge could be applied to, we varied whether Agent B was an adult or a 2-year-old child.

Our crucial condition is the False Belief condition, where the agent falsely believes that there are two objects and hence she believes she is making a real choice. We argued that this condition would be the strongest test of the person-specific encoding. In this case infants attribute a preference to the agent, and this preference attribution is based on the preceding belief attribution. Since there are two mental state attributions taking place, this could allow for an even stronger person-specific encoding.

Our hypothesis was that infants would not encode inferred information as person-specific facts; rather they would apply it in a general manner to the object and attribute a selective preference to other agents as well.

2. Study 1: Method

2.1. Participants

A total of 81 infants were recruited through newspaper advertisement. Of these, 15 infants were excluded due to technical errors (8), or crying or fussiness (7). The final sample consisted of 66 infants, with mean age of $M = 307$ days ($SD = 11$ days), 37 boys and 29 girls. Infants were accompanied by their parents, who gave their informed consent for participating in the study and received a toy as a "thank-you" gift.

2.2. Stimuli

Each infant was shown a series of videos presented in PsyScope. After a 10-s attention grabber, a series of 7 videos were shown (see Fig. 1), with the following structure. Five sequences were as part of the orientation trials[1] – (1)

[1] These video sequences in the orientation trials were similar to the orientation events of Luo (2011) with the following modifications. In the script of Luo, the agent herself put one of the objects behind an opaque (or transparent) screen, and then while she was away, a hand withdrew the object form behind the screen. So when the agent returned she could (in case of the transparent screen) or could not see (in case of the opaque occluder) that the second object disappeared. We wanted to make sure that infants follow that the agent could not be responsible for the withdrawal of the object (which would be a possible perceived solution in Luo's scenario). Therefore, in our version first the model saw that two objects were behind transparent screens, and then an opaque occluder was lowered and blocked the visual access of the protagonist for one of the objects. So, during the withdrawal of the second object,which was the next step in the orientation trials, the model was visible with her two hands resting in front of her. We added this modification because this would allow even more information for infants to encode the belief formed by the agent person-specifically, as she is constantly part of the scene.

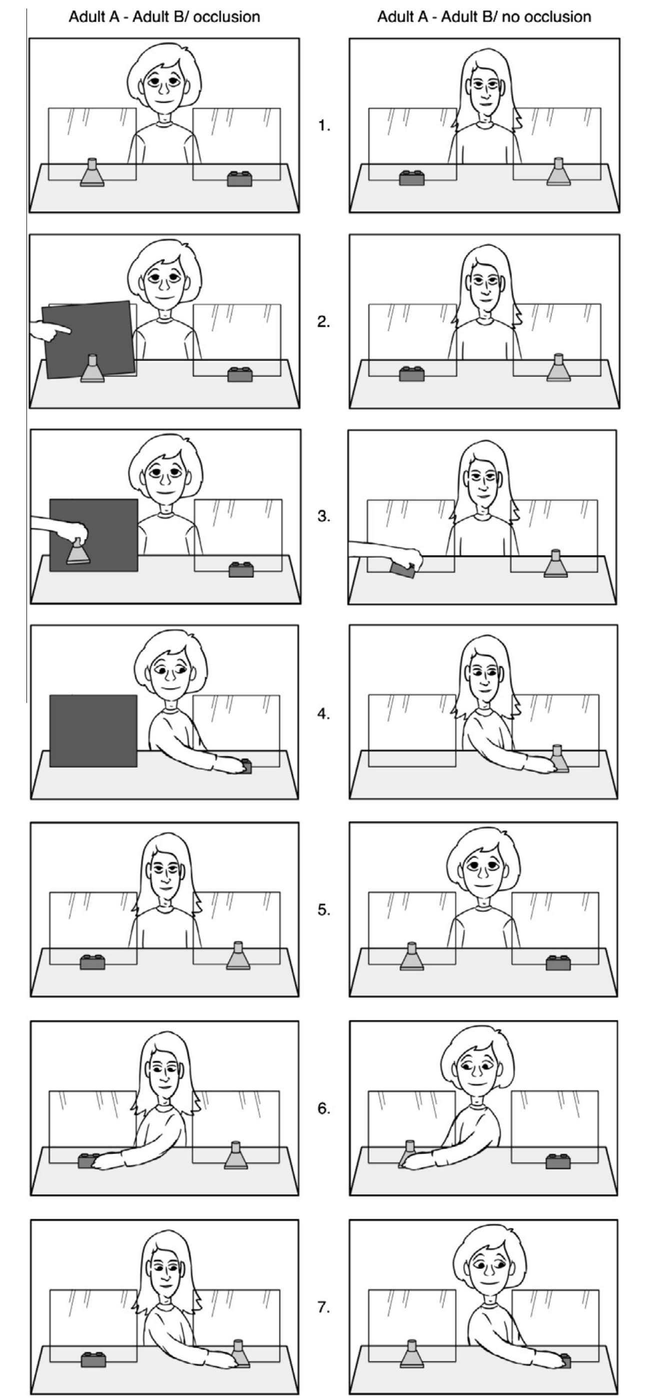


Fig. 1. Schematic drawings of the events during orientation and test trials in the Adult–Adult/occlusion, and the Adult–Adult/no occlusion (control) conditions. (1) Familiarization with Agent A, two objects visible, (2) occlusion of one object (only in the occlusion condition), (3) a hand, not visible to the Agent, reaches in and removes object A (this step being visible to the Agent only in no occlusion condition, but visible to the infant in all conditions), (4) Agent A grasps object B, (5) familiarization with Agent B, with both objects present again, and the locations switched, (6 and 7) Agent B grasps object A and object B (order of target object counterbalanced across infants). Adult–Child/occlusion condition was identical to Adult–Adult/occlusion condition, with the modification that Agent B was a 2-year-old child who performed the same acts as adult Agent B.

familiarization with Agent A, with two objects visible to both the infant and the agent, (2) occlusion of Object A (only in the occlusion conditions) from the agent but not the infant, (3) removal of Object A (this step being always visible to the infant, not visible to the agent in the Adult A–Adult B/occlusion and Adult A–Child B/occlusion conditions, but visible also for the agent in the Adult A–Adult B/no occlusion), (4) Agent A touches Object B (without Object A being present), (5) familiarization with Agent B, with both objects present but the object locations switched. Further two sequences were part of the test trials: (6) Agent B touches Object B (consistent choice), (7) Agent B touches Object A (inconsistent choice). The presentation of test events 6 and 7 were counterbalanced across subjects in each condition. At the beginning of each video (0.5 s before the onset), ashort sound was played to direct the infant's attention to the screen.

2.3. Materials

In the videos, Agent A sat behind a table and was visible from the waist up. The objects used in the videos were a $6 \times 6 \times 4$ cm yellow cuboid and a $6 \times 6 \times 6$ cm green pyramid with a 2 cm peak on the top. In the first setting, the two objects were placed behind two transparent screens, allowing both objects to be visible to the agent. In the occlusion conditions one object was occluded by a red cardboard placed next to the transparent screen, creating one single opaque screen between the agent and the object. During removal, a hand reached behind the screen and removed the object from the scene. This step was invisible to Agent A in the occlusion conditions but visible in the no-occlusion (control) condition. During the test phase the location of the objects was switched,and another agent (Agent B) was introduced. In the two test videos both objects were in front of Agent B, both behind a transparent screen and therefore visible for both the agent and the infant.

2.4. Design

A between-participants design was used, with the condition as the independent variable. Participants were randomly assigned to one of three conditions: (1) Adult A–Adult B/occlusion (23 infants), (2) Adult A–Child B[2]/occlusion (22 infants), or (3) Adult A–Adult B/no occlusion control group (21 infants). In Adult A–Adult B/occlusion condition both agents were adults, and the agent during familiarization did not see the removal of object A, due to the occluder placed between them and the object. In Adult A–Child B/occlusion condition the same setting was used, however, the second agent introduced in the 5th step was a 2-year old child. As a consequence, in these two conditions the agent made a false choice in the end of the orientation sequences (since the agent was not aware of not making a real choice due to the lack of the second object). In the third, Adult A–Adult B/no occlusion (control) condition object A was visible to the agent during all the orientation videos, therefore making it possible for agent A to observe the removal of the object.

[2] We use the letters A and B to refer to the first and second agent introduced (therefore "Child B" refers to the second agent being a child.

2.5. Procedure

Each infant sat on the lap of his or her parent, approximately 50 cm from the monitor. Two cameras were set so the experimenter could see online both the infant and the videos presented to him/her, and a split screen recording was created for offline coding. Videos 1–4 were approximately 15 s long and were presented consecutively to orient the infant. Video 5 was a familiarization trial with the objects' location switched, and the new agent (without action). During the test phase (after Videos 6 and 7) the screen froze and showed the last frame of the video until the infant looked away continuously for 2 s.

Recordings were then coded offline and looking times for inconsistent (IC) and consistent (C) videos were analyzed with a looking time program. A second observer coded 60% of the recordings. Inter-observer agreement was high ($r = 0.837$, $p = 0.01$).

2.6. Pretest

To test whether there were saliency differences between the two objects, the fixation patterns for the two objects used in the videos were measured using an eye-tracker (Tobii X50). Ten infants who were not part of the subsequent experiments (3 boys, 7 girls, mean age 309 days) were presented two pictures containing both objects in Clearview 2.5.1 (in the second picture the locations were switched) and fixation times were measured. A 2×2 repeated measures ANOVA with Identity of object (green cube vs. yellow pyramid) and Side (left vs. right) as factors showed no significant differences between fixation times to the two objects. There was no main effect of either Identity of object ($F(1,9) = 2.704$, $p = 0.135$), or Side ($F(1,9) = 2.861$, $p = 0.125$), and there was no significant interaction ($F(1,9) = 0.922$, $p = 0.362$). The result of this pretest proved that there were no baseline preferences with respect to the objects used in the video stimuli.

3. Study 1: Results

We conducted a repeated measures ANOVA in each condition (see Fig. 2), with Event (Consistent vs. Inconsistent) as a factor and Order (Inconsistent first vs. Consistent first) as a grouping variable. Preliminary tests showed no effect of sex, therefore this variable was omitted from further analyses in all three conditions.

In the case of the Adult A–Adult B/occlusion condition, the analysis revealed significant main effect of Event ($F(1,21) = 7.03$, $p = 0.015$, $\eta^2 = 0.251$), showing that infants looked significantly longer during inconsistent than during consistent events.

In the case of the Adult A–Child B/occlusion condition, the analysis revealed again a significant main effect of Event ($F(1,20) = 7.511$, $p = 0.013$, $\eta^2 = 0.273$), with longer looking times during inconsistent events. There was no main effect of Order, and no Order x Event interaction in either of the experimental conditions (Adult A–Adult B/ occlusion condition and Adult A–Child B/occlusion condition).

In both the Adult A–Adult B/occlusion condition and the Adult A–Child B/occlusion condition infants looked longer at Inconsistent (IC) events than at Consistent (C) events. (Mean looking times in Adult A–Adult B condition: $M_{IC} = 14.89$ s, $SD = 10.01$; $M_C = 11.14$ s, $SD = 6.6$; in Adult A–Child B condition: $M_{IC} = 15.58$ s, $SD = 9.96$; $M_{IC} = 11.29$ s,$SD = 6.52$).

In the Adult A–Adult B/no occlusion control condition there was no main effect of Event ($F(1,17) = 0.001$, $p = 0.977$), suggesting that looking times between Consistent and Inconsistent events did not differ significantly. Again, there was no main effect of Order and no Event x Order interaction. In this control (Adult A–Adult B/no occlusion) condition, looking times between the two types of test events were similar ($M_{IC} = 9.53$ s, $SD = 6.34$; $M_C = 9.43$ s, $SD = 5.72$).

A mixed type ANOVA with Condition Type (Test vs. Control) as between-subjects factor and Event (Consistent vs. Inconsistent) as within-subjects factor yielded a significant main effect of Event ($F(1,64) = 6.187$; $p = 0.015$, $\eta^2 = 0.088$) and a significant Event x Condition Type interaction ($F(1,64) = 5.601$; $p = 0.021$, $\eta^2 = 0.08$). These results confirm that infants looked longer at the inconsistent choice events only in the test conditions but not in the control condition.

4. Study 2

Despite the remarkable perceptual difference between the two agents (especially in the Adult A–Child B condition), it is still possible that the infants cannot perceptually distinguish between them. To exclude the possibility that infants expected an identical choice from Agent B because they could not distinguish her from Agent A, we conducted a control study.

4.1. Participants

We tested fourteen 10-month-old infants, 10 of whom contributed to the final sample, and 4 were excluded due to crying (3) or technical error (1). Mean age was 300 days (SD = 9 days), 5 boys. Infants were accompanied by their parents, who gave their informed consent for participating in the study and received no financial reward but a symbolic toy for participation.

4.2. Materials

We used a similar procedure as Buresh and Woodward (2007), who tested whether 13-month-olds could discriminate the two agents in their studies. In order to guarantee that infants in this control experiment had the same amount of exposure to Agent A as in the other conditions, we used the same familiarization videos (steps 1–4) as in our occlusion conditions of Study 1, that ended in Agent A choosing one of the objects in step 4. After this, the location of the objects remained identical, but the previously removed object was present again (as in the original test phase), and two test events followed. In one condition the same agent (Agent A) remained, but wearing a different

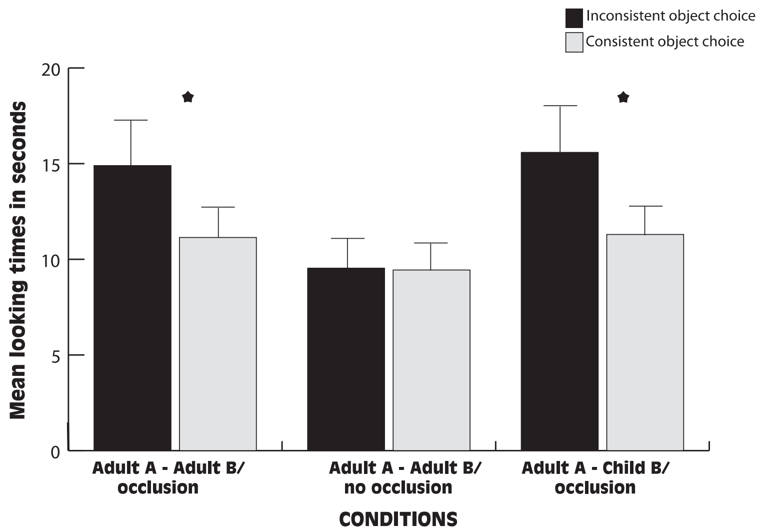


Fig. 2. Mean looking times of infants in the three conditions. Error bars represent standard errors. An asterisk (*) denotes $p < .05$.

shirt, and chose again the same object as in familiarization phase. In the other condition Agent B was present during test phase, and similarly to Agent A, chose the same object (with the location of objects still being constant). Infants sawboth test events in a within-subjects design, and we hypothesized that longer looking times to the test event with Agent B reflects novelty preference towards Agent B.

Since Agent A was wearing different clothing during the test phase than during the familiarization phase, if infants noticed the difference it could not have been merely due to noticing the different outfit of the actors. Moreover, the choice performed was identical across the familiarization and the two test videos. Hence, the only essential difference between them was the identity of the agent.

We used the Adult A–Adult B agent-pair because we argue that this is the harder discrimination to make; if infants can distinguish one female adult from another female adult, they must be able to distinguish between an adult female and a 2-year-old boy.

4.3. Procedure

The procedure was identical to Study 1. After testing, videos were coded offline and looking times for Same Agent and New Agent test videos were analyzed with a looking time program.

4.4. Results

We conducted a repeated measures ANOVA, with Agent (New Agent vs. Same Agent) as a factor and Order (Same Agent first vs. New Agent first) as a grouping variable. Preliminary tests showed no effect of sex, therefore this variable was omitted from further analyses.

The analysis revealed significant main effect of Agent ($F(1,8) = 6.264$; $p = 0.037$, $\eta^2 = 0.439$), showing that infants looked significantly longer during New Agent ($M_{NA} = 18.42$, $SD = 8.77$), than during Same Agent ($M_{SA} = 13.11$ s, $SD = 7.2$) videos. There was no effect of Order, and no Agent $\times$ Order interaction.

Thus, the looking time patterns revealed a novelty preference for Agent B, suggesting that infants were able to distinguish between the two agents.

5. Discussion

The goal of the present study was to investigate whether infants can bind the content of encoded mental states of a social partner to the respective person. Our results suggest that 10-month old infants can use the perceptual and knowledge states of others, without necessarily tracking which agent has a specific knowledge. Hence infants do not interpret the inferred preferential choice of an agent as related to that person; rather, they seem to infer that the content of the preferential choice could also be appropriate for other agents. Thus belief computation and preference attribution can be used not only to learn *about others* but also to learn *about the world* through the lenses of others, in the sense that they might learn about the object.

Our data suggest that infants could follow whether an agent does or does not have visual access to a scenario (e.g., not seeing that one of the objects was removed), and could also infer the agent's belief based on this visual access. Thus, they have also computed the agent's preference as a function of this (false) belief. Infants attributed preferential choice only when the agent has initially seen two objects in the scenario, but then did not see the removal of one of the objects, and thus infants have also computed the belief of the agent of two objects being present as a source of action prediction. Hence, the present results confirm those of Luo (2011, and also Luo & Baillargeon, 2005; Luo & Baillargeon, 2007) in the sense that for infants the availability of contrastive choice is crucial for inferring preference.

Moreover, these results confirm that infants can rely on the perspectives and knowledge of other agents to set up the preconditions for preference inferences. When attributing choice, infants did not rely on their own visual access to the scenario; rather, they took into account the agent's conflicting perspective and epistemic state – whether according to the agent's knowledge there are two objects or one – which was the reverse of what the infants could themselves see. It can be thus said that the results of the present paradigm speak for a mental state attribution that is functionally similar to belief attribution in adults. Proper belief attribution is characterized by Rakoczy (2012) as being inferentially connected to other mental states in order to guide rational action selection. In our case infants do, in fact, have to integrate the attributed beliefs with the preference in order to interpret the agent'sactions and have expectations regarding her (or another agent's) behavior.[3]

Our results show, nevertheless, that infants use the knowledge of another agent to predict not only the actions of this agent exclusively, but also the actions of other agents. Hence, the present results question the early availability of person-specific preference attributions, and refine the interpretation of early mindreading competencies of infants proposed by Luo (2011). Luo has found that 10-month-old infants interpret a person's choice of toys based on her true or false beliefs about which toys were present. According to her interpretation, these results indicate that *like adults*, even preverbal infants can consider others' mental states when making inferences about their actions. This would suggest that infants encode the attributed mental states as belonging to that specific person only. On the contrary, results of the present study showed that infants did not handle the emergentinformation based on someone's visual access in a person-specific way. They have used the acquired information to predict the actions of other agents accordingly.

Our pattern of results is in line with the findings of Moore (1999). He found that after habituating infants to an agent pointing to one of two objects, infants expected another agent as well to point to the same object. On the contrary, Buresh and Woodward (2007) found that infants did not expect a new person to approach the same object that the previous person did.

A possible explanation for the different findings could be that because the two agents, both in the Moore (1999) study and in the Adult A–Adult B/occlusion condition in the present study, were similar in appearance, infants may not have noticed the agent switch (whereas in the Buresh and Woodward study the gender difference between the two actors was salient). Study 2 addressed this issue and confirmed that infants at 10 months of age are able to perceptually distinguish between the two agents. In this study we tested whether they have a novelty preference to the new agent performing the same choice. According to our results infants can in fact discriminate between the two agents, therefore this factor cannot explain the results obtained.

In summary, based on the findings of the present study, we suggest that 10-month-olds can compute the visual access of others and attribute (true and false) beliefs accordingly. Furthermore, infants can use the inferred mental states in their evaluation and prediction of forthcoming actions and their outcomes. However, it is possible that infants are unable to integrate the information of the *source* of the mental states with their *content*. A possible interpretation of this interesting pattern of results is that infants are not yet able to track and relate the mental states of others in a person-specific manner.

We do not wish to imply that there can be attribution of propositional attitudes without an agent. Preference attribution requires an agent to be present in order for the preference attribution to be triggered. As such, an agent *is* necessary for mental state attribution. However, the question is whether infants in this case store this relation (this is what we refer to as binding), or it can be substituted by an 'agent–placeholder'. We argue that early on infants do not take into account that the attributed mental state was bound to a unique person.

This difficulty in information binding may also have an advantage for young infants. We suggest that infants can use content information they derive from the belief ascription to the social partner. Without source differentiation the content becomes part of a shared knowledge base that is applicable to other agents as well. An object that was chosen by one person can be categorized as a preferable/good object – a piece of information that can be generalized as relevant to conspecifics. Gergely, Egyed, and Király (2007) also proposed such an object-centered approach as an action interpretation strategy available for infants at the age of 14 months. Additionally, Egyed et al. (2013) found that 18-month-olds could switch between the object-centered and person-specific interpretations in a preference attribution paradigm. As such, it is plausiblethat young infants, since they cannot store and retrieve person-specific source information, can use an object-centered approach early in life to gather a universally shared knowledge base. Thus, the potential advantage of this early, but limited competence (i.e., the lack of binding of mental states to agents) is that it can support the recognition of generally preferred objects as common goals, and as such it could serve an important role in promoting joint action and cooperation.

Our study taps onto the hotly debated question whether infants possess full-blown Theory of Mind abilities. Some recent approaches argue that infants are only capable to attribute 'belief-like' informational states (see Apperly & Butterfill, 2009) or 'subdoxastic' states (Rakoczy, 2012), rather than 'proper beliefs'. In our view the beliefs (and other mental states) attributed by infants might not call for a different terminology that distinguishes them from proper beliefs (c.f. Onishi & Baillargeon, 2005; Surian et al., 2007). We claim that belief attribution observed in infants differs from adult-like ToM abilities, but these could be seen as points on a continuum rather than discrete categories. The main goal of our current study is to focus on the mechanisms of mental state ascription,rather on the terminological distinctions between proper beliefs and belief-like (or subdoxastic) states. The mechanism of mental state attribution has various features that can be subject of change during the lifespan. We suggest that

[3] Since looking time studies by their nature don't allow for testing the further criteria of 'proper beliefs' proposed by Rakoczy (2012) (whether the formed beliefs are accessible to consciousness, or inferentially promiscuous), we cannot come to a firm conclusion regarding the exact nature of the beliefs attributed by the infants.

one of these is the binding the belief contents with the owner of the belief.

A possible developmental trajectory of theory of mind abilities could be understood in light of the potential binding of mental states to the corresponding agent. Specifically, we propose that such person-specific belief encoding may emerge after the end of the first year of life, and then will become dominant strategy. Our results show that 10-month-olds possibly lack this ability, and results of Kovács et al. (2012) are in line with this pattern. In a follow-up study of Kovács et al. (2010), Kovács et al. (2012) found that even 14-month-old infants had difficulty tracking the agent that a belief belongs to, and were only able to do so if the agents were named in the beginning of the experiment. This, together with the results of Gergely, Egyed, and Király (2007), suggests that around 13–14 months under certain circumstances infants are ableto bind the mental state contents to the corresponding agents, but it still might not be the dominant stance. Later on, during the second year of life, both strategies are available in parallel, as suggested by Egyed et al. (2013). Our interpretation of their view is that object-centered information encoding arises as a result of the lack of mental state binding in the first year of life. Later on this process – that is, transmitting information generalisable across agents – is still available, but it is mostly triggered by specific cues, like ostensive, communicative signals. After the second year of life, the person-specific belief encoding can become more prominent, possibly resulting in processes similar to the full-blown Theory of Mind abilities. Promising evidence comes from studies using tasks in which children make active choices based on situative inferences (in situations involving communication, Southgate et al., 2010;active helping, Buttelman et al., 2009, and helping and correction with communicative pointing, Knudsen & Liszkowski, 2011; Knudsen & Liszkowski, 2012) – this in fact can reflect that children's belief attribution is more flexible and can guide their control of action.

Note that our results cannot be due to perceptual differences in the test events. Namely, that switching agents would be distracting or overwhelming for the infants. While in the Inconsistent choice trials, infants saw events with at least two salient differences between familiarization and test (new model grasping another toy), in the Consistent choice trials, there was only one salient difference (the new model grasping old toy). This difference in itself could result in a similar pattern of looking behavior. However, based on the above-mentioned perceptual difference, the predicted looking behavior of the Adult A–Adult B/no occlusion (control) condition would be similar to the two occlusion conditions, as in this regard they have the same structure of events. Since the results in this (control) condition differed significantly from the other two conditions, the perceptual differences cannot be responsible for the obtained data pattern.

Furthermore, this pattern of results cannot be explained merely by the combination of screens in the different condition. First, in her study Luo (2011) ruled out the possibility that infants' attribution of preference to the agent in one, but not in the other condition (shown by difference in pattern of looking times) could be merely the result of the different occluders in the two conditions. In Study 2 of Luo (2011) the agent either (a) falsely believed that there was only one object or (b) had a true belief that there were two. Results showed that infants' attribution of preference depended on the agent's knowledge about the objects and not the particular arrangements of occluders in the scene. Second, if in the present study merely the combination of screens (one transparent and one opaque screen) would determine looking times in the test events of FalseBelief condition (e.g., in the consistent choice test event), in Study 2 we should not have observed different looking times in the two test events (since both are preceded with familiarization including one transparent and one opaque screen, and both include consistent choice in test phase).

In sum, the main objective of the present study was to test whether infants understand preferential choice and others' perspective as a person-specific disposition in their action interpretations. We found that after a preference demonstration, where infants had to compute the preference based on someone's false belief, infants predicted the same object choice for a new agent. This finding leads to the proposal that there is a graspable gap between the early understanding of mental states and full-blown theory of mind capacities. Early competencies comprise (a) the capability to infer another person's belief upon this person's visual access to a situation and (b) the ability to use previously inferred mental representations for action prediction. Other studies that claimed to reveal early Theory of Mind competencies could also be interpreted in light of these two aspects of ToM (see Kovács et al., 2010; Onishi& Baillargeon, 2005; Surian et al., 2007; and Luo, 2011). Thus, what young infants may lack is the binding of the content of mental states to the person from whom that specific content was learned – a critical component for a person-specific mental state representation and action prediction system.

Acknowledgements

This research was supported by a Bolyai Research Grant from the Hungarian Academy of Sciences to Ildikó Király. The European Union and the European Social Fund have provided financial support to the project under the Grant agreement TÁMOP 4.2.1/B-09/1/KMR-2010-0003. The project received additional funding from OTKA-T049615 and KMOP-4.2.1/B-10-2011-0002. The authors thank Á. Kovács, G. Csibra, M. Hernik, O. Mascaro, Z. Káldy and two anonymous reviewers for helpful suggestions and comments on this paper, D. Tatone for the drawings; the infants and their parents for their participation, and everyone who helped in any part of the project.

References

Apperly, I., & Butterfill, S. (2009). Do humans have two systems to track beliefs and belief-like states? *Psychological Review, 116*(4), 953–970.
Buresh, J. S., & Woodward, A. L. (2007). Infants track action goals within and across agents. *Cognition, 104*, 287–314.
Caron, A. J. (2009). Comprehension of the representational mind in infancy. *Developmental Review, 29*, 69–95.

Csibra, G. (2008). Goal attribution to inanimate agents by 6.5-month-old infants. *Cognition, 107*, 705–717.
Egyed, K., Király, I., & Gergely, Gy. (2013). Communicating shared knowledge in infancy. *Psychological Science, 1–6*. http://dx.doi.org/10.1177/0956797612471952.
Gergely, G., & Csibra, G. (2003). Teleological reasoning in infancy: The naive theory of rational action. *Trends in Cognitive Sciences, 7*(7), 287–292.
Gergely, G., Egyed, K., & Király, I. (2007). On pedagogy. *Developmental Science, 10*(1), 139–146.
Gergely, G., Nádasdy, Z., Csibra, G., & Bíró, S. (1995). Taking the intentional stance at 12 months of age. *Cognition, 56*, 165–193.
Henderson, A. M. E., & Woodward, A. (2012). Nine-month-old infants generalize object labels, but not object preferences across individuals. *Developmental Science, 15*(5), 641–652. http://dx.doi.org/10.1111/j.1467-7687.2012.01157.x.
Knudsen, B., & Liszkowski, U. (2011). 18-Month-olds predict specific action mistakes through attribution of false belief, not ignorance, and intervene accordingly. *Infancy*. http://dx.doi.org/10.1111/j.1532-7078.2011.00105.x.
Knudsen, B., & Liszkowski, U. (2012). Eighteen- and 24-month-old infants correct others in anticipation of action mistakes. *Developmental Science, 15*, 113–122. http://dx.doi.org/10.1111/j.1467-7687.2011.01098.x.
Kovács, Á., Téglás, E., & Endress, A. D. (2010). The social sense: Susceptibility to others' beliefs in human infants and adults. *Science, 330*, 1830–1834.
Luo, Y. (2011). Do 10-month-old infants understand others' false beliefs? *Cognition, 121*(3), 289–298.
Luo, Y., & Baillargeon, R. (2005). Can a self-propelled box have a goal? Psychological reasoning in 5-month-old-infants. *Psychological Science, 16*, 601–608.
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.
Luo, Y., & Beck, W. (2010). Do you see what I see? Infants' reasoning about others' incomplete perceptions. *Developmental Science, 13*(1), 134–142.
Luo, Y., & Johnson, S. C. (2009). Recognizing the role of perception in action at 6 months. *Developmental Science, 12*, 142–149.
Moore, C. (1999). Intentional relations and triadic interactions. In P. D. Zelazo, J. W. Astington, & D. R. Olson (Eds.), *Developing theories of intention: Social understanding and self-control* (pp. 43–61). Mahwah, NJ: Erlbaum.
Onishi, K. H., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science, 308*, 255–258.
Perner, J. (1991). *Understanding the representational mind*. Cambridge, MA: MIT Press.
Rakoczy, H. (2012). Do infants have a theory of mind? *British Journal of Developmental Psychology, 30*(1), 59–74.
Sodian, B., & Thoermer, C. (2004). Infants' understanding of looking, pointing, and reaching as cues to intentional action. *Journal of Cognition and Development, 5*, 289–316.
Song, H., Baillargeon, R., & Fisher, C. (2005). Can infants attribute to an agent a disposition to perform a particular action? *Cognition, 98*(2), B45–B55.
Southgate, V., Johnson, M. H., & Csibra, G. (2008). Infants attribute goals even to biomechanically impossible actions. *Cognition, 107*, 1059–1069.
Surian, L., Caldi, S., & Sperber, D. (2007). Attribution of beliefs by 13-month-old infants. *Psychological Science, 18*, 580–586.
Woodward, A. L. (1998). Infants selectively encode the goal object of an actor's reach. *Cognition, 69*, 1–34.

Retrospective attribution of false beliefs in 3-year-old children

Ildikó Király[a,b], Kata Oláh[a], Gergely Csibra[b,c], Ágnes Kovács[b]

[a]Eötvös Loránd University, MTA-Momentum Social Minds Research Group, Budapest
[b]Central European University, Budapest
[c]Birkbeck, University of London

WORD COUNT: 5556

Acknowledgments
This study was supported by the grant OTKA 116 779 to Ildikó Király. The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013), ERC Grant n° 609819 SOMICS, and an ERC Starting Grant n° 284236 to Á. Kovács.

We thank Dorottya Árva, Júlia Baross, Iulia Savos and Zsófia Kalina for their help with data collection and Fruzsina Elekes and Dora Kampis for their valuable comments on the manuscript.

Corresponding author:
Ildikó Király
Cognitive Psychology Department, Social Minds Research Group
Eötvös Loránd University
Izabella 46.
Budapest -1064, Hungary
kiralyi@caesar.elte.hu

Research highlights

- We make a distinction between prospective belief tracking and retrospective belief attribution, and tested whether mechanisms of the latter type are available for young children.
- We found that 36-month-olds could retrospectively infer the content of someone's beliefs by combining present information with relevant events retrieved from episodic memory.
- 18-month-olds showed no evidence of adopting such inferences, though they could rely on prospective tracking of false beliefs.
- As soon as mechanisms of episodic memory are available, it can contribute to social cognitive processes, such as the attribution of mental states.

Abstract

We investigated whether young children would attribute beliefs to others in a retrospective manner, based on episodic retrieval of the details of the events that brought about the beliefs. We developed a task in which prospective belief tracking would not, but retrospective attribution mechanisms would have allowed children to infer that someone had a false belief. Eighteen- and 36-month-old children observed a displacement event, which was witnessed by a person wearing sunglasses. Having later discovered that the sunglasses were opaque, 36-month-olds correctly inferred that the person must have had a false belief about the location of the objects, and used this inference in resolving her referential expressions. Eighteen-month-olds failed in this task, suggesting that they cannot retrospectively attribute beliefs or revise belief attributions. However, an additional experiment provided evidence for prospective tracking of false beliefs in 18-month-olds, where they had been informed about he opacity of sunglasses in advance. This dissociation reflects that 18-month-olds rely primarily on prospective belief tracking, while 36-month-olds can also flexibly compute beliefs retrospectively, based on episodic memories, well before they pass explicit false belief tasks.

Key words: theory of mind, episodic memory development, prospective and retrospective processes in mindreading

Humans are undoubtedly ultra-social beings; they live their lives in an almost continuous flow of interactions (Boyd & Richerson, 1996). This ubiquitous sociality imposes an enormous socio-cognitive demand: in order to engage in communication, collaborations, or any event that is governed by socially formed concepts, like norms or customs, they need to be able to take into account the mental states of their social partners. Accordingly, everyday functioning requires humans to become experts in monitoring others' minds to predict and interpret their behavior — an ability also termed as theory of mind (ToM).

Despite this general view, there is scarce empirical evidence on the dynamics and the characteristics of the mechanisms that allow for computing others' mental states, and on the development of such mechamisms. The typical paradigms used for testing ToM competencies focus on measuring the attribution of false beliefs at a specific time point: at the end of the scenario. In the standard location-change false-belief task, the participant is exposed to the following event sequence: a character, Sally puts her chocolate into a basket (location A) and leaves. Another character, Ann, changes the location of the chocolate to a box (location B). Then, Sally returns for her chocolate. In the explicit version of the task, at this moment, participants are prompted to answer direct questions regarding Sally's impending action, which require them to take into account her beliefs about the situation (Baron-Cohen, Leslie, & Frith, 1985; Wimmer & Perner, 1983). Young children usually fail in this task. However, implicit versions of this task developed in the last decade have provided ample evidence that infants, similarly to adults, can track a character's beliefs, even without being explicitly asked to do so, as reflected by their looking times (Onishi & Baillargeon, 2005; Surian, Caldi, & Sperber. 2007; Kovács, Endress, & Téglás, 2010), anticipatory looks (Southgate, Senju, & Csibra, 2007; Rubio-Fernandez, 2013), or active behavior (Kundsen & Liskowski, 2012; Buttelmann, Carpenter, & Tomasello, 2009). However, whether the tasks were implicit or explicit, previous studies relied on protocols that did not allow disentangling the different cognitive mechanisms involved at the different stages of the scenarios.

Crucially, the dynamics of theory of mind processes is widely neglected: for example, it is unknown at what point of the event stream beliefs are computed and attributed in false belief tasks. Taking the standard location-change false-belief paradigm as an example, belief attribution could take place at the beginning of the story, when Sally puts her chocolate into the box, at the end, when the participant is prompted to predict Sally's behavior, or in between, for example when Ann relocates the chocolate. Crucially, these options differ not only in timing but also in the types of inferences and computations they demand. Attributing a (true) belief at the beginning of the story, when the protagonist's perceptual access to a state of affairs is recorded, requires only the maintenance of this attributed belief, despite changes of reality, as the events unfold, to succeed in the task (Kovács, 2016). This is 'prospective' attribution of belief because it may or may not have any immediate use for the observer, but it can be stored and maintained in case it is required in further inferences. Such a prospective mechanism of belief attribution does not even have to track the truth value of the belief for enabling passing a false-belief task, and does not necessarily require encoding the source event that led to belief attribution.

In contrast, if belief attribution takes place at the end of the story, when the content of the relevant belief is needed for action prediction, recovering the content of the belief must be based on a memory search targeting all relevant information that can potentially contribute to the identification of such a content. This search may be triggered spontaneously in implicit tasks (e.g., by the reappearance of the actor whose belief is relevant — when Sally returns to find her chocolate), or by the direct question regarding the protagonist's beliefs or actions in explicit tasks. While this 'retrospective' mechanism of belief attribution does not require continuous tracking and maintenance of attributed beliefs, it can only be performed successfully if all relevant details of past events are faithfully preserved and accessible when needed. For instance, in order to pass (explicit or implicit version of) the Sally-Ann task by retrospective belief attribution, one should recall the episode when Sally put her chocolate into the basket, trace the intervening events related to her and/or the chocolate (she did not see the replacement), and infer that Sally still believes her chocolate is in the basket. Retrospective, memory-based belief inferences will also allow to retrieve belief-relevant information that seemed irrelevant at the time of encoding and hence could not have been taken into account in prospective belief attribution. Such episodic retrieval could also serve as an important mechanism for belief *revision*: in case novel information comes up regarding the past context that induced prospective belief attribution, one can retrospectively re-compute the content of already attributed beliefs (cf. Klein et al., 2009).

Thus, these two mechanisms of belief attribution are likely not mutually exclusive, but they work in an integrated manner. If, for example, Sally's belief is attributed when Ann relocates chocolate, it might be based on retrospective recalling of what happened before (Sally saw the chocolate in the basket), and the resulting belief should be prospectively maintained until it is exploited for action prediction. Importantly, in the commonly used false belief tasks these computational strategies cannot be disentangled because they predict similar outcomes in terms of participants' behavior. Successful performance in all these tasks could simply be based on prospective belief attribution and maintenance of these attributed beliefs throughout the event. In fact, since retrieving past episodes poses difficulty for young children (Fivush & Nelson, 2004; Hayne, 2004), it is a plausible assumption that their successful performance in implicit false-belief tasks relies on prospective attribution mechanisms. The purpose of the present study was to test whether and when retrospective attribution mechanisms are available to children in implicit tasks. To achieve this aim, we had to develop a task that cannot be solved by purely prospective belief computations.

We developed a new paradigm[1] by extending the referential disambiguation ToM task of Southgate, Chevallier, and Csibra (2010). The crucial extension we introduced was a belief revision phase in between the belief induction phase and the test. The task had the following structure: In the belief induction phase, the experimenter hid two novel objects into two boxes and, while wearing sunglasses, she ‘witnessed’ as the location of the two objects were swapped. This scene could result in the prospective attribution of a true belief (TB) to the experimenter about the respective location of the objects, if the sunglasses are transparent. In the belief revision phase, while the experimenter was away, the participants explored her sunglasses, which turned out to be either opaque or transparent. In the condition where the sunglasses turned out to be opaque, children were expected to retrospectively update her belief content from true to false (by recalling that she was wearing the opaque sunglasses when the location change took place), and re-compute the content of the attributed belief regarding the location of the objects. We label this condition TB-FB, indicating that, to succeed, children had to retrospectively change the status of the attributed belief from true to false. In the condition where the sunglasses were transparent, retrospective revision of the attributed belief was not necessary (TB-TB condition).

In the following test phase, the experimenter returned, pointed to one of the boxes and asked for an object. The dependent measure of the study was whether children, in response to this request, gave her the object from the referred or from the other (non-referred) box. In line with Southgate et al. (2010), we built our predictions on the following consideration: when the experimenter pointed to the box containing one of the novel objects, children would not interpret the gesture as referring to the box itself; rather, they map it to the object hidden inside. Importantly, this referent mapping is dependent on the attributed belief: the

[1] In describing this paradigm, we assume that children passed the original task by Southgate et al. (2010) by relying prospective belief attribution. However, in the General Discussion will return to alternative interpretations of the task.

experimenter's gesture must refer to the object about which she (truly or falsely) believes to be located in that particular box. If children in our task retrospectively update the TB to a FB when the sunglasses turn out to be opaque during the belief revision phase (TB-FB condition), they should choose the non-referred box, while they are expected to choose the referred box in the TB-TB condition, when such an update was not required.

Experiment 1

In Experiment 1, we tested whether memory-based retrospective belief attribution mechanisms were available for 18- and 36-month-old children. The younger group represents the age when infants have been shown to pass interactive false-belief tasks (Buttelmann, Carpenter & Tomasello, 2009; Southgate et al., 2010), and the older group targeted the age when episodic memory capacities seem to emerge (Eacott & Crawley 1998; Scarf, Gross, Colombo, & Hayne, 2011).

Methods

Participants. The planned sample size was 40 18-month-old infants and 40 36-month-old children equally distributed to the TB-TB and TB-FB conditions. An additional 3 18-month-olds and 1 36-month-old were excluded and replaced because of experimenter error during the procedure. Some children did not make a choice or chose both boxes during the test: 9 18-month-olds (4 in the TB-TB, 5 in the TB-FB condition) and 4 36-month-olds (2 in the TB-TB and 2 in the TB-FB condition). Since these participants completed the task, they were not replaced. Thus, the final sample that produced evaluable data included 31 18-month-old infants (TB-TB condition: 16; TB-FB condition: 15; mean age=18.1 months; range: 17.5 - 18.5 months) and 36 36-month-old children (TB-TB condition: 18; TB-FB condition: 18; mean age= 36.3 months, range: 35.0 - 36.9 months).

Materials. A toy egg and a toy carrot were used in the warm-up trials as objects to be found. In addition, two novel objects (a blue one and a yellow one) were constructed specifically for the test trial of this study. Two cardboard boxes (a green one and an orange one) with an attached lid were used to hide these objects during the procedure. A pair of ordinary (transparent) sunglasses was used in the familiarization phase, and a different but similar looking pair of sunglasses was used in the test phase, which was either transparent (in the TB-TB condition) or opaque (in the TB-FB condition).

Procedure. The procedure was a modified version of the task introduced by Southgate et al. (2010).

Familiarization. Children were seated on the floor with their parent and were shown a pair of ordinary (transparent) sunglasses by Experimenter 1 to make sure that they were familiar with the object and its use. This phase lasted for about 30 seconds.

Warm-up trials. Experimenter 1 (E1) wore a pair of sunglasses on her head as if it was a ‘hairband’. She was kneeling in between the two cardboard boxes, which were 100 cm apart and 120 cm from the child. She gave the child the egg and the carrot, and they were allowed to play with them for roughly 10 seconds. E1 then took the objects, placed one in each box, and asked the child to retrieve one of the objects, followed by the other one (both referred by their name). This game continued until the child correctly chose the requested object twice in a row from two different boxes.

Test trial. The test trial consisted of three phases: a belief induction phase, a belief revision phase, and a test phase.

Belief induction phase. E1 gave the children the two novel objects, who were allowed to explore them for about 10 seconds. These objects were not labeled in this exploration phase. E1 then placed one object in each box and closed the lids. The location of the objects across the boxes was counterbalanced across infants. At this point, Experimenter 2 (E2) asked

E1 to put on her sunglasses. E1 then lowered her sunglasses onto her nose and sat back, but stayed in the room facing the subsequent events. E2 then deceptively approached the boxes (gesturing ‘shush’ towards the child, following the protocol of Southgate et al., 2010), switched the objects, closed the boxes, and asked E1 to leave the room with her for a while. Before leaving, E1 removed her pair of sunglasses and left it in front of the infant. Both experimenters then left the room.

Belief revision phase. At this point the child was encouraged by the parent to try on the sunglasses. Before the experiment, the parents had been told that, when they are left alone, they should ask their child to try on the sunglasses and verify together whether they could see through them. The parents had been informed in advance whether the sunglasses were opaque or transparent in order to avoid explicit signals of their own surprise or difference in their behavioral reactions in the two conditions. All the parents in the sample followed the instruction and ensured that their children noticed whether they could see through the sunglasses. The children could have drawn different conclusion from this experience in the two conditions: in the TB-TB condition the E1’s sunglasses were transparent, but in the TB-FB condition they were opaque.

Test phase. After being away for approximately 45 seconds, E1 returned to the room, greeted the infant, and sat on the floor behind the two boxes. E1 then pointed at one of the boxes (counterbalanced across infants) and said (in Hungarian), “Do you remember what I put here? I put a sefo here. Shall we play with the sefo?”, alternating gaze between the infant and the referred box twice. E1 then grasped both boxes, extended her arms towards the child and simultaneously opened the sides of both boxes to which the child was facing, without looking inside of them, whilst looking at the child. At this point, the contents of the boxes became visible only to the child, but not to E1. E1 then said to the infant, “Can you give me the sefo?”, while looking directly at the infant, and not looking to either box. E1 repeated the question until the child began to approach one of the boxes, pointed towards one of the boxes, or until 180 seconds had passed.

Coding. The sessions were video-recorded and coded off-line. The single binary depended measure of the experiment was the choice that children made in response to E1's request. The first response towards one of the boxes, after E1 had said, 'Can you give me the sefo?' was coded as the child's choice, and was categorized as choosing the referred or the non-referred box. Both reaching and pointing responses were accepted as valid choices. All sessions were coded also by a second observer, who was blind to the experimental condition, because only the recordings of the test phases of the experiments were available to her. Interrater agreement was 96% (Cohen's Kappa: 0.91.)

Results and Discussion

The number of infants who chose the referred and the non-referred box in each condition is depicted in Figure 1. Among 18-month-olds, 12 infants chose the referred box and 4 infants chose the non-referred one in the TB-TB condition. In the TB-FB condition, infants displayed a similar performance: 11 infants chose the referred box, while 4 infants chose the non-referred one. Among 36-month-olds, 16 participants chose the referred box and 2 chose the non-referred one in the TB-TB condition. Importantly, however, in the TB-FB condition, only 6 children chose the referred box, while 12 chose the non-referred one.

A 2x2 (age x condition) log-linear analysis revealed that the pattern of answers across the conditions differed in the two samples significantly: $G^2 = 13.98$, $df = 4$, $p < .01$. Follow-up Fisher's exact tests confirmed that the number children choosing the referred box differed significantly between the TB-TB and TB-FB conditions in the 36-month-old sample ($p = .002$, two-tailed), while there was no significant difference between the conditions in 18-

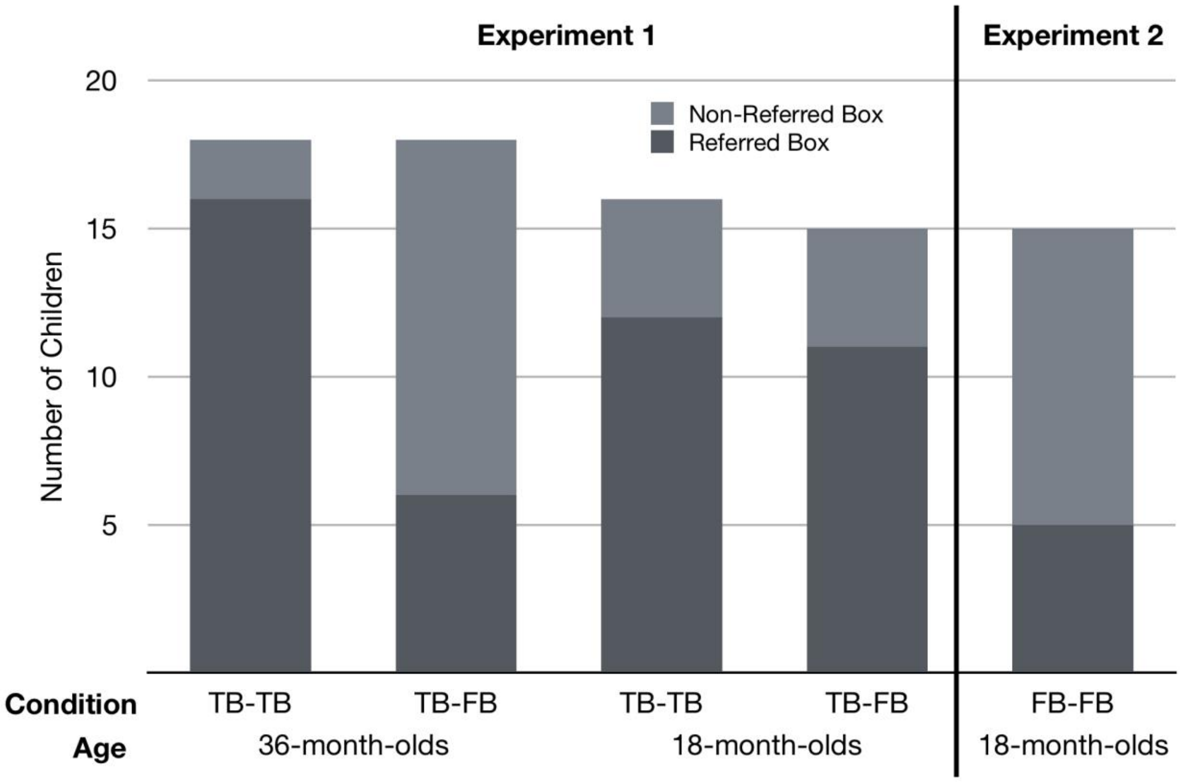


Figure 1. Number of children choosing the referred or non-referred box as a function of condition and age in Experiments 1 and 2.

month-olds ($p = 1.000$).

These results show no evidence that 18-month-old would have considered their experience with the sunglasses as relevant to their response to E1's request. However, 36-month-olds behaved differently in the two conditions, suggesting that they were able to identify that, in the TB-FB condition, the information revealed about the opacity of the sunglasses during belief revision phase was relevant for E1's belief state. Consequently, they recalled that E1 had been wearing sunglasses during the location change event, retrospectively re-computed her belief about the location of the objects, and used this information to respond to her request.

The lack of difference in performance between the two conditions in 18-month-olds does not point to the involvement of such a retrospective re-computation mechanism. Their failure might be due to prospectively maintaining the already attributed belief during the belief revision phase (i.e., attributing true beliefs to E1 in both conditions). However, one might argue that there can be a more parsimonious interpretation of the performance of this age group: infants may have simply followed the referential request of the model (by giving her the referred object in both conditions) and did not pay attention to her potential belief content in either condition. Experiment 2 investigated this question.

Experiment 2

Considering that Southgate et al. (2010) found that 17-month-olds resolved ambiguous referential request by appealing to the (true or false) belief state of the interlocutor, 18-month-olds failure in Experiment 1 was unlikely due to their inability to take into account false beliefs per se. However, it is possible that 18-month-olds did not consider the opacity of the sunglasses as causally relevant properties in assessing the belief states of the actor (though see Senju, Southgate, Snape, Leonard, & Csibra, 2011). In Experiment 2, we tested whether information about the opacity of the sunglasses *before* the encoding of E1's belief would lead infants to correctly and prospectively infer that she would have a false belief. If 18-month-olds pass this test, it would indicate that immaturity of belief revision mechanisms (rather than deficient belief attribution mechanisms) made them ignore this causally relevant information about the sunglasses in Experiment 1.

Methods

Participants. Twenty 18-month-old infants were recruited for this experiment. Because of experimenter error, 1 infant was excluded and was replaced by an additional participant. Out of the 20 infants, 5 did not make a choice during the test phase. The remaining 15 infants (mean age = 18.0 months, range = 17.5 - 18.5 months) made the final sample for this experiment.

Materials. The same props were used as in Experiment 1, with the exception of the sunglasses. A single *opaque* pair of sunglasses was used in this study.

Procedure. We created a new FB-FB condition by making infants know in advance that the sunglasses worn by E1 were opaque. The procedure used in this experiment was identical to Experiment 1 with the following exceptions. Infants were shown a pair of opaque sunglasses to explore during familiarization, and E1 wore the very same sunglasses during the belief induction phase. Then E1 left the room for about 45 seconds, leaving the pair of sunglasses behind. In this experiment, the parents were not instructed to explore the sunglasses with their child. Thereby the belief revision phase was skipped while the delay between the belief induction and test phases was kept identical to that of Experiment 1.

Coding. The coding protocol followed exactly the one in Experiment 1. The agreement between the primary coder and secondary blind coder was 93% (Cohen's Kappa: 0.87).

Results

In this experiment, only 5 infants chose the referred box, while 10 infants chose the non-referred one (Figure 1). The results from this experiment were compared with those of the two conditions in 18-month-olds from Experiment 1. A 2x3 chi-square test found that there was a difference in the pattern of choices between the three conditions (*chi-square* =

7.29, *df* = 2, *p* = 0.029). Follow-up Fisher's exact tests confirmed that the number of infants choosing the referred box differed significantly between the TB-TB and FB-FB conditions (*p* = 0.022), and also between the TB-FB and FB-FB conditions (Fisher's exact *p* = 0.033). This pattern of results suggests that, in contrast to Experiment 1, infants could take into account the (false) belief state of E1 when responding her request. This finding essentially replicates that of Southgate et al. (2010).

General Discussion

Tracking what others know and believe plays an important role in human communication because utterances have to be designed in production, and interpreted in comprehension, in the context of the mental states of one's interlocutor (Sperber and Wilson, 2002). It has been previously demonstrated that toddlers can disambiguate referential expressions by appealing to the belief content of the communicator, even when this belief is false (Southgate et al., 2010). We confirmed this finding in Experiment 2, where 18-month-old had information about the fact that the experimenter did not have perceptual access to a location change of the experimenter, and were able to use the derived state of the attributed belief in interpreting her referential expressions in the test phase. Indeed, this pattern of results provided further evidence that 18-month-olds were able to apply their self-experience (of the opacity of the sunglasses) to assess the experimenter's lack of visual access (see also Senju et al., 2011).

Importantly, however, in Experiment 1, 3-year-olds (but not 18-month-olds) displayed an additional capacity by demonstrating that they could rely on false belief attribution even without the opportunity to directly observe the communicator's lack of perceptual access generating her false belief. While false belief attribution in both Experiments 1 and 2 required inferential processes, 3-year-olds could not have succeeded in the task without invoking memories of specific past events into their inferences. In particular, at some point of the procedure, they must have recollected the location-swap event during the belief induction phase to combine this memory with the later acquired information about the sunglasses in order to be able to derive the belief content of the experimenter. Thus, the results of our study cannot be entirely explained by prospective attribution mechanisms. As such, these results prove the existence of retrospective belief attribution mechanisms in 3-year-olds. However, our findings clarify neither the nature of these mechanisms nor when exactly they operate during the events like the one we produced in Experiment 1. In our view, there are three alternative answers to these questions, differing in the assumptions as to when initial attribution takes place and what events trigger these attribution processes.

First, in principle it is possible that all responses in our tasks, or even in the majority of false-belief tasks, were based on purely retrospective mechanisms. If children possess sufficiently accurate mechanisms of recalling relevant details of past events, they could calculate the belief content of the communicator at the time when they need it; i.e., when they have to predict her action or interpret her request. According to this option, both 18-month-olds and 3-year-olds, and in all conditions, calculated the content of the relevant belief of the experimenter when she pointed to a box and asked for the ‘sefo’. This belief attribution process was the same in the TB-FB and TB-TB conditions, except that the perceptual access of the experimenter to the location-change event had to be evaluated differently depending on the subsequently acquired information concerning the transparency or opacity of the sunglasses. If this interpretation of our findings is correct, 18-month-olds’ failure in Experiment 1 was not due to memory limitation or to the absence of retrospective attribution mechanisms, but to an inferential deficiency: they did not recognize that the information they learnt about the sunglasses was relevant to the evaluation of the experimenter’s perceptual access to the earlier event that fixed the belief content. While on the basis of our study we cannot exclude this explanation, we find this option, which completely eliminates prospective attribution processes, unlikely: building a belief attribution system entirely on retrospective mechanisms would require highly reliable, fast and accurate episodic memory. Considering that even adults' episodic memory fails to meet these standards (Schacter, Guerin, & St. Jacques, 2011; Cochran, Greenspan, Bogart, & Loftus, 2016), and young children's ability to recall past events is much weaker than that of adults (Bauer & Leventon, 2013; Bauer, Wenner, Dropik, & Wewerka, 2000; Mullaly & Maguire, 2014), the wide range of findings on early belief attribution calls for mechanisms that do not exclusively depend on memory.

A more plausible alternative way to think about retrospective attribution mechanisms is that they always work on, and modulate, already attributed (true or false) beliefs[2]. According to this option, the primary mechanism of belief attribution is the prospective route: children always attribute (true) beliefs when they observe an agent's perceptual access to some relevant facts, and then maintain these attribution across events, updating the belief content when it is necessary. Thus, children in the standard location-change false-belief tasks (and 18-month-olds in Experiment 2) initially attribute a true belief of object location to the protagonist (the experimenter in our study), and then maintain this belief attribution when the content of the belief becomes false in the absence of the protagonist (or in the lack of perceptual access of the experimenter, wearing opaque sunglasses, in Experiment 2). In the present context, when 3-year-olds learnt about the opacity of the sunglasses in the TB-FB condition of Experiment 1, and subsequently re-computed the content (and/or the status) of the experimenter's belief about the location of the objects, they did not attribute a new belief to her but rather *revised* a prospectively attributed belief. In order to be able to do so, they must have stored not only the content of this belief but also some relevant facts about the

[2] This interpretation corresponds to the terminology that we adopted in the description of the procedure and the conditions of Experiment 1.

source of this attribution (i.e., E1's perceptual access to object location) and history of updates (e.g., when location change occurred within the stream of events). Then either at the point when they found out that the sunglasses were opaque, or at the point when they had to evaluate the referential expression of the experimenter, they retrospectively revised the source information and the update history of the attributed belief in light of what they learnt about the sunglasses, and re-calculated the content of the experimenter's belief accordingly. They did not have to perform such a revision in the TB-TB condition. This interpretation of retrospective attribution suggests that 18-month-olds' failure in Experiment 1 was due to the inaccessibility of source and update information either because of failing to retrieve it from memory or because of failing to encode it in the first place. In either case, the toddlers were unable to revise their prospective attribution and erroneously relied on it in the interpretation of the referential expression during the test phase.

A third possible way of interpreting the role of retrospective attribution system in theory of mind is to link it to belief update mechanisms that operate on own beliefs. According to this option, prospective true belief attribution is not mandatory, and both prospective and retrospective attribution mechanisms are triggered when the contents of own beliefs are updated or revised. In the standard, location-change false-belief task, the informational access of the protagonist to the location of the marble does not have to be recorded by creating a separate (meta-)representation; it can be represented by simply tagging the child's own representation of the location by the protagonist, indicating that she has access to this information (cf. Martin & Santos, 2016). When the marble is relocated and content of this representation is updated accordingly, the tagging is then also updated. If the protagonist has perceptual access to the change, the tagging is maintained on the representation; if she does not have access, a new representation is created and attributed to her with the old content. This latter process is a prospective attribution of a false belief, triggered not by perceptual access but the lack of it. Such prospective FB attribution would explain 18-month-old success in the FB-FB task of Experiment 2 (as well as in Southgate et al., 2010). However, in Experiment 1, children had to update their representation of current reality not only when the locations of the objects across the boxes were swapped but also when they learnt that the sunglasses, which they initially believed to be transparent, were opaque. This revision process might have triggered the search for additional representations, linked to the updated information, to be revised. Indeed, it has been suggested that one function of episodic memory is to allow us to revise our beliefs on the basis of new information related to the original source of those beliefs (Klein et al., 2009). Such a search might have led 3-year-olds to memories related to the person who had worn the sunglasses, and allowed them to retrospectively re-evaluate her perceptual access to the location change event. As a consequence, they could remove the tagging of the experimenter from their own representation of true reality, and could create a new representation by attributing to her the false belief with the content of the location of the objects before the swapping took place. Such belief attribution would thus have both retrospective and prospective elements, and while it is not a revision of an already attributed belief, it is triggered by the revision of own beliefs. In this interpretation of the results, the 18-month-olds might have failed to attribute a false belief in Experiment 1 because their search for to-be-revised information related to the opacity of the sunglasses did not lead them to the memory of the particular event during which the experimenter had worn the sunglasses and they had had to update their own representation of reality. In other words, weak or unreliable memory traces, or immature recollection processes could explain their failure.

The question of which of these alternative explanations explains our findings best is beyond this paper and will have to await for further investigations. However, our study demonstrates that retrospective belief attribution mechanisms are available to children from at least 3 years of age, and that purely prospective mechanisms are not sufficient to explain the host of findings in the relevant literature on mindreading. In addition, this pattern of results suggests that updating and revision of attributed beliefs relies on flexible manipulation of representations and metarepresentations even in so-called ‘implicit’ ToM tasks. Such a conclusion is inconsistent with views that explain young children’s performance in such tasks by associative learning (Perner & Ruffman. 2005) or by tracking agent-object relations (Butterfill & Apperly. 2013).

Furthermore, our study provides evidence not only for specific attribution mechanisms in 3-year-olds but also for the functioning of episodic retrieval processes that are necessary for retrospective belief revisions. This finding is consistent with a growing body of evidence that, when children’s memory of the original event is controlled for, 3-year-old children perform well on various versions of the so-called ‘spoon test’ (Atance & Sommerville, 2013; Scarf, Gross, Colombo & Hayne, 2013; Suddendorf, Nielsen & von Gehlen, 2011), which was suggested to be the litmus test of episodic time travel (Tulving, 2005). For example, in the study of Scarf et al. (2013), 3- and 4-year-old children had the opportunity to dig up a treasure case in a sandbox. When they uncovered the treasure case, children found that it was locked. Later children returned to the lab and were asked to select one of three items (a key and two distractor objects) to take to the sandbox with them. Three-year-olds performed above chance if the delay between the events was 15 minutes or less: they tended to take the key with them (4-year-olds performed above chance even with a 24-hour delay). While the spoon test demonstrates that children can identify and select information as relevant for an upcoming or reoccurring event based on a memory of a past event, this achievement can be based on the encoding of some semantic information extracted from that past event (locked treasure case in the sandbox). This information is then recalled when some related novel information (key for unlocking) is obtained in a context that promises revisiting the original scene. Thus, this task does not necessitate episodic retrieval.

In contrast, our paradigm required children to infer the episodic common ground with the interlocutor, which could not have been achieved by recalling semantic information about her, about sunglasses, or about object location. In other words, having learnt that the sunglasses were opaque, children had to retrieve the specific event when the model had worn those glasses, and only within the frame of the original episode they could infer the consequences of not seeing the location change. This retrospective attribution process, we suggest, is in line with the proposal that episodic memories enable updating the interpretation and inferential consequences of a past event in light of newly acquired information (Klein et al., 2009). Our study thus has provided so far the clearest evidence that 3-year-old children can recollect events (at least within a minute delay) in sufficient details to be used for inferences in social interactions.

References

Atance, C. M. & Sommerville, J. A. (2014). Assessing the role of memory in preschoolers' performance on episodic foresight tasks. *Memory, 22*(1), 118–128. DOI:10.1080/09658211.2013.820324

Baron-Cohen, S., Leslie, A.M., & Frith. U. (1985). Does the autistic child have a 'theory of mind'? *Cognition, 21*(1): 37–46. DOI:10.1016/0010-0277(85)90022-8

Bauer, P. J. & Leventon, J. S. (2013). Memory for one-time experiences in the second year of life: implications for the status of episodic memory. *Infancy, 18*(5), 755-781.

Bauer, P.J., Wenner, J.A., Dropik, P.L., & Wewerka, S.S. (2000). Parameters of remembering and forgetting in the transition from infancy to early childhood. *Monographs of the Society for Research in Child Development*, *65*, Serial No. 263.

Boyd R., & Richerson P. J. (1996). Why culture is common, but cultural evolution is rare. *Proceedings of the British Academy, 88*, 77–93.

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

Butterfill, S., & Apperly, I. (2013). How to construct a minimal theory of mind. *Mind & Language, 28:* 606– 637.

Cochran, K.J., Greenspan, R.L., Bogart, D.F., & Loftus, E. F. (2016). Memory blindness: altered memory reports lead to distortion in eyewitness testimony. *Memory & Cognition*, *44*(5), 717-726.

Eacott, M.J., & Crawley, R.A. (1998). The offset of childhood amnesia: memory for events that occurred before age 3. *Journal of Experimental Psychology: General, 127*(1), 22-33.

Fivush, R. & Nelson, K. (2004). The emergence of autobiographical memory: A social cultural developmental theory. *Psychological Review, 111*(2), 486–511. DOI:10.1037/0033-295X.111.2.486

Hayne, H. (2004). Infant memory development: Implications for childhood amnesia. *Developmental Review, 24*, 33–73. DOI:10.1016/j.dr.2003.09.007

Klein, S.B., Cosmides, L., Gangi, C.E., Jackson, B., Tooby, J. & Costabile, K.A. (2009). Evolution and episodic memory: An analysis and demonstration of a social function of episodic recollection. *Social Cognition, 27*(2), 283-319.

Knudsen, B., Liszkowski. U. (2012). 18-Month-Olds predict specific action mistakes through attribution of false belief, not ignorance, and intervene accordingly. *Infancy* 17: 672–691.

Kovács, Á.M. (2016). Belief files in theory of mind reasoning. *Review of Philosophy and Psychology, 7*(2), 509-527. DOI: 10.1007/s13164-015-0236-5

Kovács, Á., Téglás, E., & Endress, A. (2010). The social sense: susceptibility to others' beliefs in human infants and adults. *Science, 330*, 1830-1834.

Martin, A. & Santos, L. (2016). What cognitive representations support primate theory of mind? *Trends in Cognitive Sciences*, *20*(5), 375-82. DOI: 10.1016/j.tics.2016.03.005

Mullaly, S.L. & Maguire, E.A. (2014). Learning to remember: the early ontogeny of episodic memory. *Developmental Cognitive Neuroscience, 9*, 12-29.

Onishi, K. & Baillargeon, R. (2005). Do 15-month-olds understand false beliefs? *Science, 308*, 255-258.

Perner, J. & Ruffman, T. (2005). Infants' insight into the mind: how deep? *Science*, *308*, 214-216.

Rubio Fernandez, P. (2013). Perspective tracking in progress: Do not disturb. *Cognition, 129*, 264–272.

Scarf, D., Gross, J., Colombo, M., Hayne, H. (2013). To have and to hold: Episodic memory in 3- and 4-year-old children. *Developmental Psychobiology, 55*(2), 125–132. DOI: 10.1002/dev.21004

Schacter, D.L., Guerin, S.A., & St. Jacques, P. (2011). Memory distortion: an adaptive perspective. *Trends in Cognitive Sciences, 15*(10), 467-474.

Senju, A., Southgate, V., Snape, V., Leonard, M., & Csibra, G. (2011). Do 18-month-olds really attribute mental states to others? A critical test. *Psychological Science*, *22*, 878–880.

Southgate, V., Chevallier, C., & Csibra, G. (2010). Seventeen-month-olds appeal to false beliefs to interpret others' referential communication. *Developmental Science, 13*, 907–912.

Southgate, V., Senju, A., & Csibra, G. (2007). Action anticipation through attribution of false belief by 2-year- olds. *Psychological Science, 18*, 587–592.

Sperber, D. & Wilson, D. (2002). Pragmatics, modularity, and mind-reading. *Mind & Language, 17*, 3-23.

Suddendorf, T., Nielsen, M., & von Gehlen, R. (2011). Children's capacity to remember a novel problem and to secure its future solution. *Developmental Science, 14*(1), 26–33. DOI:10.1111/j.1467-7687.2010.00950.x

Surian, L., Caldi, S., & Sperber, D. (2007). Attribution of beliefs by 13-month-old infants. *Psychological Science, 18*, 580–586.

Tulving, E. (2005). Episodic memory and autonoesis: Uniquely human. In H. Terrace & J. Metcalfe (Eds.), *The Missing Link in Cognition* (pp. 4–56). New York: Oxford University Press.

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

Consciousness and Cognition 41 (2016) 93–103

Contents lists available at ScienceDirect

Consciousness and Cognition

journal homepage: www.elsevier.com/locate/concog

Evidence for spontaneous level-2 perspective taking in adults

Fruzsina Elekes [a,b,c,*], Máté Varga [d], Ildikó Király [b,e]

[a] *Doctoral School of Psychology, Eötvös Loránd University, Izabella str. 46, Budapest 1064, Hungary*
[b] *Institute of Psychology, Eötvös Loránd University, Izabella str. 46, Budapest 1064, Hungary*
[c] *Cognitive Development Center, Central European University, Október 6. str. 7, Budapest 1051, Hungary*
[d] *Department of Cognitive Science, Budapest University of Technology and Economics, Műegyetem rkp. 3., Budapest 1111, Hungary*
[e] *Institute for Advanced Study, Central European University, Október 6. str. 7, Budapest 1051, Hungary*

ARTICLE INFO

Article history:
Received 10 December 2015
Revised 11 February 2016
Accepted 12 February 2016

Keywords:
Spontaneous level-2 perspective taking
Aspectuality
Awareness
Social partner
Two systems approach in ToM

ABSTRACT

Social interactions are fostered by humans' propensity to compute their partner's perspective online. However, due to the mindreading system's limited capacity perspective taking (PT) was argued to occur spontaneously only for level-1, but not level-2 perspectives. We propose that level-2 perspectives (containing aspectual information) can also be computed spontaneously if participants have reason to assume that the partner is indeed aware of the objects' aspectual properties. Pairs of adult participants took part in the modified version of Surtees, Butterfill, and Apperly's (2012) number verification paradigm. Participants had prior information on their partner's task, which either called for processing aspectual properties or did not. The partner's inconsistent perspective was found to interfere with RT-s providing evidence for spontaneous level-2 PT. However, such interference only occurred when the partner's task involved processing the perspective dependent object feature, suggesting that PT was sensitive to the other's awareness of the to be represented information.

1. Introduction

Visual perspective taking refers to the ability to mentally map how a certain scene looks from another person's point of view. Being part of the mind reading system, perspective taking (PT) provides the basis for attributing knowledge or beliefs to others and thus lays the foundation for smooth social interactions (Aichhorn, Perner, Kronbichler, Staffen, & Ladurner, 2006; Apperly, 2008; Wimmer, Hogrefe, & Perner, 1988). While its relevance is widely recognized, the features and functioning of PT are strongly debated. In recent literature it has been argued that visual perspective taking is not a unitary capacity either in terms of the computed representation, or regarding the mechanism that leads to that representation (Apperly & Butterfill, 2009; Flavell, Everett, Croft, & Flavell, 1981; Rakoczy, 2012). The opposition to this view claims that there is only one mindreading system that, at times, recruits other cognitive faculties as well, but uses the same concepts regardless (Carruthers, 2015a). Our findings indicate that the division between mindreading systems is not as rigid as proposed by the former approach.

Based on empirical findings in preschoolers, Flavell et al. (1981) proposed that two types of information could be achieved regarding the visual perspective of others. Level-1 PT refers to representing *whether* an agent can see an object, while level-2 PT means representing *how* exactly that object appears to him, that is, under what aspect the agent sees the object. This distinction indicates that there is a qualitative gap between knowing what the other does and does not see, and being able to

* Corresponding author at: Doctoral School of Psychology, Eötvös Loránd University, Izabella str. 46, Budapest 1064, Hungary.
E-mail address: elekes.fr@gmail.com (F. Elekes).

http://dx.doi.org/10.1016/j.concog.2016.02.010
1053-8100/

represent the scene as it is visible to him/her. Underlying this notion it has been demonstrated that the ability to compute someone else's level-1 perspective develops earlier in life than the representationally more complex capacity of level-2 PT (Moll & Meltzoff, 2011; Moll & Tomasello, 2006; Sodian, Thoermer, & Metz, 2007).

Despite the bias often shown in children and adults towards egocentrism (for a review see, Samson & Apperly, 2010), there is evidence that level-1 perspective taking can emerge in a speeded way and without instruction to do so (Samson, Apperly, Braithwaite, Andrews, & Bodley Scott, 2010). In the number verification paradigm used by Samson et al. (2010), participants had to verify the amount of dots presented on the walls of a virtual room either from their own perspective, or from the perspective of an avatar located in the room. The amount of dots visible to the avatar either matched or did not match that visible to the participant, potentially creating a conflict with one's own perspective.

The results suggested (Samson et al., 2010) that adults computed the avatar's perspective online despite the fact that it was irrelevant for decision-making. This, in turn, interfered with participants' decisions when the avatar's perspective was inconsistent with their own (altercentric intrusion). Importantly, similar altercentric interference emerged when participants only had to make judgments based on their own perspective throughout the experiment. This rules out the possibility that the high inhibition demands of switching back and forth between perspectives, or the situation that trained participants to place themselves into the other's perspective played a role in the effect. Similar to adults, school-aged children also showed altercentric intrusions in this paradigm (Surtees & Apperly, 2012). Finally, computation of the other's perspective was found to be independent of parallel cognitive load, indicating that the process was indeed cognitively efficient (Qureshi, Apperly, & Samson, 2010).

Unlike level-1 PT, level-2 PT has not been reported to occur in a spontaneous way. Surtees, Butterfill, and Apperly (2012) presented subjects with single numerals that were either symmetric/unambiguous in nature (0,8) or asymmetric/ambiguous (6,9). The numbers were presented either lying on the table between the participant and the avatar (asymmetric stimuli looked different from the two perspectives), or were displayed on the wall (all stimuli looked the same independent of perspectives). Participants had to perform a number verification task from their own or the avatar's perspective. In this case, the avatar's inconsistent perspective did not interfere with egocentric perspective judgments, suggesting that adults did not compute how the scene looked from the avatar's perspective spontaneously.

Before outlining current views on the cause of the above difference, an important distinction has to be drawn between automatic and spontaneous processing. Although there are many different approaches to automaticity, an automatic cognitive process is thought to be independent of both the participant's overt goal, and of any covert goals he might have (Carruthers, 2015a). On the other hand, while spontaneous processes are also independent of overt goals or external prompting, they do depend on participants' covert goals (for example, the general motivation to understand others, Carruthers, 2015a), or on contextual factors (Back & Apperly, 2010).[1] Samson et al.'s (2010) findings were interpreted as evidence for the "relatively automatic" computation of level-1 perspectives (Qureshi et al., 2010; Surtees & Apperly, 2012). For the sake of definitional clarity we will continue to refer to these results as spontaneous PT.

The difference in spontaneity of computation found between level-1 and level-2 PT might bring us closer to understanding the mechanism behind these abilities. It has been argued that the two skills, level-1 and level-2 PT, differ in terms of their reliance on perspective computation, and relatedly, in the degree to which they are social in nature. According to some, the ability to figure out which objects someone does or does not see (level-1 PT) might not even require reasoning about perspectives at all (Aichhorn et al., 2006; Michelon & Zacks, 2006). Aichhorn and colleagues (2006) argue that differing perceptual experiences have to refer to the same objects or scene in order to qualify as *perspectives* on those objects, while in level-1 decisions the different percepts can be interpreted as resulting from a difference in the objects that are looked at. Without using the concept of seeing, perceptual access to an object can be judged based on the spatial relation between the other person's eyes and the target object (Aichhorn et al., 2006; Michelon & Zacks, 2006). Empirical findings support this proposal. Adults are quicker to make explicit level-1 decisions when the avatar is close to the target object and are slower when the distance is greater, but the speed of computation is not affected by the angular disparity between participant and avatar (Michelon & Zacks, 2006; Surtees, Apperly, & Samson, 2013a). This indicates that the information (see/does not see) is reached through tracing the person's line of sight, which line takes longer to "draw" if the path is longer.

The idea that tracking visual access to certain objects does not involve representing the perspectives of social agents gains further support from a different line of investigation as well. Studies show altercentric interference in Samson et al.'s (2010) number verification task also when the avatar is replaced by a less or non-social, directional stimulus (Nielsen, Slade, Levy, & Holmes, 2015; Santiesteban, Catmur, Hopkins, Bird, & Heyes, 2014). Based on this, Santiesteban et al. (2014) argue that the phenomenon referred to as level-1 perspective taking is driven by domain general factors, like attentional cueing, rather than Theory of Mind. The effect in Nielsen et al.'s (2015) study was, however, stronger in the social condition compared to the less and non-social conditions. Furthermore, the effect correlated with self-reported measures of Theory of Mind in the social, but not in the other two conditions. This suggests that processes specific to the social domain also contribute to spontaneous level-1 PT.

As opposed to tracing someone's line of sight, representing appearances from another point of view, that is taking someone's level-2 perspective, presumes understanding that the same object from a different angle may give rise to different percepts. Hence, seemingly contradictory contents regarding the same referent (e.g. the object's perceived identity) may all be

[1] A related term is involuntariness (Bargh, 1989). While automaticity/spontaneity refers to the features of launching a process, involuntariness of computation indicates that a process will necessarily be performed to the end if started, it cannot be down-regulated or controlled even if the perceiver is aware of its' operation.

true, given they belong to different perspectives (Perner, Stummer, Sprung, & Doherty, 2002). Accordingly, when making explicit perspective decisions, adults were influenced by the angular disparity between their own and the avatar's perspective, responding to greater angles more slowly (Surtees, Apperly, & Samson, 2013b, 2013a). This indicates that the mechanism behind explicit level-2 PT resembles mental rotation, where participants mentally rotate themselves in space to align to the other's perspective (Surtees et al., 2013b). This process builds heavily on working memory and is thus cognitively demanding in nature.

The above difference between the characteristics of level-1 and level-2 PT is in line with Apperly and Butterfill's (2009) two systems account, which postulates that the mindreading system, and perspective taking as part of it, consists of two kinds of processes. One process is cognitively demanding and slow but flexible in nature, while the other is quick and efficient at the expense of being inflexible. The latter is called the Minimal ToM and it enables the online guidance of behavior in social contexts (Apperly & Butterfill, 2009; Butterfill & Apperly, 2013). However, the quick and efficient system has a content specific limit, being unable to represent representations *as such*. Instead, it uses certain principles to track mental states without actually representing them. Briefly, the online mindreading system can track *relations* between agents and objects (i.e. level-1 perspectives), but it fails when information on object *appearance* (i.e. level-2 perspectives), or propositional mental states have to be computed.

Apperly and Butterfill's (2009; Butterfill & Apperly, 2013) theory has been criticized heavily recently (Carruthers, 2015a, 2015b; Christensen & Michael, 2015). Carruthers (2015a, 2015b) argues that the data can be accounted for by assuming the existence of one mindreading system that can work automatically, spontaneously, or deliberately depending on context. Its functioning is automatic (or at least spontaneous) whenever mental state attribution does not necessitate the use of other cognitive capacities, like working memory or executive functions. Whenever other cognitive faculties have to be recruited (like the working memory requirement of mental rotation in the case of level-2 PT), it is the encoder's explicit goal that determines whether he will allocate cognitive resources into computing the other's mental states. Hence, computation in this case will not be automatic. Importantly, in opposition to Apperly and Butterfill (2009) and Butterfill and Apperly (2013), the model does not assume a qualitative difference between representations attributed automatically and intentionally.

In addition to its sensitivity to the cognitively demanding nature of perspective computation, the mindreading system has a further restriction according to Carruthers. Namely, that due to its limited capacity, the system cannot provide a complete model of the *whole* visual scene. Perspective attributions are only made for those objects that the other person is aware of (Carruthers, 2015a). Awareness of certain objects is indicated by, for example, the other's gaze direction. Thus, the one system model of mindreading predicts that automatic computation of another's perspective will occur whenever that does not require executive functions, *and* appropriate guidance is available for the encoder to judge the other's awareness of the target stimuli. The lack of automatic PT might be caused by either or both of these factors.

For the sake of argument, let us assume that under certain conditions (say, sufficient background knowledge on the object's aspectual nature to enable shortcutting the mental rotation process), level-2 PT can be achieved without exerting a high load on working memory. Under these conditions, the factor that determines if the mindreading system will be engaged automatically is what the encoder knows about the other's awareness of the stimuli. Carruthers (2015a) argues that for level-1 PT the mindreading system selects *which objects* to make attributions about. However, even if the same concept of seeing is used for level-1 and level-2 PT, as hypothesized by Carruthers (2015a, 2015b), level-2 PT involves representing more detailed information (the object's appearance from another aspect) than level-1 PT. We propose that in the case of level-2 PT, the partner's awareness is taken into account on the level of object *features*. Hence, cues indicating that the other is aware of the *presence* of an object might not be sufficient to launch the process.

Merely looking in one direction does not obligatorily result in perceiving objects in that trajectory (Mack & Rock, 1998) and more specifically, looking at an object in space does not necessarily lead to consciously perceiving all of its features (Levin & Simons, 1997). For example, one might look at an object that has perspective dependent properties, e.g. form, and attend only to its non-perspective dependent features, e.g. texture or color. Importantly, adults' expectations have been found to be sensitive to the distinction between looking in one direction and perceiving information present there (Teufel & et al., 2009). Based on these considerations, we assume that, as a precondition of spontaneous level-2 perspective taking, the other's awareness of the stimuli is taken into consideration on the level of object *features*, not objects in general. Limiting perspective computation to the information that the partner is aware of is a beneficial strategy, as the partner's behavior will most likely be guided by that subset of information.

Attention is one of the cognitive processes that modulates conscious perception by highlighting certain bits of visual information and filtering out others (Broadbent, 1958); and while it might not be sufficient in itself, attention certainly is necessary for information to reach awareness (Cohen, Cavanagh, Chun, & Nakayama, 2012). Importantly, when reasoning about others, knowledge about the partner's goal can be used to determine what pieces of information he/she pays attention to, which, in turn, is a reasonably good indicator of what the partner is aware of out of all of the information that he/she has perceptual access to.

We hypothesized that if adults had prior information that the partner was not only looking at an object but also had awareness of the aspectual, perspective dependent properties of that object, they would compute the partner's level-2 perspective spontaneously. On the other hand, no PT was expected if participants' had no reason to presume awareness of those features.[2] The modified version of Surtees et al.'s (2012) number verification paradigm was used to test the proposal. Instead of using avatars as "social" partners, pairs of participants took part in the experiment sitting on opposite sides of a short table on which stimuli were displayed. Subjects participated in one of two groups. In the perspective-dependent (PD) group, both participants performed a number verification task. This ensured that the participant knew that the partner was attending to and was consequently aware of the object's form, that determines the number it represents, and potentially looks different from different aspects. In the non-perspective-dependent (NPD) group, however, the participant performing the number verification task knew that the partner's task (an n-back task) did not entail encoding aspectual information about the object.

We predicted interference from the other's inconsistent perspective for mutually attended stimuli in the PD group, as there participants had knowledge that the partner was necessarily aware of the aspectual object property. On the other hand, no interference was expected in the NPD group, due to the lack of evidence of the confederate's awareness of perspective dependent properties. Interference in the current study will be interpreted as an indicator of spontaneous, rather than automatic PT, as it can be argued that a live interaction makes the computation of the partner's perspective relevant without external prompting to do so.

2. Study 1

2.1. Method

2.1.1. Design

A 2 × 2 × 2 (Jointness [individual, joint] × Symmetry [symmetric number, asymmetric number] × Task [PD, NPD]) factorial design was used with Jointness and Symmetry as within subject factors and Task as a between subject variable. We used a within subject individual control instead of the "wall" trials of Surtees et al. (2012), as we placed participants into real 3D space, where that could not be implemented.

2.1.2. Participants

Data was collected from 54 university students (47 females, $M_{age} = 21.85$, $SD = 4.56$). Participant pairs were randomly assigned to groups. Eighteen participants were assigned to the PD group and 36 to the NPD group. As our question was whether the other's inconsistent perspective interfered with number verification we did not analyze data from the *n*-back task. Thus, 18 of the 36 people in the NPD group merely served as pairs for the participants. Participation was rewarded by course credit. The experiment was approved by the ethical committee of our university. All participants signed an informed consent form before starting the experiment.

2.1.3. Materials

Following Surtees et al. (2012) four numbers, two symmetric (0,8) and two asymmetric (6,9), were drawn using Matlab R2013a. The style of the visual stimuli resembled the "digital" numbers (for an illustration see, Table 1). The numbers were displayed in two different widths and two different colors (blue and green) which were varied randomly to increase the overall perceptual variability. The numbers were 11.5 cm in height, and 5.2 or 6.2 cm in width on a 21.5 in. flat monitor.

The audio stimuli were presented in a female voice with neutral and descending intonation. In the case of both symmetric and asymmetric numbers one of the two words contained one syllable while the other contained two syllables in Hungarian, the native language of the participants. The length of the audio stimuli was 760 ms for "eight" (nyolc), 760 ms for "zero" (nulla), 600 ms for "six" (hat), and 760 ms for "nine" (kilenc). The audio and visual stimuli pairings provided four types of events defined by the symmetry of the visual stimuli and correspondence between visual and audio stimuli (for details, see Procedure).

2.1.4. Procedure

Upon arriving to the lab, participants were introduced to each other and were seated next to each other in the reception room to sign the informed consent form and listen to instructions. The two groups, PD and NPD tasks, differed only in the instructions that were given (see, Appendix). In the PD group participants had to decide whether the number heard was the same as the number seen (number verification task). In the NPD group, one of the participants performed the number verification task, while the other had to decide whether the color of the stimulus currently on screen was the same as it was of the previous stimulus (*n*-back task). In contrast to Surtees et al. (2012), in the current experiment participants only had to make decisions based on their own perspective (self trials) to make the expected interference effect a stringent measure of perspective taking, rather than biased by the additional effects of task-switching.

After completing the consent forms, the experimenter escorted participants to the test room and seated them at opposite sides of a short table, facing each other. A 21.5 in. flat monitor was used to provide stimuli for both participants, and was laid

---

[2] Note, that spontaneous PT is necessarily limited to situations where there is a possibility to compute the other's level-2 perspective without exerting a high load on working memory or executive functions (Carruthers, 2015a). We argue that this is possible if one has pre-existing conceptual knowledge on the object's aspectual properties.

Table 1
Mean RT-s in milliseconds (with Standard Deviations in parentheses) in Study 1 and 2 as a function of Jointness, Symmetry, Correspondence and Task.

	Individual				Joint			
	Symmetric		Asymmetric		Symmetric		Asymmetric	
	Corres	Noncorres	Corres	Noncorres	Corres	Noncorres	Corres	Non corres
Study 1 (0/8 6/9)								
PD	685.2 (109.97)	752.91 (129.11)	677.85 (135.9)	692.07 (122.04)	696.69 (116.03)	772.16 (112.09)	730.62 (134.15)	740.48 (102.12)
NPD	704.46 (128.31)	786.13 (152.66)	724.45 (138.38)	720.32 (155.44)	713.56 (160.61)	795.54 (190.39)	729.74 (182.12)	738.26 (174.65)
Study 2 (0/8 6/9)								
PD	684.77 (121.25)	757.58 (122.87)	694.76 (135.08)	700.99 (129.74)	700.26 (92.28)	785.53 (94.1)	737.34 (132.96)	753.8 (118.41)

Note: The numbers (0/8, 6/9) are examples of the visual stimuli used in different widths and colors. PD refers to the perspective-dependent, NPD refers to the non-perspective-dependent group.

on the table between them. This way, symmetric stimuli (0, 8) looked the same irrespective of perspective, while asymmetric stimuli (6, 9) looked different from the two opposing perspectives. The audio stimulus was displayed through a pair of loudspeakers placed at equal distance from the two participants.

The script that was used for stimulus presentation and response recording was written in PsychoPy 1.81. Each trial started with a fixation cross presented at the center of the screen. The onset of the audio stimulus was 500 ms after the fixation cross appeared. The visual stimulus was presented after 300 ms following the onset of the audio stimulus and remained on the screen as long as the program received an answer (button press) from one or both participants depending on condition. Half of the participants used their right hand to indicate "yes" answers, while half used their left hand. Responses were made on Cedrus type response boxes.

Surtees et al.'s (2012, p. 79) presented the picture right after the audio ended ("participants viewed successive fixation stimuli ... followed by a 1800 ms auditory stimulus ... and then the test picture"). The aim of the shorter stimulus onset asynchrony in the current experiment (the visual image appeared before the audio ended) was to make it less likely that participants could form a mental image of the number presented in the auditory modality before they saw the picture. This mental image could then be matched in visual features with the visually presented number, without actually deciding what number the visual character depicted. We call this the visual matching shortcut. Using this strategy, participants would not engage in the number verification decision that is expected to be influenced by the other's perspective ("What number is depicted on the screen?").

The experiment consisted of three main test phases and an individual practice block containing 16 trials for both members of the pair. The test started with 16 practice trials for participant A, followed by the individual condition for the same participant. After this, participant B took part in the practice phase, followed by the joint condition. Finally, participant B proceeded with the individual condition. This setup ensured that half of the participants had the individual trials first, while the other half the joint trials. Each test phase was divided into two blocks with a 30-s-long rest phase between them. For the individual blocks, when only one of the participants performed the task, the other turned 180°, sitting with her back to the participant.

All three test phases contained a total of 112 trials, half of those depicted symmetrical visual stimuli (0, 8), half depicted asymmetrical visual stimuli (6, 9). Visual stimuli were paired with either corresponding or noncorresponding audio, resulting in the following four event types: symmetric corresponding (0 – "zero" and 8 – "eight"), symmetric noncorresponding (0 – "eight" and 8 – "zero"), asymmetric corresponding (6 – "six", 9 – "nine"), and asymmetric noncorresponding (6 – "nine", 9 – "six").

2.1.5. Data analyses

Only data collected from the number verification task was analyzed that is, both participants' responses in the PD group and one participant's response from each pair in the NPD group. When analyzing RT-s, incorrectly answered trials and outlier data points were removed. An outlier data point was defined as an RT that differed by more than two standard deviations from the mean RT of the given participant. The percentage of correctly answered trials (hit rate) was also calculated and analyzed.

2.2. Results and discussion

2.2.1. Mean reaction times

We expected to find a Jointness $\times$ Symmetry interaction for both corresponding and noncorresponding trials in the PD, but not in the NPD group. As a first step, we performed a $2 \times 2 \times 2 \times 2$ mixed ANOVA (Jointness [individual, joint] $\times$

Symmetry [symmetric, asymmetric] × Correspondence [corresponding, noncorresponding] × Task [PD, NPD]) on the mean RT-s (see Table 1, and Fig. 1) with Task as a between subject factor. A Jointness × Symmetry × Task interaction would confirm our predictions and allow us to run separate ANOVA-s for the two experimental groups.

The main effect of Correspondence, $F(1,34) = 19.88$, $p < .0001$, $\eta_p^2 = .369$, and Symmetry, $F(1,34) = 32.71$ $p < .0001$, $\eta_p^2 = .490$, reached significance, with shorter RT-s on corresponding and asymmetric trials. The main effect of Jointness approached significance, $F(1,34) = 3.22$, $p = .082$, $\eta_p^2 = .087$, responses being quicker on individual trials. We found a significant Jointness × Symmetry interaction, $F(1,34) = 7.97$, $p = .008$, $\eta_p^2 = .190$. As expected this was further qualified by a significant three way interaction between Jointness, Symmetry and Task, $F(1,34) = 6.18$, $p = .018$, $\eta_p^2 = .154$. Additionally, the interaction of Symmetry and Correspondence was also significant, $F(1,34) = 16.62$, $p < .0001$, $\eta_p^2 = .328$. The Symmetry × Correspondence interaction was specifically caused by the longer RT-s for symmetric noncorresponding stimuli and importantly it emerged independently of both Task and Jointness. This indicates that whatever caused the Symmetry × Correspondence interaction it did not interfere with the specific comparisons that were targeted.

To further clarify these data and to test if the expected RT pattern had emerged for both the corresponding and the noncorresponding trials 2 × 2 (Jointness [individual, joint] × Symmetry [symmetric, asymmetric]) repeated measures ANOVA-s were performed for these in both groups separately. In line with our predictions, in the PD group the Jointness × Symmetry interaction was significant for the corresponding trials, $F(1,17) = 7.41$, $p = .014$, $\eta_p^2 = .304$, and also approached significance for the noncorresponding trials, $F(1,17) = 3.92$, $p = .064$, $\eta_p^2 = .187$. Additionally, the main effect of Jointness was marginally significant for both the corresponding, $F(1,17) = 3.62$, $p = .074$, $\eta_p^2 = .176$, and the noncorresponding trials, $F(1,17) = 3.82$, $p = .067$, $\eta_p^2 = .183$. The main effect of Symmetry was significant for the noncorresponding trials, $F(1,17) = 12.47$, $p = .003$, $\eta_p^2 = .423$.

On the other hand, in the NPD group the Jointness × Symmetry interaction did not reach significance either for the corresponding, $F(1,17) = 0.73$, $p = .790$, $\eta_p^2 = .004$, or for the noncorresponding trials, $F(1,17) = 0.36$, $p = .555$, $\eta_p^2 = .021$, suggesting that the other's perspective did not interfere with decision making in this case. Again, the main effect of Symmetry was significant only for the noncorresponding trials, $F(1,17) = 22.37$, $p < .0001$, $\eta_p^2 = .568$. Overall, the data suggest that our findings were consistent over both corresponding and noncorresponding trials.

2.2.2. Hit rate

A 2 × 2 × 2 × 2 mixed ANOVA (Jointness [individual, joint] × Symmetry [symmetric, asymmetric] × Correspondence [corresponding, non-corresponding] × Task [PD, NPD]) was conducted on hit rates (for descriptive statistical data see Table 2).

Jointness had a significant main effect on performance, participants being more successful on individual trials, $F(1,34) = 4.831$, $p = .035$, $\eta_p^2 = .124$. Additionally, we found a significant Jointness × Correspondence interaction, $F(1,34) = 4.506$, $p = .041$, $\eta_p^2 = .117$. Corresponding decisions ("yes" answers) were more error prone in the joint compared to the individual trials than noncorresponding ("no") decisions. A tendency level interaction of Jointness and Task was also found that was caused by a decrease of performance from individual to joint trials in the PD group, but not in the NPD group, $F(1,34) = 3.483$, $p = .071$, $\eta_p^2 = .093$. The Jointness, Symmetry, Task three-way interaction was not significant, $F(1,34) = 0.028$, $p = .868$. These results converge with the RT findings in that Jointness tended to worsen performance only when the partner had to process the perspective dependent stimulus features. This however, did not differentially affect symmetric and asymmetric trials.

3. Study 2

As, to our knowledge, this is the first empirical support for spontaneous level-2 perspective taking in adults it was deemed important to test the robustness and replicability of this effect. Thus, the aim of Study 2 was to replicate the interference effect found in the PD group in Study 1 with a different group of participants and slightly modified visual stimuli.

As a consequence of the previously used digital font it might have been easier to make a decision concerning asymmetric numbers than symmetric numbers, as the former (6,9) differed in two line segments, while the latter (0,8) differed in one line segment (for an illustration see Table 1). To eliminate any possible effects of this difference, in Study 2 we displayed analog numbers where 0 and 8 are also easily distinguishable.

3.1. Methods

3.1.1. Design and participants

A total of 18 people participated in the study in pairs ($M_{age} = 22.78$, $SD = 3.67$, 11 women). All participants took part in the PD group where the within subject individual condition served as a control.

3.1.2. Materials

Only the visual stimuli were modified. Instead of the digital characters, we used analog numbers drawn in Matlab R2013a (for an illustration see Table 1). As in Study 1, the numbers were displayed in two different widths and two different colors, varied randomly. The numbers were 11.5 cm in height, and 5.2 or 6.2 cm in width. The visual images were blue and pink and

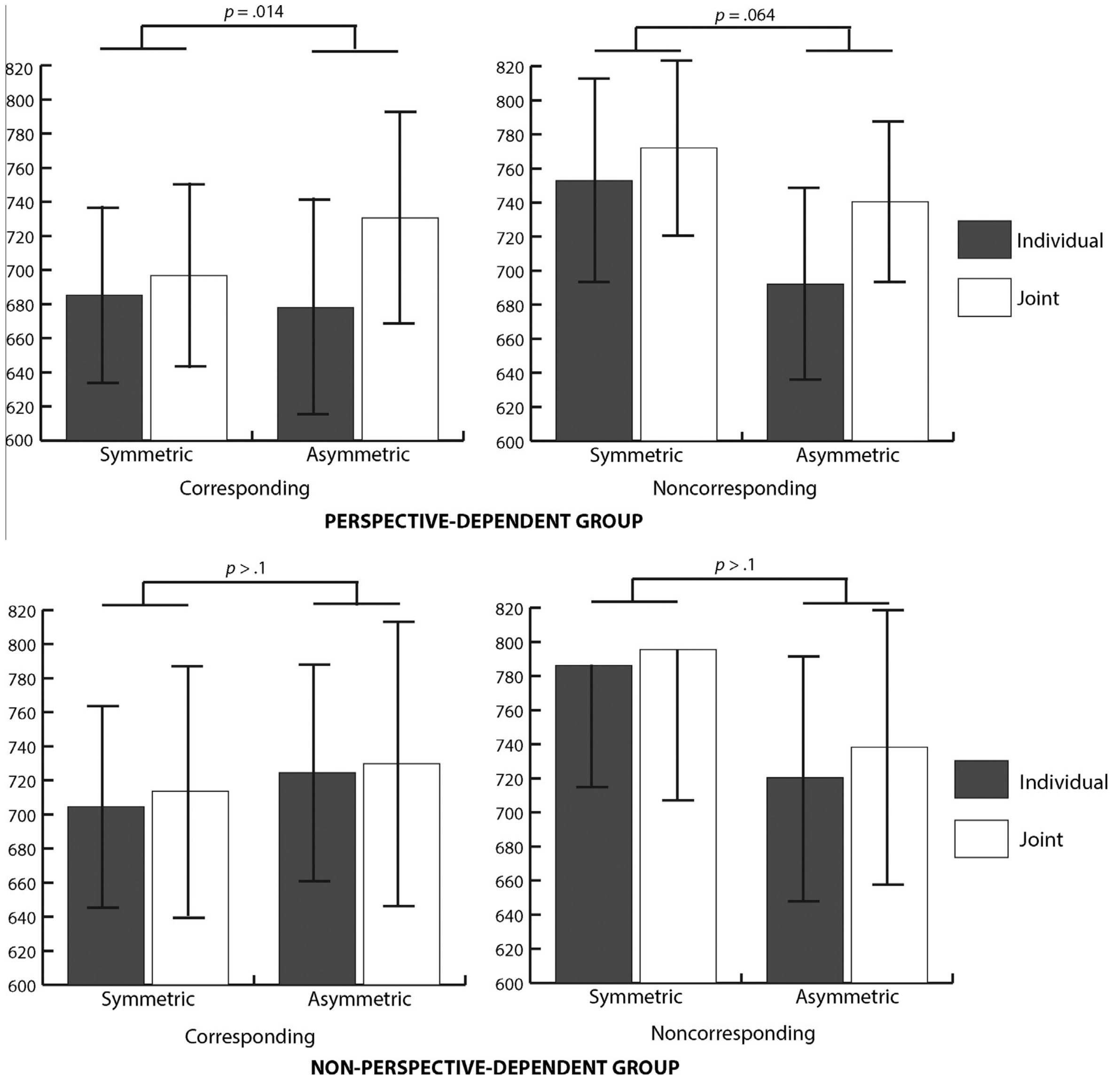


Fig. 1. Mean RT-s as a function of Jointness, Symmetry, Correspondence and Task. Error bars represent 95% Confidence Intervals.

Table 2
Mean percentage of correct responses (with Standard Deviation in parentheses) in Study 1 and 2 as a Function of Jointness, Symmetry, Correspondence and Task.

	Study 1							
	Individual				Joint			
	Symmetric		Asymmetric		Symmetric		Asymmetric	
	Corres	Noncorres	Corres	Noncorres	Corres	Noncorres	Corres	Noncorres
PD	95.23 (4.38)	92.4 (5.35)	93.73 (5.35)	94.35 (5.6)	89.55 (7.61)	91.91 (7.71)	89.84 (9.46)	88.94 (4.35)
NPD	92.95 (4.46)	90.36 (7.57)	94.78 (3.89)	92.5 (5.65)	92.87 (5.58)	91.74 (5.54)	92.01 (7.64)	92.71 (6.44)
	Study 2							
PD	94.49 (4.95)	88.54 (8.31)	93.98 (5.94)	93.57 (5.53)	93.83 (6.8)	94.06 (4.25)	91.47 (8.37)	92.29 (6.93)

participants had to press green and red buttons to indicate their answers (corresponding and noncorresponding respectively).

3.1.3. Procedure

The procedure was the same as in Study 1.

3.2. Results and discussion

3.2.1. Mean reaction times

A 2 × 2 × 2 ANOVA (Jointness [individual, joint] × Symmetry [symmetric, asymmetric] × Correspondence [corresponding, noncorresponding]) was run on the mean RT-s (for descriptive statistical data see Table 1). This revealed the main effect of Correspondense, corresponding decisions ("yes" answers) made more quickly, $F(1,17) = 14.513$, $p = .001$, $\eta_p{}^2 = .461$. The Symmetry × Correspondence interaction was found to be significant, $F(1,17) = 11.933$, $p = .003$, $\eta_p{}^2 = .412$. Just as in Study 1, this did not interact with Jointness. Crucially, we could replicate the Jointness × Symmetry interaction, implying that the other's perspective interfered with decision making more for the asymmetric than for the symmetric numbers, $F(1,17) = 9.922$, $p = .006$, $\eta_p{}^2 = .369$. Additionally, the main effect of Jointness approached significance, RT-s being marginally shorter in individual trials, $F(1,17) = 3.959$, $p = .063$, $\eta_p{}^2 = .189$.

Finally, we tested whether the same pattern emerged for both the corresponding and the noncorresponding trials, and found the expected interaction in both cases, corresponding: $F(1,17) = 4.417$, $p = .051$, $\eta_p{}^2 = .206$, noncorresponding: $F(1,17) = 5.345$, $p = .034$, $\eta_p{}^2 = .239$). Additionally, on the noncorresponding trials the main effect of both Jointness, $F(1,17) = 5.348$, $p = .034$, $\eta_p{}^2 = .239$, and Symmetry, $F(1,17) = 18.918$, $p < .0001$, $\eta_p{}^2 = .527$, was significant.

3.2.2. Hit rate

A Jointness [individual, joint] × Symmetry [symmetric, asymmetric] × Correspondence [corresponding, noncorresponding] ANOVA was performed on the hit rates (Table 2). In line with our expectations the analysis revealed a significant interaction between Jointness and Symmetry, $F(1,17) = 4.976$, $p = .039$, =.226. For asymmetric trials participants were more accurate in the individual than in the joint blocks, whereas the reverse was true for symmetric trials. Hence, hit rate provides converging evidence for the spontaneous computation of the partner's level-2 perspective. Additionally, the Jointness × Correspondense interaction was also found to be significant, $F(1,17) = 4.558$, $p = .048$, $\eta_p{}^2 = .211$, participants were better at making noncorresponding ("no") responses in the joint blocks, and corresponding ("yes") responses in the individual blocks. As the Jointness × Correspondence × Symmetry three-way interaction was significant on a tendency level, $F(1,17) = 3.809$, $p = .068$, $\eta_p{}^2 = .183$, we performed separate 2 × 2 Jointness [individual, joint] × Symmetry [symmetric, asymmetric] ANOVA-s for corresponding and noncorresponding trials. The Jointness × Symmetry interaction, signaling PT, was found to be specific to the noncorresponding trials, $F(1,17) = 11.432$, $p = .004$, $\eta_p{}^2 = .402$. In sum, indication of spontaneous level-2 PT was found in participants' accuracy as well, however, this was found to be more limited than the effect on RT-s.

4. General discussion

It has been consensual both in the empirical research and in various theoretical approaches that level-2 PT is a cognitively demanding, thus deliberate process. The findings presented here draw a more nuanced picture, suggesting that in certain contexts level-2 PT does emerge spontaneously. The interference effect we found in reaction times indicated that whenever participants took part in a task that required processing the perspective dependent object features a representation was formed about how the object looked from the partner's perspective. The representation emerged independently of participants' overt goal to make self perspective based decisions, and formed quickly enough to interfere with those decisions. This happened despite the fact that throughout the experiment participants only had to make judgments based on their own perspective. That is, they did not have to switch back and forth between tasks or perspectives, and thus, were not trained by the experimental situation to take the other's perspective as quickly as they could. The interference effect found in RT-s gained partial support from the hit rate measure as well.

Importantly, our findings showed that the fact that a fellow human being was actively making decisions regarding the same object from a different perspective was not sufficient for the interference effect to occur. Perspective interference only emerged when participants had evidence that their partner was processing the perspective dependent object feature as well. Recently, Carruthers (2015a) argued that, due to its limited capacity, the mindreading system cannot possibly give a complete model of the whole environment. Rather, it limits its operation to objects that are looked at by the confederate, that is, to objects that the agent is *aware of*. Our results indicate that cues more subtle than direction of gaze are used as a selection criterion. By varying the task of the partner, we manipulated which *stimulus feature* the other was aware of. Our findings suggest that spontaneous level-2 PT is limited to situations when someone from the other perspective is actually attending the perspective dependent features.

There is a related possibility that could have contributed to the effect and which can also be derived from the one system model of Carruthers (2015a). Namely, for spontaneous processes to be launched no external prompting or instruction is needed, however, the agent has to have the covert goal to perform that computation. Varying the partner's task also meant varying whether it was the same as one's own. Pursuing the same task as the other could have induced the implicit goal to understand and predict the other's actions to a greater extent than having different tasks. The above two explanations can both be accommodated with the one system model of ToM and can work in an additive way. Future studies will have to disentangle the role that these two factors played.

Evidence for spontaneous level-2 perspective taking is unique in the literature, and thus, the current studies have important implications for the existing models on Theory of Mind. Namely, the representation that caused interference necessarily contained more detailed information about the partner's perspective than merely tracking whether he had perceptual access to the object. Perspective interference could have resulted from representing either that whatever the partner saw looked *different* from what the participant himself could see, or from representing *how exactly* the object appeared to the partner.

Note, however, that according to the two systems account, the minimal ToM, that is responsible for online computation, cannot handle appearance related information (Apperly & Butterfill, 2009; Butterfill & Apperly, 2013). In its original form, the one system account of Carruthers (2015a, 2015b) predicts no spontaneous level-2 perspective taking either, albeit for a different reason. The one system model claims that any perspective content can be formed spontaneously given that its' computation does not require cognitively effortful processes. As explicit level-2 perspective judgments had been found to rely on a demanding, mental rotation like mechanism (Surtees et al., 2013a; Surtees et al., 2013b), level-2 PT was thought to be incompatible with spontaneity. Importantly, this model does not exclude the possibility of spontaneous level-2 PT *per se*, it excludes the possibility of a cognitively demanding computation performed spontaneously.

Two questions arise. First, do the results of the current experiments indicate that information that necessitates working memory (in the form of mental rotation) can be computed by the mindreading system spontaneously? Second, was the interfering representation actually *attributed* to the partner? Our answer to both questions is no, not necessarily.

Regarding the first question, we argue that it is possible to reach perspective information in the given paradigm without the cognitively demanding process of mental rotation. The number verification task builds on adults' relatively automatized capacity to recognize numerals. Participants most likely had associations between the concepts "six" and "nine" as they had had prior experience through education that these characters could be rotated into each other. Hence, the information, that the character appeared to depict a different number from the other's point of view, could have been acquired through accessing pre-existing conceptual knowledge, instead of using demanding mental rotation on each trial. There can, of course, be circumstances under which participants rely on mental rotation in this task (Surtees et al., 2013a). However, there is also a possibility to bypass working memory, which shortcut may have allowed participants to quickly map the scene from the other's perspective. The results presented here provide evidence that information on object appearance from a different perspective can be reached spontaneously and online. This finding can be accommodated with the one system account (Carruthers, 2015a), but contradicts the prediction of the two systems account (Apperly & Butterfill, 2009; Butterfill & Apperly, 2013).

Now, turning to the issue of attribution, while it is possible, the interference effect we found does not prove that the representation that interfered was necessarily attributed to the confederate as a mental state in the form of a proposition or otherwise. Merely reaching the conflicting information (the *content* of the partner's perspective) through the conceptual shortcut would have been sufficient to encumber decision-making. There was of course a need to identify which information (e.g. six/nine) belonged to the egocentric perspective in order to make a correct decision. However, it is possible that the content of the partner's perspective was accessed without acknowledging it to be a perspective that belonged to a *specific person*. The phenomenon that a mental state is computed without being bound to a certain agent has already been documented in ToM research (Kampis, Somogyi, Itakura, & Király, 2013). What is more, for the conflicting content to interfere, the information did not even have to be acknowledged as constituting *a perspective at all*. Accessing the object's alternative identity without recognizing that it referred to the mutually viewed object under another aspect could just as well hinder decision-making. According to our findings, what did depend on prior information about the social context was whether the content of the partner's perspective was activated in the first place: it was only accessed if the partner was assumed to be *aware* of the perspective dependent stimulus feature.

It is widely assumed that the primary function of Theory of Mind is to explain and predict behavior in order to adjust one's own behavior (either in a cooperative or a competitive way) to others'. The phenomenon we tapped into clearly improves one's chances to coordinate behavior with others. The findings indicate that given certain contextual cues (e.g. prior information that the other is also attending the perspective dependent features), adults are predisposed to activate all information they have that enables them to map what others might see about the surrounding environment. In a live social interaction the mere activation of a pre-existing piece of knowledge on perspectives that are "occupied" by someone might create a readiness for actions guided by that information. This, in turn, provides the opportunity to respond quickly and adequately to those actions. However, the findings are neutral regarding whether the interference emerged as a consequence of a meta-representation formed on the partner's perceptual experience.

By providing the first evidence for spontaneous level-2 perspective interference the results of the current experiments support those claiming that adult mindreading has to be viewed in a more dynamic way than the two systems account does, devoting attention to the role of contextual factors or cues in launching Theory of Mind processes (German & Cohen, 2012). However, the results also raise numerous questions to follow upon. For instance, instead of using avatars as social partners, we employed a live perspective-taking situation, as it is commonly implemented in studies on joint action (Sebanz, Bekkering, & Knoblich, 2006) or task sharing (Sebanz, Knoblich, & Prinz, 2003). Although, positive findings with avatars certainly show the robustness of level-1 perspective taking, some phenomena might call for more than a schematic sign of a social partner to occur. Future studies will have to clarify whether the presence of a human agent was indeed necessary for level-2 PT to happen.

Finally, the results presented here provide evidence for the spontaneity of computation, as interference emerged independently of the participants' overt goal. The interfering content was reached online, quickly enough to hinder self-perspective based decisions. The effect did, however, depend on what the participant knew about the partner's task. This is in line with German and Cohen's (2012) argumentation that the sub-processes of ToM might only be engaged when specific combinations of stimuli (involving contextual cues as well) apply. According to some, any effect that relies on such preconditions fails to qualify as automatic (Back & Apperly, 2010). Additional research is needed to determine whether the effect resists external incentives to inhibit the distracting non-relevant perspective, providing evidence for its involuntariness, and whether interference also occurs under cognitive load implying the efficiency of the process.

In summary, the results suggest that the mindreading system indeed works in a dynamic way, making use of off-line cues about the confederate's task, and prior knowledge regarding the aspectual nature of objects. Our findings indicate that the dichotomy of automatic ToM processes, tracking only relational information, and flexible mindreading processes, computing mental states, might be less clear-cut than previously thought.

Acknowledgments

We thank Ágnes Kovács for helpful comments on the manuscript and Caoimhe O'Reilly for providing language check. This work was supported by the OTKA under Grant number K-109352 and K-116779.

Appendix A. Appendix

A.1. Instructions

A.1.1. First part of instructions (both groups)

"You are going to participate in this study together, both of you will have to make decisions on simple audio and visual stimuli. There will be parts when you work alone, during this time the other person will sit with his/her back to the participant who is working, and there will be a part when you work in parallel, during this time you will sit facing each other. Both of you will have a response box to indicate your answers. One after the other you will hear numbers from a loudspeaker. During this, you will see differently colored number characters on the screen, one for each number heard from the loudspeaker."

A.1.2. Second part of instructions (perspective-dependent group)

"You will not have to deal with the color of the numbers. Your task is to decide whether the number you hear was the same as what you see. If the two numbers are the same, press the yellow button on the box, if they are different, press the red button. It is important that you answered correctly, but quickly."

A.1.3. Second part of instructions (non-perspective-dependent group)

"You (name of participant A) will not have to deal with the color of the numbers. Your task is to decide whether the number you hear was the same as what you see. If the two numbers are the same, press the yellow button on the box, if they are different, press the red button. You (name of participant B) will not have to deal with the numbers you hear, or the meaning of the numbers you see. Your task is to decide, whether the color of the character was the same as the color of the previous character. If the two colors are the same, press the yellow button on the box, if they are different, press the red button. It is important that you answered correctly, but quickly."

References

Aichhorn, M., Perner, J., Kronbichler, M., Staffen, W., & Ladurner, G. (2006). Do visual perspective tasks need theory of mind? *NeuroImage, 30*, 1059–1068. http://dx.doi.org/10.1016/j.neuroimage.2005.10.026.

Apperly, I. A. (2008). Beyond simulation-theory and theory–theory: Why social cognitive neuroscience should use its own concepts to study "theory of mind". *Cognition, 107*, 266–283. http://dx.doi.org/10.1016/j.cognition.2007.07.019.

Apperly, I. A., & Butterfill, S. A. (2009). Do humans have two systems to track beliefs and belief-like states? *Psychological Review, 116*, 953–970. http://dx.doi.org/10.1037/a0016923.

Back, E., & Apperly, I. A. (2010). Two sources of evidence on the non-automaticity of true and false belief ascription. *Cognition, 115*(2010), 54–70. http://dx.doi.org/10.1016/j.cognition.2009.11.008.

Bargh, J. A. (1989). Conditional automaticity: Varieties of automatic influence in social perception and cognition. In J. S. Uleman & J. A. Bargh (Eds.), *Unintended thought* (pp. 3–51). New York: Guilford Press.

Broadbent, D. E. (1958). *Perception and communication*. London: Pergamon.

Butterfill, S. A., & Apperly, I. A. (2013). How to construct a minimal theory of mind. *Mind & Language, 28*, 606–637. http://dx.doi.org/10.1111/mila.12036.

dc_1353_16

Carruthers, P. (2015a). Mindreading in adults: Evaluating two-systems views. *Synthese*. http://dx.doi.org/10.1007/s11229-015-0792-3.

Carruthers, P. (2015b). Two systems for mindreading? *Review of Philosophy and Psychology*. http://dx.doi.org/10.1007/s13164-015-0259-y.

Christensen, W., & Michael, J. (2015). From two systems to a multi-systems architecture for mindreading. *New Ideas in Psychology*, *40*(PartA), 48–64. http://dx.doi.org/10.1016/j.newideapsych.2015.01.003.

Cohen, M. A., Cavanagh, P., Chun, M. M., & Nakayama, K. (2012). The attentional requirements of consciousness. *Trends in Cognitive Sciences, 16*, 411–417. http://dx.doi.org/10.1016/j.tics.2012.06.013.

Flavell, J. H., Everett, B. A., Croft, K., & Flavell, E. R. (1981). Young children's knowledge about visual perception: Further evidence for the level 1–level 2 distinction. *Developmental Psychology, 17*, 99–103. http://dx.doi.org/10.1037//0012-1649.17.1.99.

German, T. C., & Cohen, A. S. (2012). A cue-based approach to 'theory of mind': Re-examining the notion of automaticity. *British Journal of Developmental Psychology, 30*, 45–58. http://dx.doi.org/10.1111/j.2044-835X.2011.02055.x.

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

Levin, D. T., & Simons, D. J. (1997). Failure to detect changes to attended objects in motion pictures. *Psychonomic Bulletin & Review, 4*, 501–506. http://dx.doi.org/10.3758/BF03214339.

Mack, A., & Rock, I. (1998). *Inattentional blindness*. Cambridge, MA: MIT Press.

Michelon, P., & Zacks, J. M. (2006). Two kinds of visual perspective taking. *Perception and Psychophysics, 68*, 327–337. http://dx.doi.org/10.3758/BF03193680.

Moll, H., & Meltzoff, A. N. (2011). How does it look? Level 2 perspective-taking at 36 months of age. *Child Development, 82*, 661–673. http://dx.doi.org/10.1111/j.1467-8624.2010.01571.x.

Moll, H., & Tomasello, M. (2006). Level 1 perspective-taking at 24 months of age. *British Journal of Developmental Psychology, 24*, 603–613. http://dx.doi.org/10.1348/026151005X55370.

Nielsen, M. K., Slade, L., Levy, J. P., & Holmes, A. (2015). Inclined to see it your way: Do altercentric intrusion effects in visual perspective taking reflect an intrinsically social process? *The Quarterly Journal of Experimental Psychology, 68*, 1931–1951. http://dx.doi.org/10.1080/17470218.2015.1023206.

Perner, J., Stummer, S., Sprung, M., & Doherty, M. (2002). Theory of mind finds its Piagetian *perspective*: Why alternative naming comes with understanding belief. *Cognitive Development, 17*, 1451–1472. http://dx.doi.org/10.1016/S0885-2014(02)00127-2.

Qureshi, A. W., Apperly, I. A., & Samson, D. (2010). Executive function is necessary for perspective selection, not Level-1 visual perspective calculation: Evidence from a dual-task study of adults. *Cognition, 117*, 230–236. http://dx.doi.org/10.1016/j.cognition.2010.08.003.

Rakoczy, H. (2012). Do infants have a theory of mind? *British Journal of Developmental Psychology, 30*, 59–74. http://dx.doi.org/10.1111/j.2044-835X.2011.02061.x.

Samson, D., & Apperly, I. A. (2010). There is more to mind reading than having theory of mind concepts: New directions in theory of mind research. *Infant and Child Development, 19*, 443–454. http://dx.doi.org/10.1002/icd.678.

Samson, D., Apperly, I. A., Braithwaite, J. J., Andrews, B. J., & Bodley Scott, S. E. (2010). Seeing it their way: Evidence for rapid and involuntary computation of what other people see. *Journal of Experimental Psychology: Human Perception and Performance, 36*, 1255–1266. http://dx.doi.org/10.1037/a0018729.

Santiesteban, I., Catmur, C., Hopkins, S. C., Bird, G., & Heyes, C. (2014). Avatars and arrows: Implicit mentalizing or domain-general processing? *Journal of Experimental Psychology: Human Perception and Performance, 40*, 929–937. http://dx.doi.org/10.1037/a0035175.

Sebanz, N., Bekkering, H., & Knoblich, G. (2006). Joint action: Bodies and minds moving together. *Trends in Cognitive Sciences, 10*, 70–76. http://dx.doi.org/10.1177/0956797612439070.

Sebanz, N., Knoblich, G., & Prinz, W. (2003). Representing others' actions: Just like one's own? *Cognition, 88*, B11–B21. http://dx.doi.org/10.1016/S0010-0277(03)00043-X.

Sodian, B., Thoermer, C., & Metz, U. (2007). Now I see it but you don't: 14-month-olds can represent another person's visual perspective. *Developmental Science, 10*, 199–204. http://dx.doi.org/10.1111/j.1467-7687.2007.00580.x.

Surtees, A. D. R., & Apperly, I. A. (2012). Egocentrism and automatic perspective taking in children and adults. *Child Development, 83*, 452–460. http://dx.doi.org/10.1111/j.1467-8624.2011.01730.x.

Surtees, A. D. R., Apperly, I. A., & Samson, D. (2013a). Similarities and differences in visual and spatial perspective-taking processes. *Cognition, 129*, 426–438. http://dx.doi.org/10.1016/j.cognition.2013.06.008.

Surtees, A. D. R., Apperly, I. A., & Samson, D. (2013b). The use of embodied self-rotation for visual and spatial perspective-taking. *Frontiers in Human Neuroscience, 7*, 1–12. http://dx.doi.org/10.3389/fnhum.2013.00698.

Surtees, A. D. R., Butterfill, S. A., & Apperly, I. A. (2012). Direct and indirect measures of Level-2 perspective-taking in children and adults. *British Journal of Developmental Psychology, 30*, 75–86. http://dx.doi.org/10.1111/j.2044-835X.2011.02063.x.

Teufel, C. et al (2009). Social cognition modulates the sensory coding of observed gaze direction. *Current Biology, 19*, 1274–1277. http://dx.doi.org/10.1016/j.cub.2009.05.069.

Wimmer, G., Hogrefe, J., & Perner, J. (1988). Children's understanding of informational access as source of knowledge. *Child Development, 59*, 386–396.

British Journal of Developmental Psychology (2017)

The British Psychological Society

www.wileyonlinelibrary.com

Level-2 perspectives computed quickly and spontaneously: Evidence from eight- to 9.5-year-old children

Fruzsina Elekes[1,2]*, Máté Varga[3] and Ildikó Király[2,4]

[1]Doctoral School of Psychology, Eötvös Loránd University, Budapest, Hungary
[2]Institute of Psychology, Eötvös Loránd University, Budapest, Hungary
[3]Department of Cognitive Science, Budapest University of Technology and Economics, Budapest, Hungary
[4]Cognitive Development Center, Central European University, Budapest, Hungary

It has been widely assumed that computing how a scene looks from another perspective (level-2 perspective taking, PT) is an effortful process, as opposed to the automatic capacity of tracking visual access to objects (level-1 PT). Recently, adults have been found to compute both forms of visual perspectives in a quick but context-sensitive way, indicating that the two functions share more features than previously assumed. However, the developmental literature still shows the dissociation between automatic level-1 and effortful level-2 PT. In the current paper, we report an experiment showing that in a minimally social situation, participating in a number verification task with an adult confederate, eight- to 9.5-year-old children demonstrate similar online level-2 PT capacities as adults. Future studies need to address whether online PT shows selectivity in children as well and develop paradigms that are adequate to test preschoolers' online level-2 PT abilities.

Statement of Contribution

What is already known on this subject?

- Adults can access how objects appear to others (level-2 perspective) spontaneously and online
- Online level-1, but not level-2 perspective taking (PT) has been documented in school-aged children

What the present study adds?

- Eight- to 9.5-year-olds performed a number verification task with a confederate who had the same task
- Children showed similar perspective interference as adults, indicating spontaneous level-2 PT
- Not only agent-object relations but also object appearances are computed online by eight- to 9.5-year-olds

The success of our everyday interactions with fellow humans calls for the ability to determine how those individuals represent the surrounding social and physical world. The capacity to attribute beliefs, intentions or percepts to others; that is Theory of Mind (ToM, Wimmer & Perner, 1983) allows humans to interpret and predict their interactional

*Correspondence should be addressed to Fruzsina Elekes, Doctoral School of Psychology, Eötvös Loránd University, Izabella u. 46. Budapest 1064, Hungary (email: elekes.fr@gmail.com).

DOI:10.1111/bjdp.12201

partners' actions. The speed with which that information can be accessed is crucial for real-time coordination, although it is widely assumed that automatically computed information lacks the conceptual richness of genuine mental states (Apperly & Butterfill, 2009). However, recent findings with adults show more complex online mindreading skills than previously reported. Specifically, recent studies show that adults can compute how objects look from another perspective in an online, uninstructed, and flexible way (Elekes, Varga, & Király, 2016; Surtees, Apperly, & Samson, 2016; Surtees, Samson, & Apperly, 2016). The current study takes the first step to assess the same skill in school-aged children.

According to Perner, Brandl, and Garnham (2003), the so-called perspective problems arise when multiple representations with different contents are formed about the same referent. That is, the representations' aspectual shape or mode of presentation differs. Perspective in this theoretical framework means that something is represented *in a certain way*, which is not limited to visual perspectives. To be able to understand such problems in their full depth, for example false beliefs or complex forms of visual perspective taking, a meta-representational integration has to be obtained that allows identifying which pieces of information belong to which specific perspective, that, on the other hand, requires the understanding of the representational nature of mental states (Perner *et al.*, 2003). The so-called two-system models claim that understanding the representational nature of mental states calls for a functionally distinct, more sophisticated ToM capacity (e.g., Apperly & Butterfill, 2009).

The two-system approach of Theory of Mind assumes (Apperly & Butterfill, 2009; Butterfill & Apperly, 2013; Rakoczy, 2012) that there are two distinct systems that are accountable for different forms of mentalizing. System 1 operates automatically, in an efficient, mandatory way and so enables online behaviour prediction. However, it can only compute *relations* between objects and agents, not genuine mental states. Representing mental states entails representing representations *as such* (i.e., understanding perspective problems). These instances of mentalizing are handled by the ontogenetically later developing system 2, which operates in a cognitively demanding consequently slow and offline but flexible manner. Accordingly, the capacity of reasoning about rich concepts of perspective has been found to develop well into the preschool years (Rakoczy, Bergfeld, Schwarz, & Fizke, 2015). Situations in which differing perspectives result in conflicting perceived *identities* of the same objects (e.g., a rubber that appears to be a toy dog), rather than different properties of the object (e.g., length of a partially occluded stick) have proven to be especially hard to handle (Sprung, Perner, & Mitchell, 2007). The difficulty regarding multiple identities is reflected in the initial behavioural responses of even those children and adults, who do have an explicit understanding of perspective problems, supporting the notion that perspective problems cannot be solved online (Low & Watts, 2013).

Of the different forms of perspective problems, in our focus, stands level-2 visual perspective taking (PT). The term level-2 perspective taking refers to the ability to map how certain objects look from different viewing angles, that is, representing not only whether another agent has visual access to certain objects (level-1 PT), but also how those objects are perceived (Flavell, Everett, Croft, & Flavell, 1981). To reason about someone's level-2 perspective, the encoder needs to acknowledge that different viewing angles lead to different percepts of the same object. For instance, the drawing of a turtle that, to the encoder, seems to stand on its feet appears to be lying on its back from the opposing perspective. A difference in visual perspective can also lead to perceived differences in object identity. For instance, the meaning of certain symbols in the Arabic notation (e.g., 6/9) depends on perspective: The same object may mean different things to perceivers who maintain different spatial positions. Similar to other types of perspective problems, the ability to reason about level-2 perspectives develops between 4 to 5 years of age (Moll, Meltzoff, Merzsch, & Tomasello, 2013) and is thought to depend on mental rotation as a mechanism (Surtees, Apperly, & Samson, 2013).

In accordance with the theoretical argument that level-2 perspectives are computed by system 2, empirical evidence also demonstrated that it remains to be an effortful and deliberate process in school-aged children (6- to 11-year-olds) and adults as well (Surtees, Butterfill, & Apperly, 2012). Level-2 PT was reported not to emerge in implicit tasks, where perspective computation is task irrelevant. The general structure of such implicit PT studies involves presenting visual scenes to participants as well as a figurative indication of a fellow social being (an avatar) whose perspective of the scene either leads to the same or conflicting perceptions. Participants perform their task regarding what they themselves see (self trials), and regarding what the avatar is assumed to see (other trials). Participants' performance reflects spontaneous PT if their self-perspective decisions are made more slowly and/or less accurately when the avatar's irrelevant perspective is in conflict with their own (altercentric intrusion).

However, evidence has started to accumulate that, in adults, level-2 PT functions in a way that is quick and independent of explicit task instructions (Elekes *et al.*, 2016; Surtees, Apperly, *et al.*, 2016; Surtees, Samson, *et al.*, 2016). These findings indicate that humans may at least be equipped with partial solutions to map how a scene is visible to others in order to guide their behaviours online. Following Carruthers' (2017) definition, we call these instances spontaneous perspective taking, as they do not play a role in participants' overt goals, but are probably still driven by covert social motivations and thus cannot be considered automatic. Genuinely automatic processes are launched in a stimulus-driven way, their execution cannot be down regulated, and are not affected by parallel cognitive load (Bargh, 1989; Carruthers, 2017; Moors & De Houwer, 2006). In the following, we give a critical analysis of recent results on spontaneous level-2 PT pointing out yet unanswered questions regarding their theoretical implications. Then, we highlight how examining the same capacity in children could contribute to our understanding of the nature of human Theory of Mind.

Investigating adults, Surtees, Samson, *et al.* (2016) found intrusions from an avatar's inconsistent perspective both for level-1 and level-2 type decisions, emerging online and without external prompting. However, level-2 PT was limited to the condition where self- and other perspective decisions were alternated randomly within a block. This, according to the authors, prevented participants to strategically disregard the partner's viewpoint as the relevance of the partner's perspective was varied from trial to trial. They take this as evidence that 'unintentional' PT is limited to level-1 contents and argue that 'only level-1 perspective taking was triggered outside of cognitive control' (2016, p 103). In our opinion, the lack of efficient strategy to disregard certain bits of information when self and other trials were alternated does not necessarily lead to an intention to process that information. Furthermore, the strategy that could be applied when trials were presented in blocks may have provided participants with a way to inhibit, or *downregulate* the ongoing process of level-2 PT *after* being triggered outside of cognitive control. Nevertheless, Surtees, Samson, *et al.* (2016) data provide evidence for spontaneous level-2 PT, and that it is affected by task features.

Elekes *et al.* (2016) and Surtees, Apperly, *et al.* (2016) reported spontaneous level-2 PT in paradigms (number verification and magnitude judgement, respectively) where, instead of using avatars as social partners, two adults took part in the task together and participants only had to make self-perspective decisions throughout the experiment. Participants in Surtees, Apperly, and colleagues were prompted to make eye contact before each trial, whereas no such instruction was applied in the other study. Importantly, in Elekes *et al.* (2016) the interference effect was found to be selective along the features of the partner's task. Level-2 PT only occurred spontaneously if participants knew that their partner was also performing a number verification task; that is, he or she had to process the perspective-dependent feature of the stimuli as well. In stark contrast, in Surtees, Apperly, *et al.* (2016) study, level-2 altercentric intrusions emerged also in the condition where the partner was known to perform a task regarding the surface features (texture) of the stimuli. While Elekes *et al.* (2016) argue that selection takes place based on participants' knowledge about their *partner's* attentional focus, Surtees, Apperly, *et al.* (2016) claim that humans represent those aspects of the partner's perspective, which are relevant for *themselves*, regardless of its relevance for their partner.

The difference between these two sets of findings suggests that PT is sensitive to yet unexplored contextual factors. For instance, prompted eye contact in Surtees, Apperly, *et al.* (2016) may have had a facilitative effect on perspective taking (eliminating its selectivity) by either highlighting the joint nature of the task, or by inducing the expectation that the referent of the partner's gaze was relevant to the encoder (Csibra & Gergely, 2009).

To summarize, the findings of three different studies converge on the fact that adults can access information on object appearances from a co-actor's perspective quick enough to interfere with their own perspective-based decisions and that this is sensitive to contextual features. Thus, the computation of level-2 perspectives bears features of both system 1 (online, uninstructed) and system 2 (flexible, represents appearance instead of agent-object relation). It seems reasonable to assume that the representation formed on others' perspectives in this case is less complex than the full-blown understanding of perspective problems, accessing merely the informational *content* of the partner's perspective, not necessarily acknowledging the perspectival nature of that information (Elekes *et al.*, 2016). However, even if interpreted in this representationally simple way, these findings pose a considerable challenge for the classical two-system view that posits a strict dichotomy between automatically computed level-1 perspectives and offline, effortfully computed level-2 perspectives.

On the other hand, the developmental literature still reflects this rigid dichotomy. Previous tests of school-aged (6- to 10-year-old) children suggested that they can automatically represent others' perceptual access to objects, that is level-1 PT (Surtees & Apperly, 2012), but fail to represent how an object appears from someone else's perspective, that is level-2 PT, in an implicit paradigm (Surtees *et al.*, 2012). The objective of our study is to take the first steps in providing converging developmental evidence to the more recent adult findings, which challenge the two-system view and potentially point to the one-system view of ToM.

Recent one-system models of mindreading (Carruthers, 2017; Westra, 2016) posit that a single ToM capacity can explain the diverse functioning of mindreading. Carruthers' (2017) one-system account claims that spontaneous PT inevitably arises for any perspective content (i.e., both level-1 and level-2) if covert motivation is present to perform the computation. However, mindreading can only be achieved spontaneously if the situation (involving one's own task and the computation to be performed spontaneously) does not evoke high cognitive load (Carruthers, 2017). When cognitive demand is high, for instance in terms of executive functions or attention, participants' overt goals determine whether resources would be allocated to perspective computation.

Thus, high cognitive load can mask people's ability to access perspective information spontaneously. In Westra's (2016) view, mindreading reaches its quick but flexible mode of operation through recruiting, amongst others, prior knowledge stored in long-term memory and integrating that with specific mindreading processes. For instance, participants may circumvent demanding mental rotation by retrieving the once (effortfully) computed percept from memory.

Motivated by recent advances on spontaneous level-2 PT in adults, the aim of our study was to test school-aged children for spontaneous level-2 PT in the number verification task in addition to a control group of adults, closely following the methodology of Elekes *et al.* (2016). Eight-year-olds were chosen to be the youngest tested age as, having spent 1–2 years in formal education in primary school, they can already read numbers. However, as spontaneous mindreading is limited to cognitively undemanding contexts (Carruthers, 2017), the mere ability to identify numbers may not be sufficient to enable spontaneity in this paradigm.

Number recognition is gradually automatized during the first years of primary school (van Galen & Reitsma, 2008; Girelli, Lucangeli, & Butterworth, 2000). Children's lack of proficiency with number recognition makes the overt (number verification) task more demanding for them compared to adults in terms of cognitive resources. The lack of expertise also rules out or weakens the possibility to use prior knowledge about the perspective-dependent nature of the stimuli, which could be used to access object appearances from the other's perspective while avoiding demanding mental rotation (Elekes *et al.*, 2016). These effects, combined with eight-year-olds' necessarily more restrained level of general cognitive resources (Brocki & Bohlin, 2004; Rueda *et al.*, 2004), motivated us to test a group of 9.5-year-olds as well for whom the task would be less challenging. Although we predicted similar level-2 perspective interference in children as in adults, due to the cognitive demand constraint, the effect may grow stronger over age.

Methods

Participants

Thirty-two Caucasian middle class children (16 eight-year-old children, M_{age} = 97.1 months, SD = 5.36 months, age range: 86.1–104.2 months, seven males; 16 9.5-year-old children, M_{age} = 113.56 months, SD = 3.56 months, age range: 109.2–119.4 months, 11 males) and 16 adults (M_{age} = 22.4 years, SD = 2.1 years, eight males) took part in our study. An additional five children were tested but excluded due to technical error (two children – one in each age group), the child's failure to cooperate (two children – one in each age group), and chance-level performance (one child in the eight-year-old group). Participants included in the analysis finished the whole procedure. Participants or their caregiver gave informed consent before the experiment. The study was approved by ethical committee of the university where the experiment was performed.

Materials

The materials of Experiment 2 in Elekes *et al.* (2016) were used. These included the visual image of two symmetric (0, 8) and two asymmetric (6, 9) digits. The visual stimuli were drawn in Matlab R2013a. Each digit was generated in two colours (blue, pink) and two widths (5.2 cm, 6.2 cm), leading to a total of 16 images. The height of all digits was 11.5 cm. The audio stimuli, comprised of the four numbers uttered in Hungarian, were recorded in a female voice in neutral, descending intonation.

Procedure

The procedure was identical to Elekes *et al.* (2016)'s Experiment 2, except that all participants' partner was one of six female confederate research assistants. In case of the adults, participants were led to believe that the confederate was a naïve participant as well. None of the participants reported doubts about their partner. For the child participants, the confederate and the experimenter greeted the child upon arriving to the laboratory. No information was provided regarding the status of the confederate.

In all age groups, the participant and the confederate were introduced to each other and were seated next to each other to listen to the instructions (see Appendix). Hence, they were informed not only about their own but also their partner's task. Both members of the pair were instructed to perform the number verification task; that is, they had to decide whether the number character depicted on the screen was the same as the number presented in the auditory modality. After the instructions were provided, the experimenter led the participant and the confederate to the test room.

During the experiment, participants sat at the opposite sides of the table, maintaining opposing perspectives on the same stimuli. Stimuli were presented on a flat screen lying on a short table. Each participant took part in both joint and individual blocks of trials. In the joint condition, participants sat facing the table and made decisions in parallel; that is, all items were responded to by both participants. In the individual condition, the passive participant turned 180 degrees and was facing away from the stimuli. Thus, the partner was present throughout the experiment, while he or she either did or did not have visual access to the stimuli depending on the condition.

The experiment started by a practice phase for participant A, which was followed by the individual test phase for the same participant. After this, participant B practised the task and then the two of them took part in the joint test phase. Finally, participant B completed the individual phase. The confederate played either the role of participant A or participant B counterbalanced across participants. Practice involved 16 trials for both participants, while the three test phases consisted of 112 trials each and were divided into two blocks with a 30-s long break between the blocks. After each test phase, the experimenter entered the room and instructed participants about the next phase.

In all test phases, half of all trials depicted symmetric visual stimuli, and half asymmetric. Visual stimuli were either paired with corresponding (0 – zero, 8 – eight, 9 – nine, 6 – six) or non-corresponding audio (0 – eight, 8 – zero, 9 – six, 6 – nine). The timing of each trial was as follows: A fixation cross appeared in the middle of the screen before each trial. The audio stimulus was presented 500 ms-s after the fixation cross, which then was followed by the visual stimulus with a fixed 300 ms interstimulus interval. The picture remained on screen until one or both participants provided an answer depending on the test phase. Participants' hands were occluded from their partner's view. Half of the participants responded to corresponding trials with their right and half with their left hand. Reaction times were measured with a Cedrus-type response box in case of the naïve participants and with a keyboard in case of the confederate.

Data analyses

Average RTs were computed, after removing incorrectly answered trials and those trials where RT differed with more than 2 standard deviations from the mean of the given participant. In addition, we calculated a psychological efficiency score (ES) by dividing average RTs by the percentage of all accurately answered trials (Townsend & Ashby, 1978) – higher values indicating more demanding decisions. The ES enables controlling for different patterns of speed-accuracy trade-offs, which may be advantageous as child and adult participants may differ in their strategies. ES is used in studies of implicit perspective taking (Bukowski & Samson, 2016; Mattan, Quinn, Apperly, Sui, & Rotshtein, 2015).

Child and adult data were analysed separately. For children, 2 × 2 × 2 × 2 mixed ANOVAs were conducted on both RTs and the ES with Jointness (individual, joint), Symmetry (symmetric, asymmetric), and Correspondence (corresponding, non-corresponding) as within-subject factors, and Age (eight-year-old, 9.5-year-old) as a between-subjects factor. Adult RT and efficiency data were entered into repeated-measures ANOVAs with Jointness, Symmetry, and Correspondence as factors.

Results

8 –to 9.5-year-olds

Reaction times

There was no overall RT difference between the two child groups, $F\,(1, 30) = 0.248$, $p = .622$, $\eta_p^2 = .008$. Children were quicker to respond to symmetric stimuli and corresponding trials as reflected by the main effect of these factors, $F\,(1, 30) = 14.911$, $p = .001$, $\eta_p^2 = .332$, and $F\,(1, 30) = 8.529$, $p = .007$, $\eta_p^2 = .221$, respectively. The expected Jointness × Symmetry interaction was significant, $F\,(1, 30) = 5.480$, $p = .026$, $\eta_p^2 = .154$, and did not interact with Age, $F\,(1, 30) = 1.225$, $p = .277$, $\eta_p^2 = .039$. The interaction comes from a steeper RT increase for asymmetric than symmetric numbers from individual to joint blocks. Symmetry also interacted with Correspondence, $F\,(1, 30) = 14.989$, $p = .001$, $\eta_p^2 = .333$.[1] For descriptive statistical data, see Figure 1 (collapsed over age) and Table 1 (in each age group separately).

Efficiency score

Child participants' decisions were found to be more efficient for symmetric than asymmetric trials, $F(1, 30) = 13.153$, $p = .001$, $\eta_p^2 = .305$, and in the individual compared with the joint condition, $F(1, 30) = 5.710$, $p = .023$, $\eta_p^2 = .106$. Crucially, the interaction of Jointness and Symmetry was found to be significant, $F\,(1, 30) = 8.977$, $p = .005$, $\eta_p^2 = .230$ – the ES increasing from individual to joint blocks more for asymmetric than symmetric stimuli. Age did not influence the perspective interference effect, $F\,(1, 30) = 0.004$, $p = .952$, $\eta_p^2 = .000$. There was, however, a significant interaction of Symmetry and Correspondence, $F\,(1, 30) = 7.971$, $p = .008$, $\eta_p^2 = .210$, and a tendency-level interaction between Jointness, Symmetry, Correspondence, and Age, $F\,(1, 30) = 4.063$, $p = .053$, $\eta_p^2 = .119$.

This four-way interaction led us to run 2 × 2 ANOVAs in the two age groups separately to check whether the targeted effect was present in both groups. Eight-year-olds showed perspective interference, $F\,(1, 15) = 7.900$, $p = .013$, $\eta_p^2 = .345$, in addition to a main effect of Symmetry, $F(1, 15) = 6.415$, $p = .023$, $\eta_p^2 = .300$. In 9.5-year-olds, the Jointness × Symmetry interaction was marginally significant $F(1, 15) = 3.093$, $p = .099$, $\eta_p^2 = .171$.

[1] *A similar interaction has been reported in this paradigm before (Elekes et al., 2016). The effect is due to the longer RTs for non-corresponding symmetric trials compared with corresponding symmetric trials, and a lack of such difference for the asymmetric trials (see Table 1). Importantly the interaction was independent of both Jointness and Age, suggesting that it did not interfere with the perspective interference effect.*

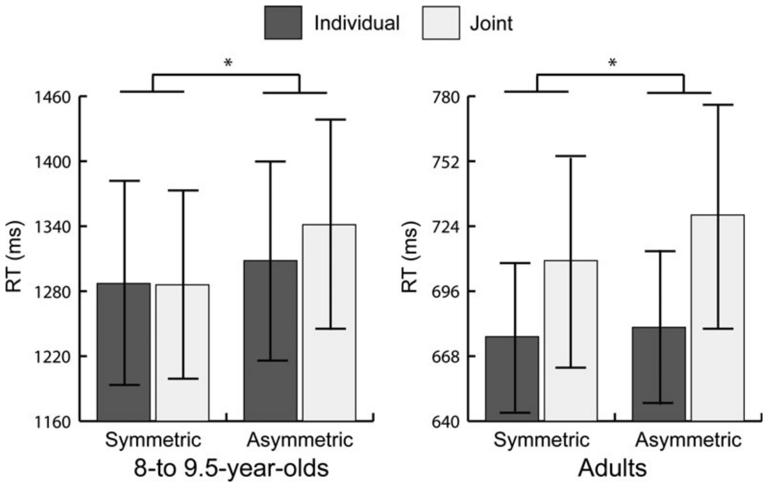


Figure 1. Mean RTs as a function of Jointness and Symmetry in children and adults. Error bars represent 95% confidence intervals. Asterisks indicate statistical significance of Jointness × Symmetry interactions, $*p < .05$.

Additionally, decisions tended to be more efficient in the individual condition, $F(1, 15) = 3.374$, $p = .086$, $\eta_p^2 = .184$, and were significantly more efficient for symmetric stimuli, $F(1, 15) = 7.291$, $p = .016$, $\eta_p^2 = .327$. See Figure 2.

Adults

Reaction times

The pattern of findings in the adult group confirms previous findings showing a significant Jointness × Symmetry interaction, $F(1, 15) = 4.668$, $p = .047$, $\eta_p^2 = .237$. Additionally,

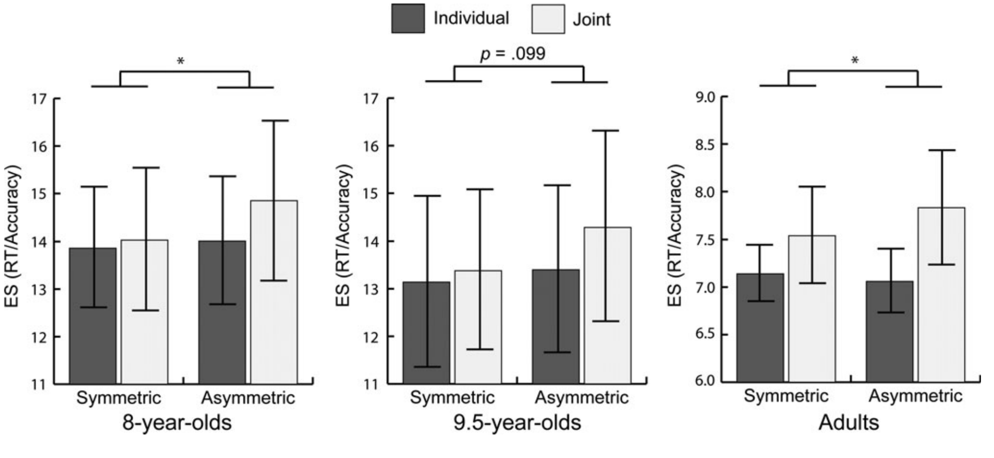


Figure 2. Mean efficiency scores in each age group as a function of Jointness and Symmetry. Error bars represent 95% confidence intervals. Asterisks indicate statistical significance of Jointness × Symmetry interactions, $*p < .05$.

Table 1. Mean RTs in milliseconds (with Standard Deviations in parentheses) as a function of Jointness, Symmetry, Correspondence, and Age

	Individual				Joint			
	Symmetric		Asymmetric		Symmetric		Asymmetric	
Age	Corresponding	Non-Corresponding	Corresponding	Non-Corresponding	Corresponding	Non-Corresponding	Corresponding	Non-Corresponding
Eight-year-olds	1,277.46 (220.99)	1,332.88 (199.64)	1,333.06 (232.41)	1,338.38 (244.32)	1,267.81 (199.85)	1,357.64 (213.41)	1,386.89 (226)	1,335.84 (236.31)
9.5-year-olds	1,223.42 (327.21)	1,314.6 (334.66)	1,300.04 (322.87)	1,261.44 (298.01)	1,198.46 (270.88)	1,319.65 (314.86)	1,317.92 (307.09)	1,324.72 (373.79)
Adults	639.74 (76.6)	713.08 (64.32)	682.39 (74.05)	678.49 (87.09)	678.98 (95.08)	739.38 (94.18)	732.56 (120.07)	725.09 (101.65)

adults' RTs were generally shorter on individual than on joint trials, $F(1, 15) = 5.192$, $p = .038$, $\eta_p^2 = .257$, and for corresponding as opposed to non-corresponding decisions, $F(1, 15) = 12.586$, $p = .003$, $\eta_p^2 = .456$. The Symmetry x Correspondence interaction was also significant, $F(1, 15) = 6.109$, $p = .026$, $\eta_p^2 = .289$.

Efficiency score

The ANOVA revealed the significant interaction of Jointness and Symmetry, $F(1, 15) = 5.006$, $p = .041$, $\eta_p^2 = .250$. The main effect of Jointness, $F(1, 15) = 7.917$, $p = .013$, $\eta_p^2 = .347$, and the interaction of Symmetry and Correspondence, $F(1, 15) = 6.143$, $p = .026$, $\eta_p^2 = .291$, were also statistically significant.

Discussion

The current study is the first to test not only adults but also school-aged children for spontaneous level-2 PT in the number verification paradigm with live interactional partners. We followed the one-system account's (Carruthers, 2017) prediction that those age groups that show spontaneous level-1 PT, for example school-aged children (Surtees & Apperly, 2012), should also be able to handle level-2 perspectives spontaneously. However, the one-system view also states that for a process to be performed spontaneously, its attentional and executive demands need to be relatively low. Applied to the current research, high processing demand could potentially mask participants' ability to access level-2 perspectives online.

Despite that due to their not yet automatized number reading skills (van Galen & Reitsma, 2008; Girelli *et al.*, 2000) and lower level of cognitive resources (attention, working memory), children face higher cognitive demand relative to their cognitive resources in the number verification paradigm than adults, we found evidence for spontaneous level-2 PT in all age groups.

The partner's conflicting perspective on jointly observed asymmetric numbers caused RTs, and ES to increase compared with symmetric stimuli both in children and adults. A separate analysis of the ES in the two child groups verified that, independently of the four-way interaction of all factors, level-2 PT was consistent over the whole sample of children. The interference effect was significant in eight-year-olds, whereas it was marginally significant in 9.5-year-olds, showing the expected pattern. The ES reinforces the findings on RTs by controlling for possible differences between groups regarding the trade-off between RT and accuracy. In addition to the level-2 PT effect, adults (and to a lesser extent 9.5-year-olds) showed a general decrease in performance from the individual to the joint condition. It may be that these groups were more sensitive to the social demands of the experimental situation than eight-year-olds that had a distracting effect on them.

Besides providing valuable findings regarding the development of spontaneous level-2 PT, our study also yields a replication for Elekes *et al.*' (2016) perspective-dependent group with adults, using confederates as partners in the task. The use of confederates, as opposed to naïve partners, provides a better controlled experimental set-up. At the same time, the presence of live partners, unlike avatars, keeps the reciprocal (though minimally) social nature of the situation.

Although the number verification paradigm was introduced as a measure of visual perspective taking (Surtees *et al.*, 2012), it relates to the capacity to take another spatial reference frame as well (Böckler, Knoblich, & Sebanz, 2011). This entails that the encoder spontaneously rotates himself or herself to another spatial frame of reference (i.e., that of the partner) leading to a conflict, instead of attributing percepts. That is, the process involves mental rotation, but lacks a meta-representational understanding of perspective. We dispute the first assumption, while do not rule out the second. Specifically, identification of rotated objects does not require mental rotation, either in adults or in children (Young, Palef, & Logan, 1980). Thus, even though the difference between visual percepts (6/9) in this paradigm is caused by different spatial perspectives on the same stimuli, the mechanism to reach the alternative content does not have to be mental rotation. On the other hand (and in line with the reasoning of Elekes *et al.*, 2016) computing a perspective content may dissociate from acknowledging it as an alternative description (representation) of the world. It is possible that the two alternative identities of the number character were extracted online, but were not contrasted as percepts referring to the same object. Thus, the effect may lack a metarepresentational nature, but still involve conflicting perspective content/meaning rather than spatial reference frames. Importantly, as long as the computation is selectively induced by certain social contexts, it serves the success of social interactions.

As current research on mindreading is, to a great degree, centred on the 'system debate', we need to relate our findings to this issue. The major claim of the two-system account is that there are two distinct systems in ToM embodying the trade-off between flexibility and efficiency. The first system functions automatically on belief-like states or registrations that do not include representing object appearances. The second system is flexible, it can handle differences in perceived object appearances and the aspectual nature of mental states, but is effortful in its operation and does not function online (Apperly & Butterfill, 2009; Butterfill & Apperly, 2013).

Probably as a response to findings on spontaneous level-2 PT (Elekes *et al.*, 2016; Surtees, Apperly, *et al.*, 2016; Surtees, Samson, *et al.*, 2016), recently, there was a shift from claiming that the latter system works offline and effortfully to a more lenient claim that it lacks genuine automaticity, which only features the first system (Low, Apperly, Butterfill, & Rakoczy, 2016; Surtees, Samson, *et al.*, 2016). However, the automaticity of the first system is challenged as well by findings that demonstrate that level-1 PT shows contextual selectivity, failing to meet the stimulus-driven nature of automatic processes. Namely, the level to which people engage in level-1 PT in the dot perspective task has been found to be influenced by participants' emotional states (Bukowski & Samson, 2016; Todd & Simpson, 2016), as well as the identity of the avatar: one that is representing the self or a stranger (Mattan *et al.*, 2015). The selectivity of implicit PT contradicts approaches (Cole, Atkinson, Le, & Smith, 2016; Santiesteban, Catmur, Hopkins, Bird, & Heyes, 2014) postulating that perspective interference reflects a 'blind', domain-general, non-social process.

The above findings on the flexibility of level-1 PT and the speed of level-2 PT converge to show that *both* forms of PT can emerge online, without instruction, and selectively in adults, raising the question whether assuming two independent systems was indeed necessary. Importantly, there is no empirical indication in our study that would allow us to make a strong stand for a single, unitary mindreading capacity. Our objective was a narrower, empirical question. The developmental literature fell behind recent empirical advances in adults in two respects. While adults had been shown to compute both level-1 and level-2 perspectives in implicit paradigms, and do so in a context-sensitive way, there seemed to be a strict dissociation between automatic level-1 and effortful level-2 PT in children. The findings reported here are the first steps to fill this void by showing online level-2 PT in school-aged children. Future research will have to test whether online PT

dc_1353_16

showed contextual selectivity in children as well, and develop paradigms that are adept to test the same capacity in preschool age.

Our findings support the notion that perspectives on object appearances can be accessed online both by adults and by eight- to 9.5-year-old children, instead of being reached through a slow and effortful process. It is possible that the closest a social cognitive process can be to automaticity is being spontaneous. Similar to automatic processes, a capacity that is quick and spontaneous may help to coordinate with other people while interactions are unfolding. However, as opposed to automatic processes, it also allows for sociocontextual factors to affect its functioning. That is, spontaneity enables selectively. Assuming that the main function of ToM capacities is to enable humans to flexibly adapt to the complex and ever changing social environment they live in, any capacity that is both quick *and* selective to environmental cues is beneficial to use.

Acknowledgements

This work was supported by OTKA, under grant number K-116779. We thank our child and adult participants for contributing to our study.

References

Apperly, I. A., & Butterfill, S. A. (2009). Do humans have two systems to track beliefs and belief-like states? *Psychological Review*, *116*, 953–970. https://doi.org/10.1037/a0016923

Bargh, J. A. (1989). Conditional Automaticity: Varieties of Automatic Influence in Social Perception and Cognition. In J. S. Uleman & J. A. Bargh (Eds.), *Unintended thought* (pp. 3–51). New York, NY: The Guilford Press.

Böckler, A., Knoblich, G., & Sebanz, N. (2011). Giving a helping hand: Effects of joint attention on mental rotation of body parts. *Experimental Brain Research*, *211*, 531–545. https://doi.org/10.1007/s00221-011-2625-z

Brocki, K. C., & Bohlin, G. (2004). Executive functions in children aged 6 to 13: A dimensional and developmental study. *Developmental Neuropsychology*, *26*, 571–593. https://doi.org/10.1207/s15326942dn2602_3

Bukowski, H., & Samson, D. (2016). Can emotions influence level-1 visual perspective taking? *Cognitive Neuroscience*, *7*, 182–191. https://doi.org/10.1080/17588928.2015.1043879

Butterfill, S. A., & Apperly, I. A. (2013). How to construct a minimal theory of mind. *Mind & Language*, *28*, 606–637. https://doi.org/10.1111/mila.12036

Carruthers, P. (2017). Mindreading in adults: Evaluating two-systems views. *Synthese*, *194*, 673–688. https://doi.org/10.1007/s11229-015-0792-3

Cole, G. G., Atkinson, M., Le, A. T. D., & Smith, D. T. (2016). Do humans spontaneously take the perspective of others? *Acta Psychologica*, *164*, 165–168. https://doi.org/10.1016/j.actpsy.2016.01.007

Csibra, G., & Gergely, G. (2009). Natural pedagogy. *Trends in Cognitive Sciences*, *13*, 148–153. https://doi.org/10.1016/j.tics.2009.01.005

Elekes, F., Varga, M., & Király, I. (2016). Evidence for spontaneous level-2 perspective taking in adults. *Consciousness and Cognition*, *41*, 93–103. https://doi.org/10.1016/j.concog.2016.02.010

Flavell, J. H., Everett, B. A., Croft, K., & Flavell, E. R. (1981). Young children's knowledge about visual perception: Further evidence for the level 1-level 2 distinction. *Developmental Psychology*, *17*, 99–103. https://doi.org/10.1037//0012-1649.17.1.99

Girelli, L., Lucangeli, D., & Butterworth, B. (2000). The development of automaticity in accessing number magnitude. *Journal of Experimental Child Psychology*, *76*, 104–122. https://doi.org/10.1006/jecp.2000.2564

dc_1353_16

Low, J., Apperly, I. A., Butterfill, S., & Rakoczy, H. (2016). Cognitive architecture of belief reasoning in children and adults: A primer on the two-systems account. *Child Development Perspectives*, *10*, 184–189. https://doi.org/10.1111/cdep.12183

Low, J., & Watts, J. (2013). Attributing false beliefs about object identity reveals a signature blind spot in humans' efficient mind-reading system. *Psychological Science*, *24*, 305–311. https://doi.org/10.1177/0956797612451469

Mattan, B. D., Quinn, K. A., Apperly, I. A., Sui, J., & Rotshtein, P. (2015). Is it always me first? Effects of self-tagging on third-person perspective-taking. *Journal of Experimental Psychology: Learning Memory and Cognition*, *41*, 1100–1117. https://doi.org/10.1037/xlm0000078

Moll, H., Meltzoff, A. N., Merzsch, K., & Tomasello, M. (2013). Taking versus confronting visual perspectives in preschool children. *Developmental Psychology*, *49*, 646. https://doi.org/10.1037/a0028633

Moors, A., & De Houwer, J. (2006). Automaticity: A theoretical and conceptual analysis. *Psychological Bulletin*, *132*, 297–326. https://doi.org/10.1037/0033-2909.132.2.297

Perner, J., Brandl, J. L., & Garnham, A. (2003). What is a perspective problem? Developmental issues in belief ascription and dual identity. *Facta Philosophica*, *5*, 355–378.

Rakoczy, H. (2012). Do infants have a theory of mind? *British Journal of Developmental Psychology*, *30*, 59–74. https://doi.org/10.1111/j.2044-835X.2011.02061.x

Rakoczy, H., Bergfeld, D., Schwarz, I., & Fizke, E. (2015). Explicit theory of mind is even more unified than previously assumed: Belief ascription and understanding aspectuality emerge together in development. *Child Development*, *86*, 486–502. https://doi.org/10.1111/cdev.12311

Rueda, M. R., Fan, J., McCandliss, B. D., Halparin, J. D., Gruber, D. B., Lercari, L. P., & Posner, M. I. (2004). Development of attentional networks in childhood. *Neuropsychologia*, *42*, 1029–1040. https://doi.org/10.1016/j.neuropsychologia.2003.12.012

Santiesteban, I., Catmur, C., Hopkins, S. C., Bird, G., & Heyes, C. (2014). Avatars and arrows: Implicit mentalizing or domain-general processing? *Journal of Experimental Psychology: Human Perception and Performance*, *40*, 929–937. https://doi.org/10.1037/a0035175

Sprung, M., Perner, J., & Mitchell, P. (2007). Opacity and discourse referents: Object identity and object properties. *Mind & Language*, *22*, 215–245. https://doi.org/10.1111/j.1468-0017.2007.00307.x

Surtees, A. D. R., & Apperly, I. A. (2012). Egocentrism and automatic perspective taking in children and adults. *Child Development*, *83*, 452–460. https://doi.org/10.1111/j.1467-8624.2011.01730.x

Surtees, A. D. R., Apperly, I. A., & Samson, D. (2013). Similarities and differences in visual and spatial perspective-taking processes. *Cognition*, *129*, 426–438. https://doi.org/10.1016/j.cognition.2013.06.008

Surtees, A., Apperly, I. A., & Samson, D. (2016). I've got your number: Spontaneous perspective-taking in an interactive task. *Cognition*, *150*, 43–52. https://doi.org/10.1016/j.cognition.2016.01.014

Surtees, A. D. R., Butterfill, S. A., & Apperly, I. A. (2012). Direct and indirect measures of Level-2 perspective-taking in children and adults. *British Journal of Developmental Psychology*, *30*, 75–86. https://doi.org/10.1111/j.2044-835X.2011.02063.x

Surtees, A., Samson, D., & Apperly, I. A. (2016). Unintentional perspective-taking calculates whether something is seen, but not how it is seen. *Cognition*, *148*, 97–105. https://doi.org/10.1016/j.cognition.2015.12.010

Todd, A. R., & Simpson, A. J. (2016). Anxiety impairs spontaneous perspective calculation: Evidence from a level-1 visual perspective-taking task. *Cognition*, *156*, 88–94. https://doi.org/10.1016/j.cognition.2016.08.004

Townsend, J. T., & Ashby, F. G. (1978). Methods of modeling capacity in simple processing systems. In J. Castellan & F. Restle (Eds.), *Cognitive theory*, (Vol. *3*, pp. 200–239). Hillsdale, NJ: Erlbaum.

van Galen, M. S., & Reitsma, P. (2008). Developing access to number magnitude: A study of the SNARC effect in 7- to 9-year-olds. *Journal of Experimental Child Psychology*, *101*, 99–113. https://doi.org/10.1016/j.jecp.2008.05.001

Westra, E. (2016). Spontaneous mindreading: A problem for the two-systems account. *Synthese*, 1–23. Advance online publication. https://doi.org/10.1007/s11229-016-1159-0
Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, *13*, 103–128. https://doi.org/10.1016/0010-0277(83)90004-5
Young, J. M., Palef, S. R., & Logan, G. D. (1980). The role of mental rotation in letter processing by children and adults. *Canadian Journal of Psychology/Revue canadienne de psychologie*, *34*, 265. https://doi.org/10.1037/h0081063

Received 30 June 2016; revised version received 30 June 2017

Appendix: Instructions

You are going to participate in this study together, both of you will have to make decisions on simple audio and visual stimuli. There will be parts when you work alone, during this time the other person will sit with his or her back to the participant who is working, and there will be a part when you work in parallel, during this time you will sit facing each other. Both of you will have a response box to indicate your answers. One after the other, you will hear numbers from a loudspeaker. During this, you will see number characters on the screen, one for each number heard from the loudspeaker. Your task is to decide whether the number you hear was the same as what you see. If the two numbers are the same, press the green button on the box; if they are different, press the red button. It is important that you answered correctly, but quickly.

IV. CULTURAL LEARNING AND NAÏVE SOCIOLOGY

The establishment of long-term knowledge base intertwines with the acquisition of cultural knowledge. The acquisition of shared, cultural knowledge demands that a) already young children should be able to recognize what is shared knowledge; b) already in the phase of acquisition children should be able to select information with respect to its potential relevance in relation to shared, cultural knowledge.

4.1 Thesis 10. The emotional and affiliative motives behind social categorization are preceded by a cognitive, epistemic function of identifying culturally knowledgeable individuals both for (1) acquiring knowledge, and (2) obtaining access to and maintaining a shared representational space in the service of successful interactions.

Oláh, K., **Király, I.** (in preparation). Selective Imitation of Conventional Tool-Users by 3-Year-Old Children

4.2 Thesis 11. Cultural knowledge represents an organized system. Supposedly, children understand that an attributed knowledge base might cover and unite different domains. Based on this, we hypothesize that children are able to form unified expectations induced by different cues on the background knowledge of a partner.

Oláh, K. Elekes, F., Bródy, G., **Király, I.** (2014). Cues of shared Knowledge induce social category formation, *PLoS ONE 9*(7): e101680. doi:10.1371/journal.pone.0101680.

4.3 Thesis 12. Naïve sociology contributes to the flexibility of social learning: when receiving a novel bit of information from a carrier of shared knowledge (e.g. cultural group), that piece of information is treated as part of the culturally shared representational space that the child intends to acquire.

Oláh, K., Elekes, F., Pető, R., Peres, K., **Király I.** (2016). 3-Year-Old Children Selectively Generalize Object Functions Following a Demonstration from a Linguistic In-group Member: Evidence from the Phenomenon of Scale Error. *Frontiers in Psychology, 24* June 2016, http://dx.doi.org/10.3389/fpsyg.2016.00963

Pető, R., Elekes, F, Oláh, K., **Király, I.** (in preparation). Learning How to Use a Tool –mutually exclusive tool-function mappings are selectively acquired from linguistic in-group models

Selective Imitation of Conventional Tool-Users by 3-Year-Old Children

RUNNING HEAD: IMITATION OF CONVENTIONAL TOOL-USERS

Katalin Oláh[1,2] & Ildikó Király[1,3]

[1]Institute of Psychology, Eötvös Loránd University

[2]Insitute of Cognitive Neuroscience and Psychology, Hungarian Academy of Sciences

[3]Central European University

Author note:

This study was supported by the grant OTKA 109 352 (PI: Ildikó Király). The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) / ERC grant agreement n° 609819.

Corresponding author:

Katalin Oláh

olah.katalin@ppk.elte.hu

Abstract

This study investigated whether toddlers would selectively imitate a demonstrator who exhibits familiarity with cultural practices in their tool-using habits over a demonstrator who consistently uses tools in an unconventional way. 3-year-old children (n=45) watched videos depicting two models, one of whom performed too-using actions in a conventional way, while the other model deviated from social conventions. Then, both models introduced a technique to build a tower (differing in one element). Moreover, the context of the demonstration was also manipulated: in one condition, the models expressed their teaching intentions, while in the other, they performed the actions without communicative signals. Children were more willing to copy the actions of the conventionally behaving model, irrespective of the context of the demonstration.

Human social learning is characterized by two – seemingly contradictory – important features. First, knowledge transmission should happen without applying much modification to the content of the information to ensure that behavioral patterns that provide the foundations of society stabilize, even when the function of certain actions are causally opaque to the observer (see Csibra & Gergely, 2009). Second, novices should be able to filter out irrelevant information from the excess of stimuli reaching them at every moment. Young children seem to rely on a number of cues that may distinguish relevant pieces of information and behavioral patterns from irrelevant ones. For example, children from a very early age take into account the intentions underlying action demonstration from potential teachers. Thus, children will not reproduce actions that are merely accidental (Carpenter, Akhtar & Tomasello, 1998) and already infants possess a special sensitivity to detect the teaching intentions of others, helping them to identify those elements of an episode that are worth learning (as described in Natural Pedagogy theory: Csibra & Gergely, 2006; Csibra & Gergely, 2009). Empirical evidence lend support to the claim about the existence of an innate system making humans adept at detecting so-called ostensive-referential signals that highlight an episode as pedagogical (e.g. Senju & Csibra, 2008; Yoon, Johnson & Csibra, 2008; Futó, Téglás, Csibra & Gergely, 2010; Topál, Gergely, Miklósi, Erdőhegyi & Csibra, 2008). A number of studies have also shown that this sensitivity guides imitative behavior in children: unusual actions will be less likely to be copied if the action demonstration is not preceded by communicative signals (Király, Csibra & Gergely, 2013).

In addition to selection processes that target the content of the transmission process, effective social learning should also be supported by mechanisms that help novices to make judgments about the reliability of the source of information as well. Such mechanisms should guide selection between *potential informants*. A number of studies confirm the notion that children already from a young age are discriminative in who they accept information from. For example, 3-to-5-year-old children are more willing to endorse object labels provided by a familiar individual than an unfamiliar one (Corriveau& Harris, 2009a); however this initial trust evoked by familiarity is overridden by cues of accuracy (in labeling familiar objects) for 4- and 5-year-olds. Several other studies have also highlighted the importance of the past accuracy of the informant in guiding children's learning processes (e.g. Pasquini, Corriveau, Koenig & Harris, 2007; Koenig & Harris, 2005; Koening, Clément & Harris, 2004). Moreover, children also retain these impressions of reliability and continue to prefer an accurate individual as information source one week after the first exposure to the potential informants (Corriveau & Harris, 2009b). On a similar vein, 14-month-old infants have been shown to monitor reliability in emotional expression and selectively imitate models that have proved reliable in this respect (Poulin-Dubois, Brooker, Polonia, 2011). At the same age, infants attend to cues of confidence and appropriate usage of tools in deciding whom to imitate (Zmyj, Buttelmann, Carpenter & Daum, 2010). 14-month-old infants are also more willing to copy novel instrumental actions performed by adults than by children (Zmyj, Daum, Prinz, Nielsen & Aschersleben, 2012; Jaswal & Neely, 2006).

Taken together, the results described above suggest that young children pay attention to cues that provide information about the knowledgeability of potential teachers in order to selectively endorse information that is most likely to be useful and appropriate. However, adaptive social learning mechanisms in humans also have to answer the challenge that lies in the diverse nature of cultural practices and adaptive behavioral patterns. Cumulative cultural evolution has led to significant variations among social groups in the scope of adaptive behavioral patterns that support survival in a particular environment. These specific behavioral patterns and knowledge have to be transmitted through generations with the help of adaptive social learning mechanisms (Boyd & Richerson, 1996; Henrich & McElreath, 2003). We argue that under such circumstances, one challenge novices are faced with during the transmission process is that potential information sources may be reliable in one context but not in the other. Novices should not only favor informants that are confident or experienced in general, but that possess knowledge that is valid in the specific social environment they grow up in. In other words, novices should be prepared to selectively endorse information coming from members of their own social group („in-groups”). A handful of studies have already confirmed that young children indeed show selectivity based on group membership. Buttelmann, Zmyj, Daum and Carpenter (2013) have shown that infants as young as 14-month-old selectively imitate linguistic in-group members over people speaking in a foreign language. Howard, Henderson, Carrazza and Woodward (2015) reported similar findings with 19-month-old and 3-year-old children with the constraint that the younger age group only showed selectivity when the potential informants were presented on video. We propose that language cues are effective in guiding learning processes and serve as a salient cue for social categorization because they provide direct evidence about whether the informant shares cultural knowledge with the child and thus is capable of transmitting information that is valid in their social environment.

However, language may not be the only cue that informs novices about the cultural knowledgeability of the potential teacher. Language is a reliable marker as humans possess an innate sensitivity and preference to speech (Vouloumanos & Werker, 2007) and an early-developing ability to detect subtle discrepancies in it (Nazzi, Bertoncini & Mehler, 1998). Moreover, these discrepancies reliably signal the boundaries of both broader (foreign language) and narrower (foreign accent) social categories. Nevertheless, we claim that language is merely one possible – though strong – cue to possessing knowledge that is specific to the child’s social group. To test this idea, this study explores whether other possible markers of cultural knowledgeability would produce convergent results in an imitation paradigm. Our candidate marker is tool-using habits as humans already from early childhood have a special stance toward artifact functions (Kelemen & Carey, 2007; Casler & Kelemen, 2007) that make fast and efficient learning about tool-functions possible (Casler & Kelemen, 2005) and results in viewing artifact functions as normative (Casler, Terziyan & Greene, 2009). Moreover, it has been shown that children form similar representations based on the language a person speaks and the level of conventionality they exhibit in their tool-using behavior (Oláh, Elekes, Bródy & Király, 2014). Thus, we investigated whether children would selectively imitate a model whose tool-using habits conform to the cultural norms over someone who violates the cultural norms.

In addition, we also tested how ostensive communication would modulate any potential effect of the model's group membership. As described above, according to the theory of Natural Pedagogy (Csibra & Gergely, 2009), an innate sensitivity to ostensive-communicative signals foster the transmission of culturally relevant knowledge in humans by pointing out the to-be-acquired information. Thus, communicative signals and the different qualities of the teacher both serve to help children acquire culturally relevant knowledge; however, little is known about how these cues interact with each other in forming the behavior of children. One possibility is that children only attend to the communicative intentions of others if the person has been proven to be a reliable source of information. In this case, children would be equally (un)willing to imitate an unconventionally behaving model following a communicative and a non-communicative action demonstration, but they would show increased motivation to copy the actions of a conventionally behaving model after a communicative demonstration. The second possibility is that children's tendency to accept knowledge in a communicative setting is so strong that it overwrites the significance of cues of familiarity with cultural practices. In this case, children would imitate a model that gives ostensive signals irrespective of past behavior and would only differentiate based on it in the absence of such cues. To investigate the interplay of these two factors, we presented children with videos introducing two models, one of whom performed conventional tool-using actions, while the other used the same tools in an unconventional way. After that, both models demonstrated how to build a tower from building blocks either in a communicative or a non-communicative way. The two demonstrations varied in one element and we analyzed whether children would be more willing to copy the variant introduced by the conventionally behaving model. Importantly, both models either expressed their intention to teach (Communicative condition) or did not give any evidence of it (Non-communicative condition) and there were no conditions where the behavior of the models differed in this respect. This ensured that our two factors of interest (conventionality of behavior and communicativeness) were manipulated independently and thus the design would allow us to draw inferences whether one of the factors would have the power to overshadow the significance of the other.

Methods

Participants

50 3-year-old children participated in the study (mean: 39.3 months; SD: 2 months; range: 34-43 months). Children were either tested in one of two kindergartens (n=38) or in the baby lab (n=12). All children were monolingual. 5 children had to be excluded from the Ostensive condition due to passivity (1); having a distracting toy in their hand during testing (1); touching the apparatus too early (1) or not paying attention to the demonstration videos (2). The final sample consisted of 21 children in the Ostensive condition and 24 children in the Non-ostensive condition.

Materials

For the familiarization phase, two sets of videos were recorded of two protagonists. The videos depicted simple tool-using actions based on the stimuli developed for the study of Oláh et al. (2014). Each protagonist demonstrated two different tool-using actions either in a conventional way (cutting up a piece of paper using a pair scissors and having a bite of food with a fork) or in an unconventional way (cutting up a banana using a pair of scissors and combing one's hair with a fork). Each demonstration video was recorded with both protagonists in both manners. In addition, test videos were recorded with the protagonists that also had two different versions. The demonstrated action was a tower building technique, where the protagonist showed how to build a tower from three (or four) building blocks: a blue building block that was used as the base, a yellow middle section and a red top. Crucially, the middle section could be built either from a single double block or by adding two single blocks (see Fig. 1). All test videos had an ostensive-communicative and a non-communicative version and all versions were recorded with both participants. In the ostensive videos, the protagonist started the demonstration with looking into a camera, waving and saying „Hi". She finished the action by looking back up into the camera. In the non-ostensive videos, she simply started the demonstration with reaching for the first building block and did not look back at the camera in the end.

Insert Figure 1 about here

For the imitation phase, the tools used in the test videos were presented for the children. Namely, the blue building block (constituting the base of the tower), the red building block (constituting the top of the tower) and the elements that could potentially be used for the middle section: the double yellow building block and the two single yellow blocks.

Procedure

Children were tested individually either in a quiet room of the kindergarten or the laboratory. After escorting the child into the testing area, the experimenter told the participant that they would be watching short movies of two girls and that they should pay close attention to what happens. After that, the experimenter played the familiarization and the test videos. Each participant saw one of the protagonists perform both of the familiarization actions in a conventional way, while the other protagonists performed both actions in an unconventional way. With this, we wanted to create the impression that one protagonist consistently behaves according to social norms, while the other consistently deviates from them. Importantly, their actions were always performed in a confident way and were efficient in bringing about the desired goal. Children first watched the „fork action” being performed by both protagonists and then saw the second action („scissors action”) being performed by the two models in the order they appeared in the first pair of videos. After the familiarization videos, the two test videos immediately followed. Children saw one of the participants perform the tower building action with constructing the middle section from two pieces and the other protagonist building the middle part from one piece. Both participants performed the test action either in an ostensive way (Ostensive condition) or in a non-ostensive way (Non-ostensive condition). The following factors were counterbalanced across conditions: identity of the conventionally behaving model, the variant of the building technique performed by the conventionally behaving model and the order of appearance of the two models.

Coding

We coded whether children would choose to build the middle section from one block or two. Since children introduced significant variations into the building procedure, the following criterion was used: if children took either the double building block or the two separate pieces and placed them on the building block serving as base, then this was considered a clear choice irrespective of how they continued the building (in many cases, children ended up using up all the building blocks to build an even higher tower). If this element was missing or was performed in a completely different way (e.g. putting the two separate blocks on top of each other), then the behavior was coded as an alternative solution. A second coder coded 60% of the videos. All discrepancies between coders have been discussed and resolved.

Results

Statistical analyses were performed in SPSS 20.0. Our main question was whether children would be more willing to imitate a conventionally behaving model than a person violating the cultural norms. Therefore, our dependent variable was which model children imitated. After coding the videos, we observed that a large proportion of participants came up with an alternative solution. For this reason, first we analyzed all the data, including children with alternative solutions. In this analysis, we used a dependent variable with three possible values (imitating the conventional model/imitating the unconventional model and alternative solution). Additionally, we performed an analysis that included only children with clear choices (children following one protagonist).

To test the effects of condition on children’s choices between the conventionally and unconventionally behaving model, we first conducted regression analyses with choice of model as the dependent variable and condition (Ostensive-Non-ostensive), identity of the conventional model, order of presentation of the models, variant performed by the conventional model, testing location, age and sex as predictor variables. Since none of the predictor effects reached significance (all $p>0.76$ for the analyses excluding alternative solutions and all $p>0.904$ for the analyses including alternative solutions), we restricted the analyses to the factors of interest (choice of model and ostensiveness of the demonstration) for our research question and used Chi-square tests, Additionally, we conducted tests of distribution to explore whether children were generally more inclined to imitate the conventionally behaving model (Kolmogorov-Smirnov for the variable with three values and binomial test for the one with two values).

Analyses Including All Participants

The results of the Chi-square tests show that there was no difference between conditions in the number of children choosing to follow either of the models or opting for an alternative solution ($\chi^2(2)=0.277$; $p=0.87$). The results show that more than half of the children in both conditions imitated the variant introduced by the conventionally behaving model with around the same number of participants choosing to copy the unconventionally behaving model and to come up with a new method of tower building (see Figure 2). Analyzing the distribution of behavior types across conditions, we found a significant difference between the different response types ($Z=2.311$; $p<0.001$), showing that participants performed the variant they had seen from the conventionally behaving model most often.

Insert Figure 2 about here

Analyses Excluding Alternative Solutions

Due to the fact that they came up with a novel building method, 12 children were excluded from this analysis, leaving 17 children in the Non-ostensive and 16 children in the Ostensive condition. Similarly to the results of the first analysis, we found no difference in the distribution of behavior types between the conditions ($\chi^2(1)=0.113$; $p=0.737$); showing that the majority of children imitated the conventionally behaving model in both conditions (n=13 in both conditions). A binomial test showed that children were altogether significantly more likely to perform the variant introduced by the conventionally behaving model (p=0.001). The results are depicted on Figure 3.

Insert Figure 3 about here

Discussion

This study tested whether 3-year-old children would selectively imitate a model whose competence in cultural knowledge was indicated by their tool-using habits. The results confirmed our hypothesis, showing that children were more willing to learn from someone whose behavior conformed to the cultural norms over someone who violated the culturally established norms. Oláh et al. (2014) provided evidence that the same behaviors that we used for familiarization in this study are associated with language use in children's representations. Therefore, we propose that there may be a parallel in the selection mechanisms of children's learning processes exhibited in our study and those showing selectivity based on linguistic cues (e.g. Buttelmann et al., 2013; Howard, et al., 2015, Oláh, Elekes, Pető, Peres & Király, 2016). Both of these cues (language and conventionality in tool-using behavior) imply familiarity with the ways of a given culture; therefore these selection mechanisms ensure that children endorse information that will likely be useful within their own environments.

The results also show that the selectivity based on the models' prior behavior was not affected by the expression of communicative intentions in the test phase: children were just as unwilling to follow the behavior of an unconventionally behaving model in this case as they were when the models performed the actions in a non-communicative way. Thus, it seems that toddlers first identify the circle of reliable teachers and are reluctant to respond to the teaching intentions of those who fall outside of this circle. It is important to note that our study applied a forced choice method where children were always presented with a variant both from the conventionally and the unconventionally behaving model. Thus, it is possible that since children could not simply base their decisions about whom to follow on the perception of teaching intentions, they looked for other cues that could serve as guidance. We cannot be absolutely sure whether children would not imitate someone who does not keep to the cultural conventions but expresses their intentions to pass on knowledge if they are not presented with an alternative. However, similar studies on selective imitation of linguistic in-group members usually apply a between subjects method and work with a communicative demonstration and also report reduced imitation rates of an out-group member (e.g. Howard et al., 2015, but see Buttelmann et al., (2013) for the same finding in a not particularly ostensive context). Thus, given the parallels between children's reactions to linguistic out-group models and non-conformists to cultural norms, we would expect to see similar reluctance to imitate the latter following a communicative demonstration even when no alternative is presented. Note that in our design there were no conditions that directly pitted ostention againtst conventionality, that is, where one model was conventional but produced no communicative signals, whereas the other behaved in a non-conventional way but expressed their willingness to teach. The reason for this choice was the fact that a potential null-result (i.e. half of the children imitating one model, while the other half imitating the other model) would have been difficult to interpret: it could raise the possibility that both signals have equal significance in children's eyes and that is why children chose at random; however it would be difficult to separate this interpretation from other (e.g. low-level) explanatations.

We believe it very likely that the relative importance of communicative cues and cues about cultural identity undergoes significant changes in the first years of life. It is possible that both the sensitivity to ostensive-referential signals (Csibra & Gergely, 2006) and the tendency to select teachers based on perceived group membership have innate roots; however, the latter is more strongly dependent on already stored information. Therefore, it may be adaptive for younger children to learn everything presented in a communicative context and later use the accumulated knowledge as anchors in subsequent learning episodes. It may also be an efficient strategy considering that the circle of people children meet in the first months of their lives is usually much more limited than in later years and is less likely to include people who may otherwise not be a part of the wider social group of children and would therefore communicate knowledge that is not valid for children.

An important question that arises is whether the unconventional behaviors used in the familiarization phase in our study would lead children to form the impression that the person does not share cultural knowledge with themselves and consequently cannot be regarded as a member of the same cultural group. Children could have simply inferred that the person is „ignorant”, „funny” or a „rule-breaker”. A number of studies have shown that children show selective learning based on similar behavior cues implying that the knowledge of the potential teacher is not reliable (e.g. Pasquini et al., 2007; Koenig & Harris, 2005). The study by Zmyj and colleagues (2010) applied a very similar method to ours where they introduced a model whose behavior deviated from the cultural norms and importantly, who also signaled uncertainty about how to use the tools in front of him. In our study, the models always performed the actions with confidence in order to suggest that the person was not lacking knowledge, simply possessed *different* knowledge about the usage of the tools. Nonetheless, it is possible that children at this age do not differentiate between the two cases and treat both an uncertain model and a confident, but unconventionally behaving one as equally ignorant. Future research may address this question. However, even if children may not make the difference, the fact that they base their judgments about knowledgeability on conventionality of behavior is in itself informative. There may be other cues that could serve equally well as guidance about knowledgeability if the concept did not inherently include familiarity with cultural practices. For example, children could rely more strongly on cues of confidence or on the efficiency of the observed action. In our familiarization videos, the unconventional actions were always efficient in bringing about the highlighted goal. Children could also make the assumption that a person who finds a way to arrive at their goal is worth following, however this was not the case: familiarity with the mean to the goal played a crucial role. Moreover, mere familiarity was not sufficient to evoke trust as both the goal and the means were familiar in all the cases. „Unconventionality" was defined as the unexpected association of the two (otherwise familiar) elements of the actions, therefore beyond a sense of familiarity, top-down mechanisms sensitive to more subtle characteristics of the organization of behavior had to play a part. Thus, we suggest that for children (and adults as well), „knowledgeability" always includes familiarity with cultural practices. Therefore, even if children cannot explicitly postulate this, an unconventionally behaving person is not simply „stupid" but not being a good (conformist) member of a given social group as they do not share the established cultural knowledge.

To our knowledge, this study is the first one to show that language may not be the only relevant cue that provides grounds for selectivity in learning through signaling access to a specific body of cultural knowledge. We propose that tool-using habits and language both show children whether the interaction partner shares cultural knowledge with them and are reluctant to learn from someone who appears ignorant in this respect. This selectivity ensures that children accumulate knowledge that is valid and useful in their social environment and filter out irrelevant pieces of information from the exceess of stimuli reaching their cognitive system. We propose that the same sensitivity to known and unknown behavioral patterns (conformity to established practices) guiding children's learning processes helps humans navigate a world filled with multiple dimensions of subcultures in adulthood as well. .

References

Boyd, R., & Richerson, P. J. (1996). Why culture is common, but cultural evolution is rare. In Proceedings-British Academy (Vol. 88, pp. 77-94). Oxford University Press Inc.

Buttelmann, D., Zmyj, N., Daum., M., Carpenter, M. (2013). Selective imitation of ingroup over outgroup member in 14-month-old infants. Child Development 84, 422–428. doi:10.1111/j.1467-8624.2012.01860.x

Carpenter, M., Akhtar, N., & Tomasello, M. (1998). Fourteen-through 18-month-old infants differentially imitate intentional and accidental actions. Infant behavior and development, 21(2), 315-330. doi: 10.1016/S0163-6383(98)90009-1

Casler, K., & Kelemen, D. (2005). Young children's rapid learning about artifacts. Developmental science, 8(6), 472-480. doi:10.1111/j.1467-7687.2005.00438.x

Casler, K., & Kelemen, D. (2007). Reasoning about artifacts at 24 months: The developing teleo-functional stance. Cognition, 103(1), 120-130. http://doi.org/10.1016/j.cognition.2006.02.006

Corriveau, K., & Harris, P. L. (2009a). Choosing your informant: Weighing familiarity and recent accuracy. Developmental science, 12(3), 426-437. doi:10.1111/j.1467-7687.2008.00792.x

Corriveau, K., & Harris, P. L. (2009b). Preschoolers continue to trust a more accurate informant 1 week after exposure to accuracy information. Developmental science, 12(1), 188-193. doi:10.1111/j.1467-7687.2008.00763.x

Csibra, G., & Gergely, G. (2006). Social learning and social cognition: The case for pedagogy. Processes of change in brain and cognitive development. Attention and performance XXI, 21, 249-274.

Csibra, G., & Gergely, G. (2009). Natural pedagogy. Trends in cognitive sciences, 13(4), 148-153. http://doi.org/10.1016/j.tics.2009.01.005

Futó, J., Téglás, E., Csibra, G., & Gergely, G. (2010). Communicative function demonstration induces kind-based artifact representation in preverbal infants. Cognition, 117(1), 1-8. http://doi.org/10.1016/j.cognition.2010.06.003

Henrich, J., & McElreath, R. (2003). The evolution of cultural evolution. Evolutionary Anthropology: Issues, News, and Reviews, 12(3), 123-135. doi:10.1002/evan.10110

Howard, L.H., Henderson, A. M., Carrazza, C., Woodward, A.L. (2015). Infants’ and young children’s imitation of linguistic ingroup informants. Child Development 86, 259–275. doi:10.1111/cdev.12299

Jaswal, V. K., & Neely, L. A. (2006). Adults don't always know best preschoolers use past reliability over age when learning new words. Psychological Science, 17(9), 757-758. https://doi.org/10.1111/j.1467-9280.2006.01778.x

Kelemen, D., & Carey, S. (2007). The essence of artifacts: Developing the design stance. Creations of the mind: Theories of artifacts and their representation, 212-230.

Király, I., Csibra, G., & Gergely, G. (2013). Beyond rational imitation: Learning arbitrary means actions from communicative demonstrations. Journal of Experimental Child Psychology, 116(2), 471-486. http://doi.org/10.1016/j.jecp.2012.12.003

Koenig, M. A., Clément, F., & Harris, P. L. (2004). Trust in testimony: Children's use of true and false statements. Psychological Science, 15(10), 694-698. https://doi.org/10.1111/j.0956-7976.2004.00742.x

Koenig, M. A., & Harris, P. L. (2005). Preschoolers mistrust ignorant and inaccurate speakers. Child development, 76(6), 1261-1277. doi:10.1111/j.1467-8624.2005.00849.x

Nazzi, T., Bertoncini, J., & Mehler, J. (1998). Language discrimination by newborns: toward an understanding of the role of rhythm. Journal of Experimental Psychology: Human perception and performance, 24(3), 756. http://dx.doi.org/10.1037/0096-1523.24.3.756

Oláh, K., Elekes, F., Bródy, G., & Király, I. (2014). Social category formation is induced by cues of sharing knowledge in young children. PloS one, 9(7), e101680. http://dx.doi.org/10.1371/journal.pone.0101680

Oláh, K., Elekes, F., Pető, R., Peres, K., & Király, I. (2016). 3-Year-Old Children Selectively Generalize Object Functions Following a Demonstration from a Linguistic In-group Member: Evidence from the Phenomenon of Scale Error. Frontiers in Psychology, 7. doi: 10.3389/fpsyg.2016.00963

Pasquini, E. S., Corriveau, K. H., Koenig, M., & Harris, P. L. (2007). Preschoolers monitor the relative accuracy of informants. Developmental psychology, 43(5), 1216. http://dx.doi.org/10.1037/0012-1649.43.5.1216

Poulin-Dubois, D., Brooker, I., & Polonia, A. (2011). Infants prefer to imitate a reliable person. Infant Behavior and Development, 34(2), 303-309. http://dx.doi.org/10.1037/0012-1649.43.5.1216

Senju, A., & Csibra, G. (2008). Gaze following in human infants depends on communicative signals. Current Biology, 18(9), 668-671. http://doi.org/10.1016/j.cub.2008.03.059

Spelke, E. S., & Kinzler, K. D. (2007). Core knowledge. Developmental science, 10(1), 89-96. doi:10.1111/j.1467-7687.2007.00569.x

Topál, J., Gergely, G., Miklósi, Á., Erdőhegyi, Á., & Csibra, G. (2008). Infants' perseverative search errors are induced by pragmatic misinterpretation. Science, 321(5897), 1831-1834. doi: 10.1126/science.1161437

Vouloumanos, A., & Werker, J. F. (2007). Listening to language at birth: Evidence for a bias for speech in neonates. Developmental science, 10(2), 159-164. doi:10.1111/j.1467-7687.2007.00549.x

Yoon, J. M., Johnson, M. H., & Csibra, G. (2008). Communication-induced memory biases in preverbal infants. Proceedings of the National Academy of Sciences, 105(36), 13690-13695. doi: 10.1073/pnas.0804388105

Zmyj, N., Buttelmann, D., Carpenter, M., & Daum, M. M. (2010). The reliability of a model influences 14-month-olds’ imitation. Journal of Experimental Child Psychology, 106(4), 208-220. http://doi.org/10.1016/j.jecp.2010.03.002

Zmyj, N., Daum, M. M., Prinz, W., Nielsen, M., & Aschersleben, G. (2012). Fourteen-Month-Olds' Imitation of Differently Aged Models. Infant and Child Development, 21(3), 250-266. doi:10.1002/icd.750

Acknowledgments

This study was supported by the grant OTKA 109 352 (PI: Ildikó Király). The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) / ERC grant agreement n° 609819.We thank Dániel Németh, Júlia Baross, Iulia Savos, Réka Pető and Zsófia Kalina for their help with data collection and Fruzsina Elekes for her help with the preparation of the experimental design and stimulus.

Author contribution

KO and IK designed the experiment, KO performed the experiments and analyzed the data and KO and IK wrote the paper.

Additional Information

Competing financial interests

The authors declare no competing financial interests.

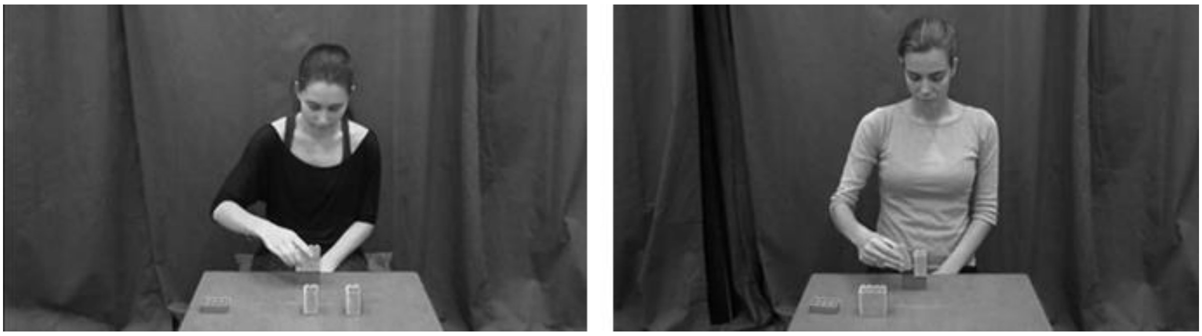


Figure 1. Two possible ways to build the tower (using two single blocks or one double block to build the middle section) performed by the two protagonist.

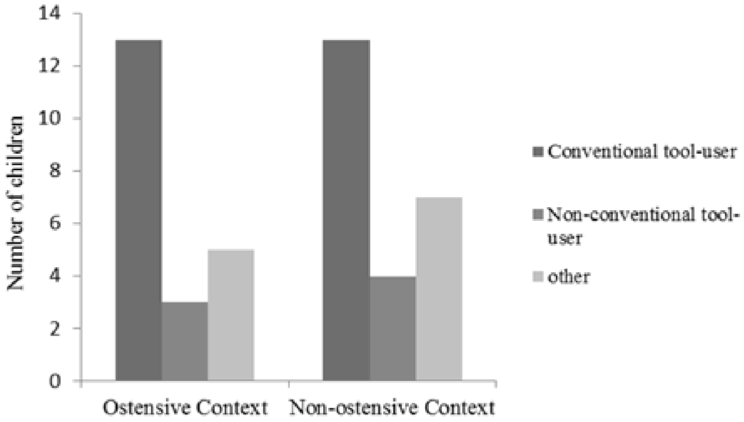


Figure 2. Number of children imitating the variants introduced by the two models or opting for an alternative solution in the Ostensive and the Non-ostensive conditions.

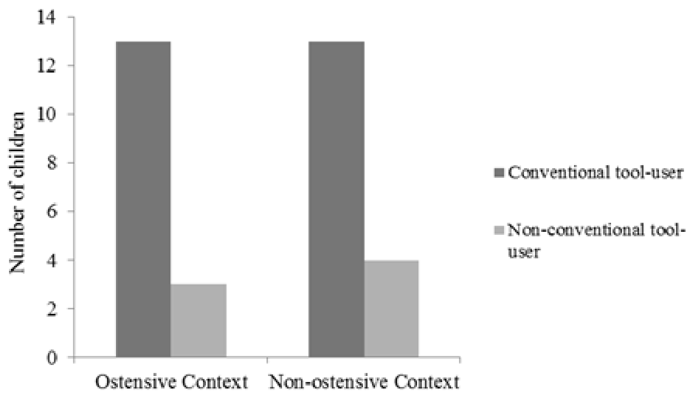


Figure 3. Number of children imitating the variants introduced by the two models in the Ostensive and the Non-ostensive conditions (excluding alternative solutions).

OPEN ACCESS Freely available online
PLOS | ONE

Social Category Formation Is Induced by Cues of Sharing Knowledge in Young Children

Katalin Oláh[1,2]*, **Fruzsina Elekes**[1,2], **Gábor Bródy**[1], **Ildikó Király**[1]

1 Department of Cognitive Psychology, Eötvös Loránd University, Budapest, Hungary, **2** Institute of Cognitive Neuroscience and Psychology, Hungarian Academy of Sciences, Budapest, Hungary

Abstract

Previous research has shown that human infants and young children are sensitive to the boundaries of certain social groups, which supports the idea that the capacity to represent social categories constitutes a fundamental characteristic of the human cognitive system. However, the function this capacity serves is still debated. We propose that during social categorization the human mind aims at mapping out social groups defined by a certain set of shared knowledge. An eye-tracking paradigm was designed to test whether two-year-old children differentially associate conventional versus non-conventional tool use with language-use, reflecting an organization of information that is induced by cues of shared knowledge. Children first watched videos depicting a male model perform goal-directed actions either in a conventional or in a non-conventional way. In the test phase children were presented with photographs taken of the model and of a similarly aged unfamiliar person while listening to a foreign (Experiment 1) or a native language (Experiment 2) text. Upon hearing the foreign utterance children looked at the model first if he had been seen to act in an unconventional way during familiarization. In contrast, children looked at the other person if the model had performed conventional tool use actions. No such differences were found in case of the native language. The results suggest that children take the conventionality of behavior into account in forming representations about a person, and they generalize to other qualities of the person based on this information.

Citation: Oláh K, Elekes F, Bródy G, Király I (2014) Social Category Formation Is Induced by Cues of Sharing Knowledge in Young Children. PLoS ONE 9(7): e101680. doi:10.1371/journal.pone.0101680

Editor: Howard Nusbaum, The University of Chicago, United States of America

Received July 24, 2013; **Accepted** June 10, 2014; **Published** July 11, 2014

Funding: The project received financial funding from the National Development Agency (URL: www.nfu.hu/?lang = en, grant number: KMOP-4.2.1/B-10-2011-0002) and the number: K-109352 and K-100695). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* Email: olah.katalin@ttk.mta.hu

Introduction

Human societies are unique among species in that our form of living entails a level of interdependence between conspecifics that cannot be found in any other animal species. Humans form alliances with other people for different purposes every day, during which they engage in a wide variety of joint actions and collaboration towards goals, such as creating artifacts, working towards scientific discoveries or – even – doing sports together [1,2]. Group-living constitutes a fundamental characteristic of the human race and group-affiliations can influence various aspects of our lives to a great extent. Therefore the ability to form cognitive representations of human groups can be seen as an evolutionary adaptive capacity of the human brain.

The phenomenon that adults have a propensity to think of fellow humans as belonging to groups has been well documented in the social psychology literature (see [3] for a review) and there is also ample evidence suggesting that even infants are able to perceive the boundaries of certain social categories. Studies conducted in the field of developmental psychology as well as cultural anthropology have demonstrated that young children or even infants are sensitive to social categories such as age [4], sex [5–9], race [10–14]. It has also been demonstrated that children as young as 2.5 years of age – similarly to adults [15] – are willing to accept even arbitrary cues of group membership when forming expectations about a person's behavior [16], however other studies suggest that some classifications have more limited power in guiding young children's behavior towards people [17].

More recently, in a new line of research, Kinzler and colleagues have shown that spoken language occupies a prominent role in young children's representations of social categories [18–20]. Their findings also indicate that this distinction is more privileged in children's eyes than some other, such as race [21], which suggests that language taps into some of the more fundamental processes underlying the propensity of category-based thinking. One of the most important findings of these experiments is that the accent with which a person speaks has a stronger effect on 5-year-olds attitude toward that person than their ethnicity.

Despite the great and long interest and the social relevance of the question of social categorization, the cognitive basis and the function of this process is still somewhat under-explained. Theories in social psychology have provided numerous accounts of how affiliating with a group may benefit an individual in – for example – boosting their self-esteem [15,22]; justifying a bigger share of resources [23–24], etc. These accounts generally tend to emphasize the competition between different social groups and assume an antagonist relationship between them. Theorists from various fields of cognitive sciences tend to grasp a different aspect of the question and offer other explanations with respect to the function of group-based thinking. Sperber and Hirschfeld [25] for example argue that the human mind has evolved a special domain to reason about social kinds in order to make sense of the extremely complicated structure of human societies. Similarly, Cosmides, Tooby, and Kurzban [26] propose that the primary function of representing social categories is to map out potential coalitional partners.

An adaptive system that is responsible for representing social groups must correspond to the core characteristics and advantages of group living. Forming short or long term bonds with other people enables us to create the uniquely complex and developed cultural niche that we live in. These products of human cooperation are also unique in that they can be preserved over time by the help of a specially evolved communicational system that allows for the efficient transmission of knowledge from one generation to the other [27–28]. As a result, each human culture is characterized by a vast body of shared knowledge that contains generic information about the world (such as "snow is cold") as well as a set of mostly arbitrary cultural norms. The very nature of these cultural norms is that while they are in great part arbitrary, acquiring knowledge of the norms specific to a certain culture is crucial in managing everyday life and in interacting with others (imagine driving on the wrong side of the road!).

We propose that the relevance of representing social groups is that it enables us to find the boundaries of culturally shared knowledge. Spoken language is obviously a part of this and can serve as a perfect indication of cultural group membership from the earliest periods of our lives as even 2-day-old infants are sensitive to language [29]. However, conducting successful interactions with others not only requires a commonly spoken tongue, but also knowledge of other social norms and conventions. Therefore, we propose that the basis of social categorization may be culturally shared knowledge in general. This notion is plausible if we consider that this distinction is genuinely meaningful and helpful in guiding behavior, whereas other distinctions, such as one based on skin-color, are in themselves empty categories. That is not to say that in certain situations some features cannot correlate with each other (see [30] for how perceptual markers of ethnicity can become associated with fundamentally norm-based social groups). However, characteristics such as race are not deterministic in this respect.

To directly investigate the question whether children are indeed sensitive to the boundaries of culturally shared knowledge, we designed two experiments where we sought to test whether children form similar representations of people based on other cues of sharing knowledge (using tools in a conventional or a non-conventional way) as they do based on language. The conventionality of tool-using behavior is an adequate index of sharing cultural knowledge as artifact functions have an inherent cultural aspect. It has been proposed that the differentiation of tools was a result of recursive tool making practices. In consequence, function knowledge includes cognitively opaque properties that could only be understood through cultural practices and culturally evolved teaching situations. Csibra and Gergely argue that communication of generic knowledge was selected as a consequence of the learnability problem induced by cognitively opaque contents in tool making practices [27,28]. Our hypothesis was that distinct cues -potentially of shared knowledge- induce an organized representation of social categories. Children were first familiarized with videos depicting a model that performed either conventional or non-conventional tool-use actions. Then, we tested whether children differentially associated a foreign language (Experiment 1) or their mother tongue (Experiment 2) to the model or a stranger based on the kind of tool-use the model had previously performed. We expected to find a difference between conditions in Experiment 1, but not in Experiment 2 as we hypothesized that the strong familiarity effect of hearing their native language would overshadow other effects.

Experiment 1

Method

Participants. Thirty (13 girls; mean age: 24.37 months, SD: 2.27 months) monolingual children between the age of 20 and 28 months participated in the study, of whom 15 were assigned in the Conventional condition and 15 in the Non-conventional condition. Participants were selected from a database of volunteer families that had previously applied for participation. Children were excluded from participation if there was at least one person in their immediate family whose native tongue was not Hungarian. An additional seven children were tested, but later excluded from the sample due to experimenter error (2); the eye-tracker did not return any data for the given measure and behavior could not be coded by visual observation (2), children could not be calibrated properly (2) and child's parents turned out to be of two different cultures (1).

Ethic statement. The experiment was conducted with the approval of the Ethical Committee of the Faculty of Education and Psychology, Eötvös Loránd University. Parents of the children signed informed consent prior to participation.

Equipment. For video stimuli presentation and data collection a Tobii T60XL eye-tracker was used with the TobiiStudio 3.2 software. The screen's size was 52×32 cm and 1920×1200 pixels. We used a five point calibration throughout the experiment. Children who were included in the final sample provided at least 80% valid eye-tracking data.

Video stimuli. For the familiarization phase, three videos were created for each condition (Conventional and Non-conventional). In each of the videos, a male model performed a goal-directed action with a chosen tool. The three videos depicted three different tool-using actions. The setup in the video always included two visible goals (e.g. a plate of food and messy hair) and two possible tools to bring the goals about (e.g. a fork and a brush). In all the videos the model first non-verbally, but explicitly demonstrated his goal by reaching for one of the goal-objects. Then he grabbed one of the tools, examined and then rejected it (shaking his head). After that he grabbed the other tool, nodded on examining it, and then used the tool to bring the goal about. After the goal had been attained, he expressed satisfaction by nodding at the outcome. Importantly, in both conditions all the goals and possible tools were familiar to the children, but in the Non-conventional condition the associations between tools and goals were unfamiliar (e.g. using a fork to brush his hair), while in the Conventional condition associations were always familiar (e.g. using a fork to eat food). The length of the videos varied between 16 and 20 seconds. For detailed description of the videos see the Appendix. Note that in both the Conventional and Non-conventional condition, the tool that was chosen by the model could efficiently bring about the outcome. Moreover, in both conditions, the model performed the action with equal confidence, and without communicating with the participant in any way (he did not look into camera, did not smile, wave, etc.). Thus, the familiarization events only differed in their level of conventionality.

Procedure. On arriving to the laboratory, children had some time to explore the room and get comfortable in the company of two experimenters, while parents were briefed about the experiment and signed informed consent. After this, one of the experimenters escorted the child and their parent into the testing room and seated them in a chair in front of the monitor of the eye-tracker (at a distance of approx. 60–70 cm) with the child sitting on the parent's lap. The experimenter assisted with calibration, but left the room once it was finished. Before the stimuli began, the model entered the room and started manipulating the computer without looking at or talking to the children. This element of the procedure was added in order to avoid the possibility that in-group and out-group effects were weakened by the fact that it was the in-group experimenter that directed children's attention to the stimuli. After the model had also left the room, children were presented with the three familiarization videos. Children always saw either three Conventional on three Non-conventional videos, depending on condition.

The familiarization phase was immediately followed by the test in which two photographs appeared side-by-side on the screen. The pictures were 21×11 cm and were positioned on the left and the right side of the screen with a 10×11 cm line in the middle, separating them. One picture depicted the model, while the other photo was taken of another young man matched in age. After 6 seconds had elapsed the voice of a man was heard from the speakers, who spoke in Swedish for 14 seconds.

Data analysis. Four groups of area of interest were created for analyses combining two factors: who was the target person (model or the other person) and time window (before the onset of the audio stimulus and after). The cut-off point was set at eight seconds, which marks the end of the first utterance; therefore by this point children already had the opportunity to judge the familiarity of the language. The length of the *before* time window was 8 seconds, while the *after* time window lasted for 12 seconds.

To test whether children differentially associated the foreign language to the two men in the test phase based on experimental condition, we analyzed the direction of their first gaze on hearing the foreign language. This measure was introduced in order to detect the potentially organized information seeking by children: if conventionality of tool use induces social categorization based on shared knowledge, children could use this information to integrate the novel stimulus as well. Children were expected to associate foreign language use with the model only after non-conventional tool use, since both cues point to the model's possessing different cultural knowledge than the participant - thus children would look first at the model. On the other hand, in case of conventional tool use, children were hypothesized to search for the source of the foreign language utterance by looking first at the novel human face, since there was a mismatch between the social categories induced by the two different cues.

We also analyzed total visit duration in both time windows to explore any possible general preference towards either of the models. In addition we analyzed total visit durations for the familiarization videos in order to exclude the possibility that children were more attentive in the Non-conventional condition due to the surprising behavior exhibited by the model.

Results. An alpha level of .05 was used for all statistical tests. Sex and age were first always entered into the analyses, but were not significant in any case and were therefore removed from all models.

An independent samples T-test was performed on the percentage of looking at the three familiarization videos with condition as a between subject variable. The analyses revealed a marginal effect of condition ($t(29) = 1.99$, $p = 0.057$), showing that children attended to the videos slightly more in the Conventional condition (Mean percentage of looking in the Conventional and the Non-conventional condition: 98 and 94 percent, respectively).

Repeated measures GLM analyses on the total visit durations (factors: target person [model vs. other], condition [Conventional vs. Non-conventional]) were performed separately for the *before* and the *after* time window in the test phase, which revealed a general preference towards the model ($F(1,28) = 13.36$, $p = 0.001$) *before* the onset of the audio stimulus (Mean looking times: Conventional condition/model: 2.84 sec; Conventional condition/other: 2.051sec; Non conventional condition/model: 2.79 sec; Non-conventional condition/other: 2.23 sec). This preference disappeared in the *after* time window (Mean looking times: Conventional condition/model: 2.59 sec; Conventional condition/other: 1.99 sec; Non-conventional/model: 2.19 sec; Non-conventional/other: 2.36 sec). See also Figure 1.

Crucially, analyzing the directions of the first fixations we found a significant effect of condition in the *after* time window ($\chi^2 = 4.821$, $p = 0.028$) showing that after the onset of the foreign language text (after time window), children were more likely to look at the model in the Non-conventional condition (10 out of 15 children), whereas the majority of children fixated on the other person first in the Conventional condition (11 out of 15 children). Results are depicted in Figure 2. No significant effect of condition was found *before* the audio stimulus ($\chi^2 = 2.4$, $p = 0.12$).

Experiment 2

Experiment 2 was designed in order to test whether similar results could be obtained using the children's native language instead of a foreign language. Since first fixations can be regarded as part of an information seeking process after a certain stimulus, we expected it to be more indicative of children's cognitive processes after events that violate their expectations. Due to the fact that hearing people speak in our native tongue is such a strong part of our every-day experiences, we hypothesized that it would not elicit any specific cognitive processing, leading to a random pattern of first fixations in the two conditions. However, this study provides important additional information for interpreting the above described results and exploring the validity of the hypothesis that cues of shared knowledge may play a part in forming representations of social groups.

Method

Participants. Twenty-six (10 girls, mean age: 24.32 month, SD: 2.28) monolingual children participated in the study with equal number of children assigned in the two conditions (Conventional and Non-conventional). The criterion for participation was the same as in Experiment 1. An additional three children were excluded from the sample due to inattentiveness.

Materials and procedure. The applied stimuli and the procedure were identical to the ones used in Experiment 1 with the exception that during the test phase, the Swedish audio text was replaced by a Hungarian (native) text. The length of the audio stimulus and the duration of the first utterance were matched to those in Experiment 1.

Results. An alpha level of .05 was used for all statistical tests. Age and sex were first entered into all of the models used in the analyses, but were later removed as they were not significant in any of the cases.

Pair-wise analysis of the percentages of the aggregated looking times during the three familiarization videos revealed a significant effect of condition, with children looking overall longer in the Conventional condition ($d(25) = 3.08$, $p = 0.009$). However, the difference between looking times was relatively small with a mean of 99 percent in the Conventional condition and 95 percent in the Non-conventional condition.

A GLM analysis on the total visit durations in the test phase yielded a significant effect of condition ($F(1, 24) = 24.34$, $p < 0.001$) and target person ($F(1, 24) = 4.34$, $p = 0.048$) in the *before* time

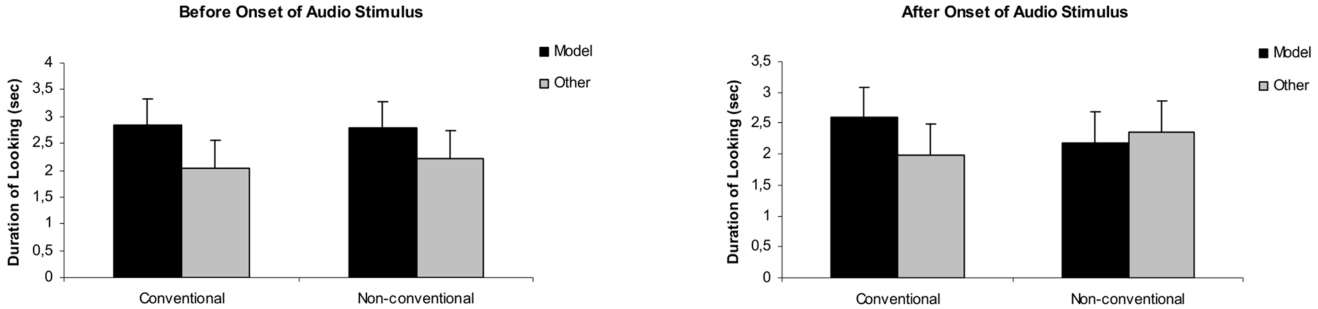


Figure 1. Total looking times in the test phase of Experiment 1. The duration of overall looking times at the two photographs (depicting the model and the other person) in the Conventional and Non-conventional condition, presented separately for the periods *before* and *after* the onset of the foreign language stimulus.
doi:10.1371/journal.pone.0101680.g001

window. Children spent more time looking at the model than at the other person and they looked longer in the Non-conventional than the Conventional condition (Conventional/model: 2.82 sec; Conventional/other: 1.99 sec; Non-conventional/model: 3.55 sec; Non-conventional/other: 2.53 sec). No effects were found in the *after* time-window (Conventional condition/model: 2.69 sec; Conventional condition/other: 2.21 sec; Non-conventional condition/model2.65 sec; Non-conventional condition/other: 2.47 sec, see also Figure 3).

Analyzing the direction of the first fixations we found an effect of the experimental condition in the *before* ($\chi^2 = 3.85$, $p = 0.05$) time window showing that most of the children in the Conventional condition fixated on the model first (9 out of 13 participants) while the majority of children in the Non-conventional condition fixated on the other person first (9 out of 13 participants). However, *after* the onset of the stimulus there was no difference between conditions ($\chi^2 = 0.16$, $p = 0.69$) with approximately the same number of children looking at the model first in both conditions (7 and 8 in the Conventional and Non-conventional condition, respectively). Results are depicted in Figure 4.

Discussion

In this experiment we investigated whether 2-year-old children form similar representations of a person based on the observed level of conventionality in their tool using habits as they do based on the language they speak. This design allowed us to test whether children would associate a foreign and a native language differentially to people based on the conventionality of his behavior. We found that children associated a foreign language to the model if he had previously performed goal-directed actions in a non-conventional way, but formed an association between the foreign language and the other person if previously the model had been seen to act in a conventional way, making it unlikely that he was the source of the foreign language utterance. On the other hand, we found no evidence of children differentially associating a native language to the two men. The latter result indicates that different characteristics along which we form judgments about a person (or possibly about social group membership) may be organized hierarchically, and language represents a stronger cue than certain other qualities (in this case tool-use, but see [21]). Children meet a vast number of people every day that differ along countless traits but with a few exceptions they share a commonly spoken tongue. Therefore hearing a text in their native language will come as no surprise to them and will probably not elicit such a

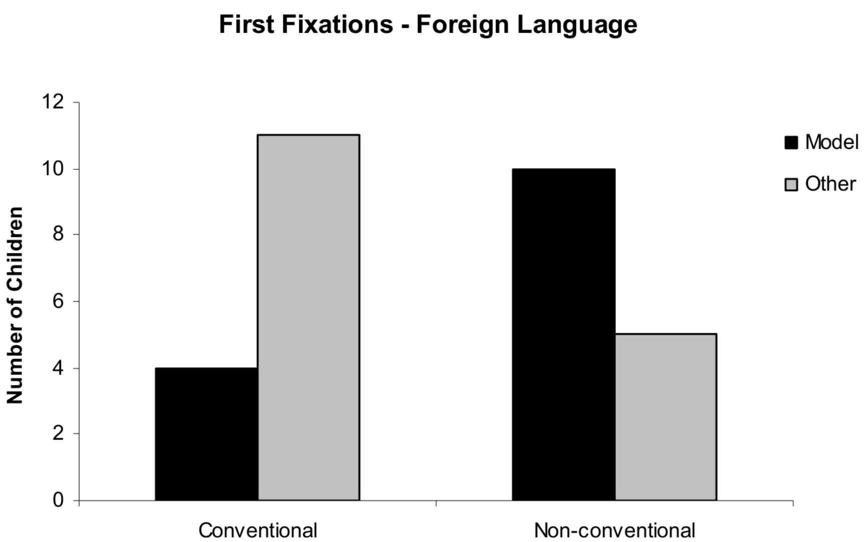


Figure 2. Distribution of first fixations after the onset of the audio stimulus in Experiment 1. Number of children looking first at the model and at the other person in the Conventional and the Non-conventional condition after hearing the foreign language text.
doi:10.1371/journal.pone.0101680.g002

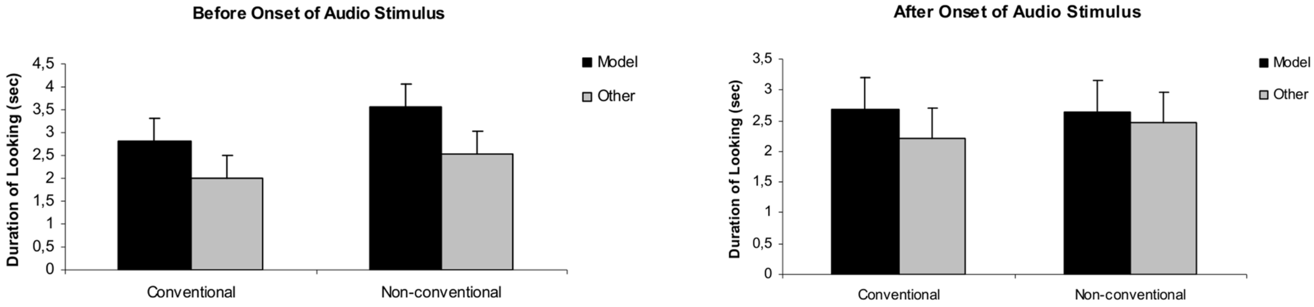


Figure 3. Looking times in the test phase of Experiment 2. The duration of overall looking times at the two photographs (depicting the model and the other person) in the Conventional and Non-conventional condition, presented separately for the periods *before* and *after* the onset of the native language stimulus.
doi:10.1371/journal.pone.0101680.g003

strong response from children. Supposedly this is reflected in the fact that we found no differences in Experiment 2.

First fixations were chosen as the subject of analyses as an indication of participants' expectation about who is more likely to be the source of the foreign language. Since children were presented with photographs, the question of to whom the voice belongs is rather ambiguous, which becomes evident to participants once they start to visually explore the static stimuli. However, the direction of the first fixation provides information about children's expectations before the realization that the visual stimuli will not help clear the ambiguity.

Analyzing the total visit durations in the test phase, we found a preference towards the model before the onset of the audio stimulus. This suggests that familiarity did have an effect in the test phase, but preferences based on mere familiarity faded away by the time the audio stimulus started and were also suppressed by the conflicting information that was provided about the model's behavioral habits (this is reflected in the between-conditions difference in first fixations in Experiment 1 and the even distribution of first fixations in Experiment 2).

Our results suggest that children take the familiarity and the conventionality of performed actions into account in forming representations about a person, and they generalize to other qualities (in this case, language) of the person based on this information. The phenomenon that even young children organize information about people systematically has been recently shown in a study where 6-month-old infants matched a non-native language to an other-race face [31]. This study, similarly to ours analyzed the looking-time patterns with photographs and spoken texts as stimuli.

Spoken language has been shown to occupy a prominent role in children's representations of humans. Children not only prefer people belonging to the same linguistic group as themselves [18], but they extend this preference to objects associated with a linguistic in-group [20], and they selectively learn from people speaking their own language [19,32]. This study has demonstrated that these representations are not constrained to the domain of language but are possibly part of a wider module designated to reason about humans in terms of familiarity and conventionality of their behavior. These characteristics are ultimately indications of whether a person is in possession of the same cultural knowledge as oneself.

Differentiating between people based on whether they share the same cultural knowledge has great adaptive value in managing every day life. In conducting interactions with others we unconsciously rely on an immense amount of shared knowledge

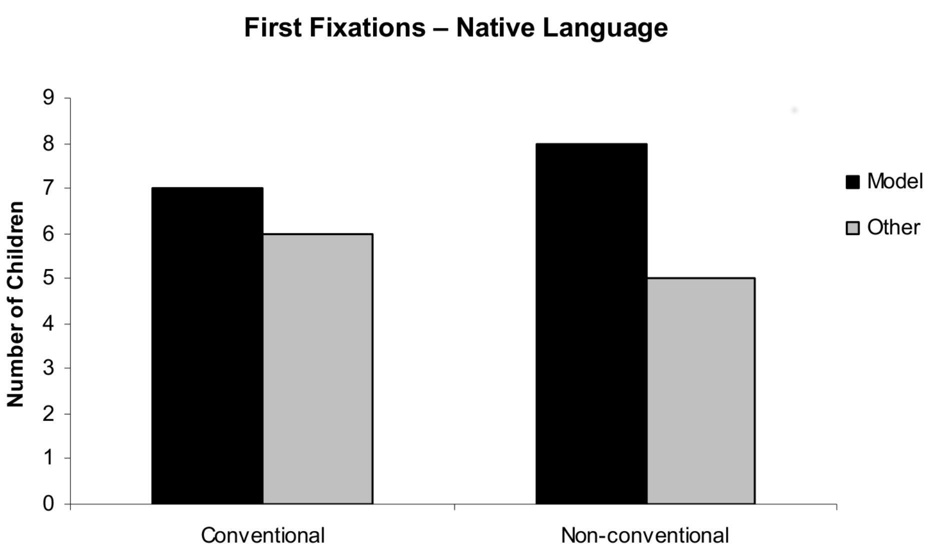


Figure 4. Distribution of first fixations after the onset of the audio stimulus in Experiment 2. Number of children looking first at the model and at the other person in the Conventional and the Non-conventional condition after hearing the native language text.
doi:10.1371/journal.pone.0101680.g004

without which our interactions would be infinitely more difficult, unsuccessful or even dangerous (think of the example of driving on the inappropriate side of the road).

From a developmental point of view, this differentiation also gains great importance as children are in the process of acquiring the necessary cultural knowledge to become a competent member of a particular society. For this, they must be able to identify reliable sources information, whose knowledge is likely to prove useful to them as well (see [20]). It is also worth noticing that children already at the age of two are extremely sensitive to violations of norms [33], which means that they are equipped with the ability to form judgments of violating or adhering to social conventions.

In sum, this study has demonstrated that young children rely on the familiarity of tool using actions in forming judgments about a person and these representations are convergent with the ones based on language use. We propose that these two characteristics are alike in that they both belong to the body of culturally accumulated and shared knowledge. This study opens the ground for further investigations aiming to test the potential role that cues of culturally shared knowledge play in representing social groups.

Appendix

Detailed description of the video stimuli

1. Food vs. Hair. The setup includes a fork and a brush (tools), a plate of food and the model's hair being messy (implication of goals). In the Conventional condition the model uses the fork to eat some potato from a plate, while in the Non-conventional condition the model uses the fork to brush his hair.

2. Liquid vs. Locket. The setup includes a key and a spoon (tools), a glass of liquid and a locket (implications of goals). In the Conventional condition the model uses the key to open the locket, while in the Non-conventional condition the model uses the key to stir the liquid.

3. Banana vs. Paper. The setup includes a knife and a pair of scissors (tools), a banana and some crepe paper (implications of goals). In the Conventional condition the model uses the scissors to cut the paper, while in the Non-conventional condition the model uses them to cut a banana.

Author Contributions

Conceived and designed the experiments: KO FE GB IK. Performed the experiments: KO FE GB. Analyzed the data: KO FE GB IK. Wrote the paper: KO FE GB IK.

References

1. Boyd R, Richerson PJ (2009) Culture and the evolution of human cooperation. Philos Trans R Soc Lond B Biol Sci 364(1533): 3281–3288.
2. Tomasello M, Carpenter M, Call J, Behne T, Moll H (2005) Understanding and sharing intentions: The origins of cultural cognition. Behav Brain Sci 28 (5): 675–735
3. Fiske ST (2000) Stereotyping, prejudice, and discrimination at the seam between the 305 centuries: Evolution, culture, mind, and brain. Eur J Soc Psychol 30(3): 299–322.
4. Bahrick LE, Netto D, Hernandez-Reif M (1998) Intermodal Perception of Adult and Child Faces and Voices by Infants. Child Dev 69(5): 1263–1275.
5. Leinbach MD, Fagot BI (1993) Categorical habituation to male and female faces: Gender schematic processing in infancy. Infant Behav Dev 16(3): 317–332.
6. Miller CL (1983) Developmental changes in male/female voice classification by infants. Infant Behav Dev 6: 313–330.
7. Poulin-Dubois D, Serbin LA, Kenyon B, Derbyshire A (1994) Infants' intermodal knowledge about gender. Dev Psychol 30: 436–442.
8. Quinn PC, Yahr J, Kuhn A, Slater AM, Pascalis O (2002) Representation of the gender of human faces by infants: A preference for female. Perception 31: 1109–1121.
9. Walker-Andrews AS, Bahrick LE, Raglioni SS, Diaz I (1991) Infants' bimodal perception of gender. Ecol Psychol 3: 55–75.
10. Baron AS, Banaji MR (2006) The Development of Implicit Attitudes Evidence of Race Evaluations From Ages 6 and 10 and Adulthood. Psychol Sci 17(1): 53–58.
11. Dunham Y, Baron AS, Banaji MR (2008) The development of implicit intergroup cognition. Trends Cogn Sci 12 (7): 248–253.
12. Hirschfeld LA. (1995) Do children have a theory of race? Cognition 54: 209–252.
13. Hirschfeld LA (1996) Race in the Making. Cambridge: MIT Press.
14. Kelly DJ, Quinn PC, Slater AM, Lee K, Gibson A, et al. (2005) Three-month-olds, but not newborns, prefer own race faces. Dev Sci 8(6): 31–36.
15. Tajfel H, Billig MG, Bundy RP, Flament C (1971) Social categorization and intergroup behavior. Eur J Soc Psychol 1(2): 149–178.
16. He Z, Baillergeon R (2012) Expectations about Co-retaliation in Toddlers. Poster Presented at the BCCCD12, Budapest, Hungary, 12–14 January 2012.
17. Kinzler KD, Spelke ES (2011) Do infants show social preferences for people differing in race? Cogn 119(1), 1–9.
18. Kinzler KD, Dupoux E, Spelke ES (2007) The native language of social cognition. Proc Natl Acad Sci U S A 104(30): 12577–12580.
19. Kinzler KD, Corriveau KH, Harris PL (2011) Children's selective trust in native-accented speakers Dev Sci 14(1): 106–111.
20. Kinzler KD, Dupoux E, Spelke ES (2012) 'Native' Objects and Collaborators: Infants' Object Choices and Acts of Giving Reflect for Native over Foreign Speakers. J Cogn Dev 13(1): 67–81.
21. Kinzler KD, Shutts K, DeJesus J, Spelke ES (2009) Accent trumps race in guiding children's social preferences. Soc Cogn 27(4): 623–634.
22. Turner J, Hogg MA, Oakes PJ, Reicher SD, Wetherell MS (1987) Rediscovering the social group: A social categorization theory. Oxford: B. Blackwell.
23. Sherif M, Harvey OJ, White BJ, Hood WR, Sherif CW (1961) Intergroup cooperation and competition: The Robbers Cave experiment.
24. Sidanius J, Pratto F (1993) The inevitability of oppression and the dynamics of social dominance. In: Sniderman PM, Tetlock PE, Carmines EG, editors. Prejudice, politics, and the American dilemma. Stanford: Stanford University Press. pp. 173–211.
25. Sperber D, Hirschfeld LA (2004) The cognitive foundations of cultural stability and diversity. Trends Cogn Sci 8(1): 40–46.
26. Cosmides L, Tooby J, Kurzban R (2003) Perceptions of race. Trends Cogn Sci 7(4): 173–179.
27. Csibra G, Gergely G (2006) Social learning and social cognition: The case for pedagogy. In: Munakata Y, Johnson HM, editors. Processes of change in brain and cognitive development. Attention and performance XXI. Oxford: Oxford University Press. pp. 249–274.
28. Gergely G, Csibra G (2006) Sylvia's recipe: The role of imitation and pedagogy in the transmission of cultural knowledge. In: Levenson S, Enfield N, editors. Roots of human sociality: Culture, cognition, and human interaction. Oxford: Berg Publishers. pp. 229–255.
29. Moon C, Cooper RP, Fifer WP (1993) Two-day-olds prefer their native language. Infant Behav Dev 16(4): 495–500.
30. McElreath R, Boyd R, Richerson P (2003) Shared norms and the evolution of ethnic markers. Curr Anthropol 44(1): 122–130.
31. Uttley L, de Boisferon H, Dupierrix E, Lee K, Quinn PC, et al. (2013) Six-month-old infants match other-race faces with a non-native language. Int J Behav Dev 1–6.
32. Buttelmann D, Zmyj N, Daum M, Carpenter M (2012) Selective Imitation of In-Group Over Out-Group Members in 14-Month-Old Infants. Child Dev 84(2): 422–428.
33. Rakoczy H, Warneken F, Tomasello M (2008) The sources of normativity: Young children's awareness of the normative structure of games. Dev Psychol 44(3): 875.

frontiers in Psychology

ORIGINAL RESEARCH
published: 24 June 2016
doi: 10.3389/fpsyg.2016.00963

3-Year-Old Children Selectively Generalize Object Functions Following a Demonstration from a Linguistic In-group Member: Evidence from the Phenomenon of Scale Error

Katalin Oláh[1,2], Fruzsina Elekes[1,3], Réka Pető[1], Krisztina Peres[1] and Ildikó Király[1]*

[1] Department of Cognitive Psychology, Institute of Psychology, Eötvös Loránd University, Budapest, Hungary, [2] Psychobiology Research Group, Institute of Cognitive Neuroscience and Psychology, Centre for Natural Sciences, Hungarian Acadamy of Sciences, Budapest, Hungary, [3] Department of Cognitive Science, Central European University, Budapest, Hungary

OPEN ACCESS

Edited by:
Natasha Kirkham, Cornell University, USA

Reviewed by:
Yuyan Luo, University of Missouri, USA
Yue Yu, Rutgers University-Newark, USA

***Correspondence:**
Katalin Oláh
olah.katalin@ppk.elte.hu

Specialty section:
This article was submitted to Developmental Psychology, a section of the journal Frontiers in Psychology

Received: *15 February 2016*
Accepted: *10 June 2016*
Published: *24 June 2016*

Citation:
Oláh K, Elekes F, Pető R, Peres K and Király I (2016) 3-Year-Old Children Selectively Generalize Object Functions Following a Demonstration from a Linguistic In-group Member: Evidence from the Phenomenon of Scale Error. Front. Psychol. 7:963. doi: 10.3389/fpsyg.2016.00963

The present study investigated 3-year-old children's learning processes about object functions. We built on children's tendency to commit scale errors with tools to explore whether they would selectively endorse object functions from a linguistic in-group over an out-group model. Participants ($n = 37$) were presented with different object sets, and a model speaking either in their native or a foreign language demonstrated how to use the presented tools. In the test phase, children received the object sets with two modifications: the original tool was replaced by one that was too big to achieve the goal but was otherwise identical, and another tool was added to the set that looked different but was appropriately scaled for goal attainment. Children in the *Native* language condition were significantly more likely to commit scale errors – that is, choose the over-sized tool – than children in the *Foreign* language condition (48 vs. 30%). We propose that these results provide insight into the characteristics of human-specific learning processes by showing that children are more likely to generalize object functions to a category of artifacts following a demonstration from an in-group member.

Keywords: scale error, object function, social category, learning, language

INTRODUCTION

Differentiating between people who belong to our social group from those who do not contributes greatly to our success in social interactions. The ability to detect the boundaries of social categories is not only vital in case of intergroup conflict, but it also helps us govern our behavior in everyday situations, such as determining what language to choose as the form of communication, how to interpret the behavior of the other person, etc. For adults, the process of categorization seems effortless and inevitable when faced with social stimuli (e.g., Taylor et al., 1978). Research with pre-school children have repeatedly shown that category-based thinking, stereotyping and in-group favoritism appear quite early in development (Aboud and Skerry, 1984).

In the past decades, ample evidence has been accumulated to support the notion that the tendency to perceive the social world as made up of groups emerges in infancy, and that this capacity may constitute a special faculty of the human mind (Kinzler and Spelke, 2007). For example, infants already at three months of age seem to be able to differentiate between female and male faces (Quinn et al., 2002), and they prefer to look at faces belonging to their own racial group (Kelly et al., 2005). Importantly, despite the early emergence of the ability to perceive the differences between various social categories, results suggest that some cues of group membership take precedence over others (Kinzler et al., 2010). A growing body of evidence highlights the importance of language in the process of categorization and developing social preferences. For example, Kinzler and Spelke (2011) have shown that 10-month-old and 2.5-year-old children do not preferentially interact with a racial in-group person, but using the same paradigm, a clear preference was observed for a native speaker over a person speaking a foreign language (Kinzler et al., 2007). In another study directly comparing the relevance of race and language in children's choices of friends, 5-year-olds were found to select people who were from a different race, but spoke with their native accent over people of the same race but speaking with a foreign accent (Kinzler et al., 2009).

For children, the importance of identifying members of the same social group lies – at least partly – in the information it may provide about the knowledgeability of the individual (Kinzler et al., 2012; Oláh et al., 2014). Language and accent supposedly prove to be such reliable cues because they usually mark the boundaries of broader cultural groups; therefore people speaking the same language likely share other aspects of cultural knowledge as well. Keeping track of the knowledge state of others is a key factor behind conducting successful social interactions with others, yet it has a special significance for infants and children who are just in the process of acquiring knowledge about the world. Children will be most successful in this endeavor if they can select trustworthy and knowledgeable informants, who will provide information that is valid and useful within the given social context. Language and accent can be good indicators for children whether someone is potentially a reliable teacher for them.

So far, a handful of studies have explored the significance of linguistic group membership in infants' and children's willingness to accept information from someone. Kinzler et al. (2012) have shown that 10-month-old infants extend the preference for native speakers to the objects they interact with. When given the possibility to choose from two toys previously introduced by a native and a non-native speaker, infants reliably choose the one introduced by the native model. Similarly, Shutts et al. (2009) showed that 12-month-old infants' choices of food were influenced by the emotional reactions of linguistic in-groups, but not that of out-groups.

Moreover, Buttelmann et al. (2013) have shown that 14-month-old infants were more likely to imitate a sub-optimal means to achieve a goal following a demonstration by an in-group member than by an out-group member. However no selectivity was observed in endorsing the object preferences of different group members. A study by Howard et al. (2015) further extends our understanding of social category based learning processes by showing that 19-month-old children only took into account the group membership of the model in an imitation task when the demonstration was administered on screen, but not in the case of live modeling. Contrarily, 3-year-old children selectively imitated an in-group member regardless of the mode of the demonstration, showing that selectivity becomes stronger with age.

A relevant question concerns children's learning about object functions, since humans' habits in using artifacts have an inherently cultural aspect. While an artifact may be appropriate to bring about several different goals, a very specific function is usually assigned to them during production. Adults and older children have a strong propensity to define object categories by the intended function, known as the *design-stance* (Dennett, 1989). Casler and Kelemen (2005) have shown that the precursors of this can be found in children as young as 2 years of age. Children of this age seem to represent objects as existing for certain purposes and view this purpose as an intrinsic property of the given object (*the teleo-functional stance*) – though they cannot yet explicitly give explanations in terms of the design-stance. It follows from such a conceptualization of artifact functions that they are not strictly or exclusively determined by the physical properties of the object, but that there is a partly arbitrary or incidental element in the process of assigning functions to objects. This arbitrary component makes object functions variable across cultures. Thus, object functions constitute a part of our cultural knowledge (e.g., whether we use a fork or chopsticks for eating).

Another important quality of object functions is that they are generally causally opaque by simple observation (Csibra and Gergely, 2009), therefore novices must rely on culturally knowledgeable individuals to pass on information about the intended function. In this study, we build on the phenomenon of scale error to investigate whether children can flexibly modulate their learning processes in response to the cultural group membership of the person demonstrating the object function.

The term scale error refers to young children's tendency to disregard the actual size of the object they are interacting with when the object category is familiar to them. As a consequence, for example, they may try to slide down a miniature slide or try to squeeze themselves into a matchbox sized car (DeLoache et al., 2004). DeLoache et al. (2004) have demonstrated this phenomenon in children aged 18–30 months in a free-play setting and suggest that it may stem from an inability to integrate information from distinct processes in visual perception and from a lack of inhibitory control. Specifically, when children encounter an object that activates the representation of a kind of object, an action plan is formed based on stored knowledge of the object category. This action plan, however, does not become inhibited by size information as it would in the case of adults or older children. DeLoache et al. (2004) propose that this may be due to the lack of integration of information processed by the ventral and dorsal visual stream (Milner and Goodale, 1995) or a dissociation between action planning and control (Glover, 2004). Since the study by DeLoache et al. (2004), a number of studies have confirmed the robustness of scale errors (e.g., Rosengren et al., 2009; Ware et al., 2010).

Casler et al. (2011) have demonstrated the same phenomenon in two-year-old children with instrumental tool-use in a structured setting. In this study, children were presented with novel and familiar object sets. In the first phase of the experiment, a model demonstrated how to use the tools to achieve certain goals. Afterwards, children received the object sets with one alteration: the original tool was replaced by one that was either too big or too small to efficiently bring the goal about. Additionally, they received a novel object that was appropriate for goal attainment, but had not been presented during the demonstration. Under such circumstances, 2-year-old children committed scale errors 31% of the time. Casler et al. (2011) argue that a proneness to committing the scale error may originate from the early emerging teleo-functional stance (Casler and Kelemen, 2005), that is, to view artifacts as existing to serve certain functions. As a consequence, the function of the tool is incorporated into the representation of the object kind and when the category representation becomes active, it inevitably activates the representation of the task the object is for.

Although committing scale errors seems to be a robust phenomenon that has been demonstrated in numerous studies, the occurrence rate of it seems to vary with age. However, results from different studies do not show a clear trend of decreasing or increasing occurrence rates with age. DeLoache et al. (2004) have found that among 18–30 month-old participants, the 20.5–24 month-old group was the most prone to scale errors (with making 1.3 scale errors on 3 object sets on average). On the other hand, Ware et al. (2006) found that when testing children between the ages of 16–24, 29–32, and 35–40 months, the latter group committed the most scale errors.

In this study, we build on the assumption that scale errors occur with tools due to the fact that function constitutes an inherent part of stored knowledge about object categories. We propose that this makes the phenomenon of scale error sensitive to the context of knowledge acquisition. Research suggest that learning about object kinds happens with the help of specialized learning mechanisms that allow the observer to efficiently gain information about a category of objects from a single demonstration (e.g., Futó et al., 2010; Butler and Markman, 2012; for a general description see the Natural Pedagogy Theory, Csibra and Gergely, 2006, 2009; Gergely and Csibra, 2006). Cues, such as eye-contact, specific intonation, and addressing prompt the learner to extract generalizable knowledge from the demonstration (as opposed to episodic information), thus contributing to the generation and enrichment of knowledge stored about object kinds. However, as described in the beginning of this review, efficient learning also requires an ability to select knowledgeable teachers, who can provide valid information. Therefore, we hypothesized that if tool functions were presented by in-group models, children would be more prone to subsequently committing scale errors since the demonstrated function would be more likely to be incorporated into the representation of the object. We followed the methods of Casler et al. (2011) with the modification that the demonstrator was either presented as a speaker of children's native language or a foreign language. We involved 3-year-old children in the study, as this is the age where both the occurrence of scale errors (e.g., Ware et al., 2006) and selectivity based on the linguistic group membership of the model (Howard et al., 2015) have been robustly demonstrated.

MATERIALS AND METHODS

The study was carried out with the approval of the Research Ethics Committee of the Faculty of Education and Psychology of Eötvös Loránd University. Participants' caregivers gave written informed consent.

Participants

Participants were 37 monolingual Hungarian children (14 girls) recruited through advertisements in the local area. Their ages ranged from 30 to 40 months, with a mean of 33.31 months ($SD = 2.69$). Children were randomly assigned to either the *Native* ($n = 17$) or the *Foreign* ($n = 20$) language condition. An additional 9 children were tested but later excluded from the sample due to passivity (3), camera failure (3), experimenter error (2), and the child was bilingual (1).

Materials

The object sets used in the study were inspired by three of the object sets used in the study of Casler et al. (2011). Each set consisted of a target object and three potential tools. There was one tool used in the demonstration phase and two presented in the test phase. One tool used in the test phase was identical to the one introduced during demonstration except that it was too big to bring about the goal, whereas the other testing tool was an alternate to the originally presented one with different perceptual features but corresponding size and affordances. The first object set could be used to paint on paper and consisted of a container with blue paint mixed with water inside (11 × 10.5 × 5 cm), a small paintbrush (19 cm long with a 3.5 × 1.5 cm head), a larger paintbrush that could not fit into the container (24 cm long with a 4.5 × 1.5 cm head) and a silicone brush (19 cm long with a 2.5 × 1.3 cm head). The second object set consisted of a yellow box (25.6 × 12.5 × 9.5) with a hole (1.5 × 1.2 cm) on top and a plastic toy inside that made a whistling noise when pushed on. The small tool used in the demonstration was a yellow wooden flat stick (14.9 cm long, 1.4 cm wide), while its larger counterpart was 29.7 × 3.9 cm in size. The alternate tool was a cylindrical stick painted red (14.8 cm, 9 mm diameter). The target action entailed inserting the tool in the hole to push on the toy inside the box and to elicit the sound. The last object set constituted of a blue box (10 × 10 × 10 cm) with a transparent tube (1.9 cm diameter) attached on top and a small ball inside the tube. The originally presented small tool was a thin wooden stick (14.5 cm long) with wooden balls (1.8 cm diameter) attached on both ends. Its larger counterpart was 25 cm long, while the balls attached were 3.5 cm in diameter. The alternate tool was a stick made out of cork (13 cm long, diameter: 1.5 cm). The target action entailed pushing the ball out of the tube with the help of the tool. For the object sets (see **Figure 1**).[1]

Procedure

Experiments were conducted by 4 female experimenters of whom 2 took turns in taking the role of Experimenter 1 (E1) and 2 took turns in taking the role of Experimenter 2 (E2); however, the roles were counterbalanced across conditions, thus each experimenter participated in both the *Foreign* and the *Native* language conditions. E1 was the person greeting the participants and administering the test trials, while E2 played the role of the demonstrator. E1 always spoke in the child's native language, whereas the language E2 used depended on condition.

Upon arrival to the laboratory, children were received by E1, who invited the child to participate in a session of free play in order to familiarize children with the environment and the experimenter. When the child seemed comfortable in the setting, E1 escorted the child and the caregiver into the testing room, where children were seated on the caregiver's lap in front of a small table. E1 then told the child that she would be back in a few seconds and left. At this point, a second female experimenter (E2) entered the room and sat down at the opposite end of the table. She started the demonstration by saying three sentences either in Hungarian (participants' native language) or in German (a foreign language to the participants). The sentences were construed in a way that they did not help the interpretation of the object function demonstration, but were not completely unrelated to the context in order to avoid confusing children in the *Native* condition. The sentences could be translated into English as follows: "Where have I put my things? They must be here somewhere. Ah, there they are!". After that, she pulled out the first object set containing the target object and the small tool. She took the tool in her hand, looked at the child with a smile, named the tool by a non-word and demonstrated the action. Then, she put away the object set and performed the demonstration with the other two object sets one after the other. When the demonstration was over, E2 left the room and E1 re-entered. E1 sat down and said to the child: "Now let's play something, shall we? Let me just see what we have here!" She then pulled out the first object set from behind a panel with two alterations compared to the initial demonstration. The tool used in the first phase was replaced by its larger counterpart that was inappropriately scaled to bring about the same goal. The alternate tool was also presented this time. The two tools were placed on the two sides of the target object. Children were allowed to interact with the object set for as long as they showed interest. After that, E1 put away the object set and presented the next one with the same alterations. All children received 3 trials, one with each object set. Children received the object sets in two pre-defined orders. The order and the side of the tools in the test phase were counterbalanced across conditions.

[1]The object sets could be categorized as either completely novel to children (the yellow box with the toy inside and the blue box with the tube on top) or familiar (painting).

Irim Tentusz Bolum

FIGURE 1 | Materials. The picture depicts the target objects with the original tool and the corresponding over-sized and alternate tools presented in the test phase. The original tool was named by a non-word (see top row) in the beginning of each demonstration.

Coding

We analyzed children's choices of tools in the first 1.5 min of the interaction phase with each object set. Only children's first choices were taken into account and we coded whether it was the over-sized (committing the scale error) or the alternate tool (not committing the scale error). Children not choosing a tool during this time were considered passive on the trial (object set) and the trial was excluded from analyses (*Native*: 2 out of 51 trials; *Foreign*: 6 out of 60 trials). An independent coder blind to the research question coded 80% of the videos. Reliability between the coders was good (Cohen's kappa: 0.86).

To test whether children were equally attentive in the *Foreign* condition as in the *Native* condition, children's looking behavior during the demonstration phase was also coded using Solomon Coder (András Péter).[2] We coded the time children spent looking at each action demonstration from the moment E2 named the object she was about to use until the moment she started putting away the object set in question. We found that children in both conditions were attentive for almost the whole duration of the demonstrations (98.6% in the *Foreign* condition and 99.29% in the *Native* condition). The difference between conditions was not significant ($t(30) = 0.72$; $p = 0.48$).

RESULTS

Statistical analyses were performed with the SPSS 20 software. We used Generalized Linear Mixed Models (GLMM) with binary regression to test for differences in the occurrence of scale errors across conditions. We used this method for analyses since the dependent variable is not continuous, but is composed of three nominal values (a choice between the oversized tool and the novel tool on three trials). Therefore a GLMM is the best option as it can treat the different trials separately and thus provides a more elaborate test of the question. We used backwards modeling, where the following variables were included in the initial model,

[2]http://solomoncoder.com

but were later removed as they were not significant: sex and age of the child, the presentation order of the object sets, side of the tools used in the test, the identities of the two experimenters, object type (novel/familiar). "Participant" was added to the model as a factor and "trial" as the repeated measure. In addition to these effects, only condition as a fixed effect was included in the final model.

Condition had a significant main effect on the amount of scale errors committed by children, with more scale errors occurring in the *Native* as opposed to the *Foreign* condition (F (1, 101) = 4.024; $p = 0.048$). On average, participants in the *Native* language condition committed the scale error on 48% of the trials, whereas the rate was 30% in the *Foreign* language condition (**Figure 2**).

The same effect of condition held with a simple comparison of proportions (occurrence rate of scale errors in the two conditions) using a χ^2 test ($\chi^2(1) = 6.81$; $p = 0.009$).

DISCUSSION

Building on the phenomenon of scale error, the present study investigated whether 3-year-old children's learning processes about tool functions would be influenced by the group membership of the person introducing the objects to them. We found that children were less prone to committing scale errors if the demonstration was performed by a person speaking in a foreign language. We propose that this result does not merely inform us about a quite specific phenomenon described in the developmental literature –that is, the occurrence of scale errors – but it reflects the special characteristics of human-specific learning mechanisms. As described in the introduction, scale errors supposedly occur because children do not treat the artifacts they encounter as individual and unique objects, but form representations of object kinds, during which the function assigned to the category of the artifact becomes a core characteristic (Casler et al., 2011). It has been suggested that a specialized learning mechanism helps children to extract kind-relevant, generalizable information from a single demonstration if the interactional partner expresses their intention of passing on knowledge (Csibra and Gergely, 2009). In this study, children supposedly committed the scale error on nearly half of the trials in the *Native* condition because they regarded the initial demonstration as an instance of teaching. Thus, the above-described genericity-bias (Csibra and Gergely, 2009) led them to retrieve the acquired knowledge (the function of the tool) in the presence of another exemplar of the same category (the over-sized counterpart of the original object) and children tried to enforce that function on the given exemplar irrespective of its actual size. It has also been shown that children are more prone to committing scale errors when the objects are named during the demonstration (Hunley and Hahn, 2016). This labeling effect with scale errors originates from the phenomenon that learning of object kinds is facilitated by naming the object (e.g., Booth and Waxman, 2002). In our study, this enhanced the proneness to committing scale errors in the *Native* condition, but children were not equally willing to accept the information from the model speaking a foreign language.

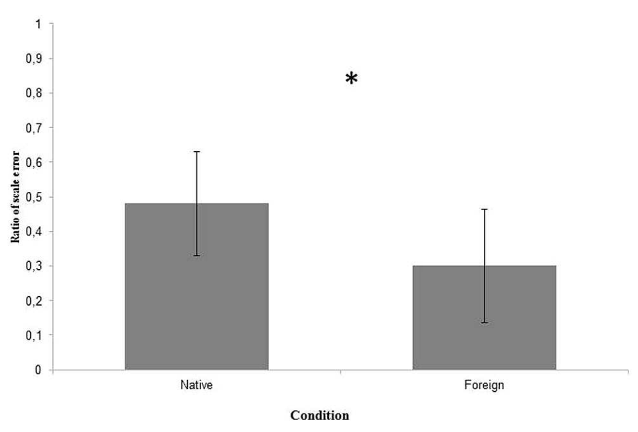


FIGURE 2 | Avarage ratio of scale errors committed in the Native and the Foreign language conditions on the three trials. The asterisk indicates that the difference is significant at the level of 0.05.

We thus propose that the decreased occurrence rate of scale errors in the *Foreign* language condition can be accounted for by the selectivity children exhibit in learning situations. That is, even though children may perceive the teaching intention exhibited by the model, a specific mistrust toward the epistemic state of the model leads them to refuse to endorse the information presented. Results suggest that even though pedagogical cues facilitate learning in general, young children are not equally willing to accept information from all sources. For example, 14-month-old infants are reluctant to imitate a model whose past behavior has turned out to be misleading (Poulin-Dubois et al., 2011) or could be seen as incompetent based on the level of confidence they exhibited (Zmyj et al., 2010). Importantly, studies have also shown that children show selectivity based on the language the potential teacher speaks (e.g., Buttelmann et al., 2013; Howard et al., 2015). Language cues can be of special importance when acquiring culturally relevant knowledge (such as tool functions) since the use of a foreign language is an indication that the person is not familiar with the ways of the given culture. Consequently children may not view them as a reliable source of information. We propose that the drop in the occurrence of scale errors reflect children's resistance to accept the foreign language speaking model as a teacher and therefore they did not extract kind-based knowledge about the objects, which subsequently led to less confusion in the test phase.

Alternatively, the decreased occurrence of scale errors in the *Foreign* condition may not reflect a mistrust in the model, but a failure to encode the teaching intentions of a foreign speaker. This could possibly be the result of an intuition that members of different cultural groups keep to different norms in their behavior; therefore children could exhibit more confusion when interpreting the signals of an out-group member. While this interpretation is not perfectly independent of the one outlined in the previous section, nor are the two necessarily mutually exclusive, the sensitivity to communicative cues may constitute such a fundamental and universal capacity of the human mind

(Csibra and Gergely, 2011), that it is unlikely to be disrupted in such a case.

An alternative explanation for our result could be that children simply paid less attention to the foreign language model and that is why scale errors occurred with less frequency. However, this explanation is not likely, as children were equally attentive during the demonstration regardless of condition and they seemed to understand the basic structures of the different tasks (they attempted to achieve the goal that was demonstrated). If children had been simply inattentive in the *Foreign* condition, then we would have expected to see instances where children were simply lost at how to interact with the novel objects. However, this was not the case; children reached for one of the tools on almost all trials in both conditions (see the section on *Coding*).

Altogether, our participants committed more scale errors than children in the study of Casler et al. (2011). While the ratio of scale errors in their study is comparable to that found in the Foreign language condition (around 30%), this number was substantially higher in the Native language condition (48%). This may be accounted for by the fact that we did not use the exact same object sets as did Casler et al. (2011). Instead of using four object sets, we settled on replicas of three of the ones used in their study. Importantly, Casler et al. (2011) have found that children committed the most scale errors when presented with novel apparatuses and novel tools (40%). In our study, two out of three object sets can be regarded as novel tool-novel apparatus sets, which could have resulted in higher overall ratios.

On the one hand, our study contributes to our understanding of how the group membership of the model influences children's learning processes. On the other, it provides a further piece of evidence to support the claim that scale errors cannot solely be explained in terms of problems of inhibitory control, but that they result from the way children form representations of tools and their functions (Casler et al., 2011). Specifically, children view artifacts as being for certain functions and they treat this artifact-function correspondence in quite a rigid way. Furthermore, our study suggests that scale errors are at least partly the result of the characteristics of human-specific learning processes that result in viewing an object as having a fixed function.

AUTHOR CONTRIBUTIONS

KO, FE, and IK conceived the experiment; KO, PR, and KP ran the experiments; KO, RP and IK analyzed tha data; KO prepared the first draft of the manuscript and FE, RP, KP, and IK provided feedback. The final version of the manuscript was submitted with the approval of all of the authors.

FUNDING

Financial support for this research was provided by the Hungarian Scientific Research Fund under grant number OTKA K-109352.

ACKNOWLEDGMENTS

We would like to thank Anna Berei for her help in conducting the experiments. We are grateful to the participating families for helping us realize this project.

REFERENCES

Aboud, F. E., and Skerry, S. A. (1984). The development of ethnic attitudes a critical review. *J. Cross Cult. Psychol.* 15, 3–34. doi: 10.1177/0022002184015001001

Booth, A. E., and Waxman, S. (2002). Object names and object functions serve as cues to categories for infants. *Dev. Psychol.* 38, 948. doi: 10.1037/0012-1649.38.6.948

Butler, L. P., and Markman, E. M. (2012). Preschoolers use intentional and pedagogical cues to guide inductive inferences and exploration. *Child Dev.* 83, 1416–1428. doi: 10.1111/j.1467-8624.2012.01775.x

Buttelmann, D., Zmyj, N., Daum, M., and Carpenter, M. (2013). Selective imitation of in-group over out-group members in 14-month-old infants. *Child Dev.* 84, 422–428. doi: 10.1111/j.1467-8624.2012.01860.x

Casler, K., Eshleman, A., Greene, K., and Terziyan, T. (2011). Children's scale errors with tools. *Develop. psychol.* 47, 857. doi: 10.1037/a0021174

Casler, K., and Kelemen, D. (2005). Young children's rapid learning about artifacts. *Develop. sci.* 8, 472–480. doi: 10.1111/j.1467-7687.2005.00438.x

Csibra, G., and Gergely, G. (2006). Social learning and social cognition: the case for pedagogy. *Process. Change Brain Cogn. Dev. Attent. Perform. XXI* 21, 249–274.

Csibra, G., and Gergely, G. (2009). Natural pedagogy. *Trends cogn. Sci.* 13, 148–153. doi: 10.1016/j.tics.2009.01.005

Csibra, G., and Gergely, G. (2011). Natural pedagogy as evolutionary adaptation. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 366, 1149–1157. doi: 10.1098/rstb.2010.0319

DeLoache, J. S., Uttal, D. H., and Rosengren, K. S. (2004). Scale errors offer evidence for a perception-action dissociation early in life. *Science* 304, 1027–1029. doi: 10.1126/science.1093567

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

Futó, J., Téglás, E., Csibra, G., and Gergely, G. (2010). Communicative function demonstration induces kind-based artifact representation in preverbal infants. *Cognition* 117, 1–8. doi: 10.1016/j.cognition.2010.06.003

Gergely, G., and Csibra, G. (2006). "Sylvia's recipe: The role of imitation and pedagogy in the transmission of cultural knowledge," in *Roots of Human Sociality: Culture, Cognition, and Human Interaction*, eds N. J. Enfield and S. C. Levenson (Oxford: Berg), 229–255.

Glover, S. (2004). Separate visual representations in the planning and control in action. *Behav. brain sci.* 27, 3-24. doi: 10.1017/S0140525X04000020

Howard, L. H., Henderson, A. M., Carrazza, C., and Woodward, A. L. (2015). Infants' and young children's imitation of linguistic in-group and out-group informants. *Child dev.* 86, 259–275. doi: 10.1111/cdev.12299

Hunley, S. B., and Hahn, E. R. (2016). Labels affect preschoolers' tool-based scale errors. *J. Exp. Child Psychol.* doi: 10.1016/j.jecp.2016.01.007 [Epub ahead of print].

Kelly, D. J., Quinn, P. C., Slater, A. M., Lee, K., Gibson, A., Smith, M., et al. (2005). Three-month-olds, but not newborns, prefer own-race faces. *Dev. sci.* 8, F31–F36. doi: 10.1111/j.1467-7687.2005.0434a.x

Kinzler, K. D., Dupoux, E., and Spelke, E. S. (2007). The native language of social cognition. *Proc. Natl. Acad. Sci. U.S.A.* 104, 12577–12580. doi: 10.1073/pnas.0705345104

Kinzler, K. D., Dupoux, E., and Spelke, E. S. (2012). 'Native'objects and collaborators: infants' object choices and acts of giving reflect favor for native over foreign speakers. *J. Cogn. Dev.* 13, 67–81. doi: 10.1080/15248372.2011.567200

Kinzler, K. D., Shutts, K., and Correll, J. (2010). Priorities in social categories. *Eur. J. Soc. Psychol.* 40, 581–592. doi: 10.1002/ejsp.739

Kinzler, K. D., Shutts, K., DeJesus, J., and Spelke, E. S. (2009). Accent trumps race in guiding children's social preferences. *Soc. cogn.* 27, 623. doi: 10.1521/soco.2009.27.4.623

Kinzler, K. D., and Spelke, E. S. (2007). Core systems in human cognition. *Prog. Brain Res.* 164, 257–264. doi: 10.1111/j.1467-7687.2007.00569.x

Kinzler, K. D., and Spelke, E. S. (2011). Do infants show social preferences for people differing in race? *Cognition* 119, 1–9. doi: 10.1016/j.cognition.2010.10.019

Milner, A. D., and Goodale, M. A. (1995). *The Visual Brain in Action*, Vol. 27. New York: Oxford University Press.

Oláh, K., Elekes, F., Bródy, G., and Király, I. (2014). Social category formation is induced by cues of sharing knowledge in young children. *PLoS One* 9:e101680. doi: 10.1371/journal.pone.0101680

Poulin-Dubois, D., Brooker, I., and Polonia, A. (2011). Infants prefer to imitate a reliable person. *Infant Behav. Dev.* 34, 303–309. doi: 10.1016/j.infbeh.2011.01.006

Quinn, P. C., Yahr, J., Kuhn, A., Slater, A. M., and Pascalis, O. (2002). Representation of the gender of human faces by infants: A preference for female. *Perception* 31, 1109–1122. doi: 10.1068/p3331

Rosengren, K. S., Gutiérrez, I. T., Anderson, K. N., and Schein, S. S. (2009). Parental reports of children's scale errors in everyday life. *Child Dev.* 80, 1586–1591. doi: 10.1111/j.1467-8624.2009.01355.x

Shutts, K., Kinzler, K. D., McKee, C. B., and Spelke, E. S. (2009). Social information guides infants' selection of foods. *J. Cogn. Dev* 10, 1–17. doi: 10.1080/15248370902966636

Taylor, S. E., Fiske, S. T., Etcoff, N. L., and Ruderman, A. J. (1978). Categorical and contextual bases of person memory and stereotyping. *J. Pers. Soc. Psychol.* 36, 778. doi: 10.1037/0022-3514.36.7.778

Ware, E. A., Uttal, D. H., and DeLoache, J. S. (2010). Everyday scale errors. *Dev. Sci.* 13, 28–36. doi: 10.1111/j.1467-7687.2009.00853.x

Ware, E. A., Uttal, D. H., Wetter, E. K., and DeLoache, J. S. (2006). Young children make scale errors when playing with dolls. *Dev. Sci.* 9, 40–45. doi: 10.1111/j.1467-7687.2005.00461.x

Zmyj, N., Buttelmann, D., Carpenter, M., and Daum, M. M. (2010). The reliability of a model influences 14-month-olds' imitation. *J. Exp. Child Psychol.* 106, 208–220. doi: 10.1111/j.1467-8624.2012.01860.x

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Learning How to Use a Tool –mutually exclusive tool-function mappings are selectively acquired from linguistic in-group models

Réka Pető* [1,2], Fruzsina Elekes* [1,2,3], Katalin Oláh[1,2] & Ildikó Király[2,3]

[1]Doctoral School of Psychology, Eötvös Loránd University, Izabella str. 46, Budapest 1064, Hungary

[2]Institute of Psychology, Eötvös Loránd University, Izabella str. 46, Budapest 1064, Hungary

[3]Cognitive Development Center, Central European University, Október 6. str. 7, Budapest, 1051, Hungary

*R. Pető and F. Elekes contributed equally to this work.

Correspondence concerning this article should be addressed to Katalin Oláh, Insitute of Psychology, Eötvös Loránd University (Izabella Str. 46, 1064, Budapest, Hungary), e-mail: olah.katalin@ppk.elte.hu, Tel: (36-1) 461-2600/5610

Word count: 6879

Abstract

The present study investigated whether 4-year-olds used language as a cue to social group membership to infer whether the tool-use behavior of a model had to be encoded as indicative of the tool's function. We built on children's tendency to treat functions as mutually exclusive (ME), i.e. their propensity to refrain from using the same tool for more than one function. We hypothesized that children would form mutually exclusive tool-function mappings only if the source of the function information is a linguistic in-group person, as opposed to an out-group person. In Experiment 1, participants (n = 39) were presented with four tool-function pairs by a model who previously spoke either in their native or in a foreign language. During the test phase, children encountered new purposes, for what they could either use the demonstrated tools' color variant or another equally suitable, thus far unseen, alternative tool. In line with our predictions, children preferred to use the alternative tool for the new function only in the cultural in-group (native language) condition (in-group: 63.3%, out-group: 42.7%). Experiment 2a replicated the initial finding using another foreign language, whereas Experiment 2b demonstrated that the lack of ME tool choice in the out-group condition did not originate from children's failure to encode the demonstration. These findings suggest that children restrict learning artifact functions from linguistic in-group models. The mutual exclusivity principle in the domain of function learning is used more flexibly than previously proposed.

Key words: function learning; selectivity; culture; mutual exclusivity

Humans' physical environment consists, in great part, of objects that were designed, planned and manufactured by fellow humans. Thus, children are surrounded by man-made artifacts or tools from the earliest of ages. Although there are several different purposes for which any given tool can be used, all tools have functions defined by social norms, and the vast majority of them have *a single* one. Social norms *prescribe* how certain tools should be used, creating a context in which tool use behavior may be considered correct or incorrect. The great challenge in function learning is to be able to acquire information by social observation rapidly (often in the absence of comprehensive causal understanding), and at the same time, to be sensitive to cues that provide guidance regarding the social validity of the observed information. Specifically, the model's (lack of) access to the in-group's shared knowledgebase should not be ignored. The study we present sheds light on how 4-year-old children cope with this challenge.

Due to their different experiences (some of which depend on personal choices, others on socio-cultural factors), different people know different things, and are thus differently reliable sources of information. There is evidence that infants monitor various cues of individual knowledgeability and show reservations to use information conveyed by previously unreliable sources (e.g. Poulin-Dubois, Brooker, & Polinia, 2011). This is especially so if the cue to knowledgeability and the to-be-learnt information belong to the same domain: Three-year-old children prefer to learn novel object labels and functions from an informant who previously named well-known objects and stated artifact functions correctly (Birch, Vauthier, & Bloom, 2008; Brooker & Poulin-Dubois, 2013). If aided by a cue to the model's certainty in his own knowledgeability (confidence) even 14-month-old infants show selectivity in learning (Zmyj, Buttelmann, Carpenter, & Daum, 2010). Moreover, during their pre-school years (3 –to 4 years), children show evidence of understanding that an individual's expertise may be limited to certain domains and avoid over-generalizing expertise to unrelated fields of knowledge (Kushnir, Vredenburgh, & Schneider, 2013).

Although observational evidence of an individual's behavior conveys the most direct proof of knowledgeability, this route is rather slow and demanding. As social group-memberships can largely influence what kinds of knowledge certain individuals have access to, using others' social categories to create expectations about what those people know may serve as a heuristic to determine expertise, increasing the efficiency of social learning. Wood, Kendall and Flynn (2013) argue that the bias to learn from models who share the learner's features (e.g. age, sex, culture) is beneficent, as familiarity signals a shared environment, where the same behaviors are relevant. In fact, some proposed that the primary aim of social categorization is to identify the borders of shared knowledge and to guide observational learning (Oláh, Elekes, Bródy & Király, 2016; Soley & Spelke, 2016; Esseily, Somogyi & Guellai, 2016). Accordingly, there is evidence that those social categories that indicate a relevant set of shared knowledge, like the model's age (Seehagen & Herbert, 2010; VanderBorght & Jaswal, 2009; Zmyj, Daum, Prinz, Nielsen, & Aschersleben, 2011) or gender (Perloff, 1982; Taylor, 2013) but not race (Krieger, Möller, Zmyj, & Aschersleben, 2016), influence children's propensity to learn.

When it comes to learning about artifact functions, the role of a larger scale organized system, culture, needs to be considered as well[1]. Although artifacts are designed to serve a given function, all of them can be used for an array of other purposes as well. The reason why forks are only used for eating is that the given social or cultural group accepts eating to be its

[1] We define *culture* in a minimal sense, referring to those groups in which the members accumulate, share and pass on knowledge to others creating a stable, shared representational space.

function, which binds the members of the group to adhere. An artifact's function is as much the product of social consensus as the result of its physical properties. Consequently, function information is only valid within the borders of the group that shares it, creating the need for monitoring where those borders are.

From the learner's point of view, cultural differences pose a more demanding situation than ignorant models. When it comes to cultural differences, individuals with different cultural backgrounds merely possess *knowledge alternatives.* These alternatives are equally valid within the social-cultural context they emerged in, but elicit puzzlement and disapproval when applied outside the circle of people who share those traditions. Thus, it is not the model's knowledgeability *per se*, but the applicability and relevance of the knowledge he/she possesses that needs to be monitored. Features like the intentionality of the model's action (Carpenter, Akhtar, & Tomasello, 1998) or the confidence of his actions (Birch, Akmal, & Frampton, 2009) are insufficient to determine whether the conveyed information was culturally valid. Cues to cultural group membership are needed.

A possible candidate that infants may use to tap the borders of shared, cultural knowledge is language use (Wood et al., 2013). Evidence suggests that humans can differentiate between their mother tongue and a foreign language from practically their first days of life (Mehler, Jusczyk, Lambertz, Halsted, Bertoncini, & Amiel-Tison, 1988) and language is known to guide affiliative relations from infancy. From 5-months onwards infants develop a preference to look at native speakers, as well as objects associated with or attended by native speakers, leading to a general preference to interact with linguistic in-groups (Kinzler, Dupoux, & Spelke, 2007; Kinzler, Dupoux, & Spelke, 2012; Marno, Guellai, Vidal, Franzoi, Nespor, & Mehler, 2016).

If however, language signals belonging to a meaningful social group, than, as a feature of categorization, it gains predictive power as well. The fact, that someone speaks the child's native language (or on the contrary, a foreign language) can lead the child to assume that he shares (or does not share) other pieces of the in-group's knowledge as well. In line with this, there is empirical indication that at 2 years of age children assume that if others express knowledge that differs from their own they will do so coherently over more than one domain: when a model used well known artifacts in an unconventional way, for incorrect outcomes, children expected that person to be the source of a foreign language utterance too, as opposed to an unfamiliar agent (Oláh, et al., 2014). This finding specifically suggests that children make inferences from function to language knowledge. Based on this, we can expect young learners to use language as a cue to guide them towards models who possess culturally valid information about functions.

Language based selectivity in social learning emerges already during infants' second year of life. Buttelman, Zmyj, Daum and Carpenter (2013) showed that 14-month-old infants imitate the unusual manner of a goal-directed action only if the model has previously been talking in the child's mother tongue as opposed to a foreign language. The studies of Howard, Henderson, Carrazza and Woodward (2015) further point out that 19-month-old children may already be able to disregard untrustworthy (non-native) informants when those are presented via video, however, it is not until 3 years of age that children can also show selectivity in their imitative behavior when information is presented to them by live models. These studies point out that from their second year of life infants already possess a tendency to preferentially follow the behavior of a native speaking model. However, the findings do not provide strong evidence for the following two claims: i) that children truly learn about functions and ii) they do this selectively from people they perceive to belong to their own cultural group.

Regarding the first point, while the procedures of both studies (Buttelman et al., 2013; Howard et al., 2015) included objects to act upon, neither employed genuine tool use behavior, hence tool functions could not have been inferred. Furthermore, imitation paradigms do not lend strong support for function learning in general due to their inability to distinguish between learning about artifacts' likeability (preferences), their physical properties (affordances) and their culturally defined, normative *function*. Whichever of the above information the child acquires through observation would likely lead to the same behavioral outcome: imitation. Finally, as Buttelman and colleagues (2013) emphasize, the cue (language use) to group-membership may just signal similarity/dissimilarity or familiarity/unfamiliarity to the child, biasing them toward copying the native speaking model's behavior on an affiliative basis, rather than being interpreted as an indicator of cultural identity. In the following we summarize what distinctive features function representations have that can be utilized to study how social categorization shapes function acquisition, and outline our study that attempts to shed light on this process.

There are features of mature function representations that can be used as a litmus test for function learning and its cultural selectivity. According to Casler and Kelemen (2005, 2007), humans' function representations reflect the so-called teleo-functional stance: they expect tools to exist for a purpose, which purpose is a socially (culturally) defined, enduring, and core property of the tool kind and is independent of the agent using it. Casler, Eshleman, Greene, & Terziyan (2011) argue that humans' tendency to view artifacts as existing *for* certain functions is such a pervasive, fundamental feature of their thinking that knowledge about a tool's function may even override knowledge about its physical affordances, leading to so-called scale errors. The term scale error refers to children's (or even adults') tendency to use tools according to the function it was designed to serve even if the given exemplar is unsuitable to fulfill that goal due to its size (Casler et al., 2011; DeLoache, Uttal & Rosengren, 2004; Ware, Uttal & DeLoache, 2010; Casler, Hoffmann & Eshleman, 2014).

This phenomenon has been used by Oláh, Elekes, Pető, Peres and Király (2016) to test the selectivity of function learning. In their study children were presented with goal directed tool-use actions either by a native or a non-native model. At test, children had the opportunity to achieve the demonstrated goals either with a novel, but suitable tool, or a disproportionately large exemplar of the demonstration tool. Three-year-olds were more likely to create artifact representations that led to scale errors (choosing the inappropriately sized demonstration tool) when the artifact was introduced by a native model. This suggests that children interpreted the tool-use action demonstrated by a native, as opposed to a non-native speaker as indication of what the tool *had to be used for* (its function) regardless of the size of the specific exemplar.

Another feature of the teleo-functional stance is that it leads to reasoning about tool-function relations as mutually exclusive: one tool – one function only. The principle of mutual exclusivity was first described in the literature of language learning (e.g. Markman & Wachtel, 1988), where it was suggested to help learners acquire new linguistic information in sub-optimal learning contexts in which the referent of a label is unclear (Markman, Wasow, & Hansen, 2003). Applied to the domain of artifacts, the principle of mutual exclusivity means that if a tool is known to have a given function then i) the same function should always be pursued by that tool (i.e. generalization), and ii) the same tool should not be used to serve other functions (i.e. dissociation, Casler & Kelemen, 2007). While the generalization component is present from 24 months (Casler & Kelemen, 2007), the dissociation component develops gradually between 2 –to 4 years of age. In an experiment by Casler and Kelemen (2005), pre-schoolers (31, 43, and 51 months olds) were shown two novel, perceptually distinct tools that were equally suitable to bring about two goals, turning on a light and crushing cookies. The experimenter highlighted their similarities and then used one of them to turn on the light, allowing the child to copy the action too. Participants were tested on the same day and a few days later and they had to indicate which tool they would use to bring about the two goals. Only the oldest (4-year-old) group showed genuine mutual exclusivity: they chose the alternative tool for the new purpose and the demonstrated tool for the demonstrated purpose both right after demonstration and after a delay.

The question remains, do children learn functions selectively from cultural in-group models at the dawn of genuine function reasoning? The findings of Oláh et al. (2016) indicate that children create tool-function mappings selectively from observing native models: they *generalize* the tool’s use to contexts where the same function needs to be obtained. However, if children indeed learn function from culturally knowledgeable sources only, the same selectivity should also be apparent in how specialized those representations are, i.e. in the *dissociation* criterion of the mutual exclusivity principle.

To explore this question, in the current study, 4-year-old children were introduced to a model who first spoke a few sentences either in the child’s native language, or in a foreign language. Then, the model used four different tools to obtain four specific purposes on different objects one by one. After the function demonstration, children themselves had the chance to achieve *a novel goal* with the help of one of two tools: the demonstration tool’s colour variant, and an alternative, but equally affordant tool. Children’s tool-choices were coded as either reflecting mutual exclusivity (choosing the alternative tool for the novel purpose), or not (choosing the demonstration tool). Four-year-olds were chosen to be the tested age group as this is the earliest age at which children reliably show mutual exclusivity in their reasoning about artifacts (Casler & Kelemen, 2005). We expected children to encode the presented information as the *function* of the tool when it comes from an in-group model, leading them to reason about that tool-function relation in a mutually exclusive manner, and choosing the alternative tool for the novel purpose. On the other hand, children were expected not to treat the observed behavior of a linguistic out-group person as indication of culturally valid function, causing them to behave in a more flexible manner, reflecting their affordance understanding.

Experiment 1

Method

Participants. Participants were 39 4-year-old (age range: 44 – 52 months) monolingual children tested either in the Baby lab of our university or in nearby kindergartens located in the same city. Children were randomly assigned to either the Native (n=20) or the Foreign (n=19) language condition (native condition: $M_{(age)}$ = 49.1 months, *SD* = 2.32, 6 boys and 14 girls; foreign condition: $M_{(age)}$ = 48.2 months; *SD* = 2.49, 10 boys and 9 girls). One additional child (in the foreign condition) was tested but excluded from the final sample due to passivity.

Materials. The object sets used in the study were created for this experiment to avoid children's already existing knowledge. In the following the four object sets are referred to by the pseudo-word that was used in the demonstration phase to label the demonstration tool (tentusz, irim, kavu and bólum). All of the four sets contained two target objects (enabling different achievable outcomes) and three tools: the demonstration tool, the demonstration tool's colour variant, and the alternative tool. The demonstration tool and its colour variant differed only in terms of their colour and thus represented two tokens of the same tool kind, whereas the alternative tool looked different. Both tool kinds could be used successfully to reach the two goals.

During the demonstration phase, one target object and the demonstration tool was used from all four object sets (top row of Figure 1.). The other target object appeared only in the test phase accompanied by the demonstration tool's colour variant and the alternative tool (bottom row of Figure 1.), which were equally suitable for achieving the given goal. All tools were created in seizes that could comfortably be used by the child participants. For a detailed description of the four object sets, see Table 1. Figure 1 shows each set, tentusz, irim, kavu, bólum – from left to right.

Table 1.

Description of the four object sets.

		Object set			
		1.	**2.**	**3.**	**4.**
Demonst-ration phase	**Pseudo-word**	"tentusz"	"irim"	"kavu"	"bólum"
	Target object	yellow box with a hole on top, beeper inside	circular box with a hole on the side, bells hanging inside	box with hole on one side, transparent top, soft toy inside	paper wrapped plastic bottle with a lamp inside
	Demonstration tool	blunt pencil with metal whorl wrap (blue)	cardboard stick with an oval top (blue)	plastic scoop (brown-yellow)	candlestick with a plastic cap (pink)
	Action goal	make a beep	ring the bells	obtain soft toy	turn on the light
Test phase	**Target object**	blue box, marble in the transparent tube on its top	polystyrene goal fixed to a cardboard, and a ball	jar with a polystyrene figure inside	plasticine
	Colour variant tool	yellow	green	white	green
	Alternative tool	wooden stick with wooden balls on each end	T-shaped piece of cardboard	plastic spoon	wooden handle fixed on an oval sole
	Action goal	extrude the	score a goal	obtain the	flatten the

	marble	with the ball	figure	plasticine

Figure 1.

Procedure. The experiment was conducted by two female experimenters and one bilingual model. The experimenter (whose person was counterbalanced across conditions) always spoke in the participants' native language whereas the model spoke either in the native or a foreign language depending on the condition.

The experiment started with a short free-play phase in a child friendly reception room. Once the child seemed to feel comfortable, the experimenter escorted her/him and the caregiver into the test room and asked them to sit down next to each other on the chairs placed in front of a small table. The experimenter then told them that she had to leave for a few seconds but she was coming back. At this point, the model entered the room and sat down at the opposite end of the table. Before starting the demonstration she said three sentences either in the participants' native or in a foreign language (Romanian): "Oh, where have I put my things? They must be here somewhere. Oh, here they are!" The sentences did not contain any important information about the object sets, but they were contextually relevant to avoid children's confusion in the native condition. It is important to emphasize that the model spoke only before she started the demonstration. Thus, children were not distracted by the foreign language in the foreign condition, and did not receive any additional information in the native compared with the foreign condition during function demonstration.

The demonstration phase started right after the linguistic in- or out-group induction described above. The model placed the object sets on the table and performed the respective tool function demonstrations one after another. She placed the given target object in front of her, held the demonstration tool in her right hand, looked at the child, named the tool with a pseudo-word and reached the goal only once. After that she put away the apparatus and introduced the second, third and fourth one.

When the model finished all four actions she left the room and the experimenter came back. She greeted the child, sat down and said: "I am back, let's play something, shall we? Look, what I have here for you!" She then brought out the object sets one by one in the same order as it was shown by the model. First, she put the target object in the middle of the table and one tool on each side of it: the demonstration tool's color variant, and an alternative tool equally suitable for goal-attainment. As the target object was also unknown for the child, the experimenter always verbalized the actual goal. (e.g.: "Let's take the ball out of the tube with the help of one of the tools.") Children could interact with the objects as long as they wished.

The order in which the object sets were presented and the side of the tools during the test phase (familiar vs. novel) were counterbalanced across conditions.

Coding. We analyzed children's first tool choices (demonstration/alternative tool). The tool was only coded as chosen if the child attempted to produce the goal with it, but this was determined independently of the success of the tool use action. If a child seized both tools but only used one of them to act on the target object, the behavior was coded for the active tool. As all children received all of the four apparatuses, each of them could perform 4 choice actions. The percentage of choosing the demonstration tool was calculated and analyzed.

Trials where children did not choose a tool for 1 minute (1 child in the foreign condition, which means 4 trials), and those where the child seized both of the tools at the same time but then did not use any of them (1 out of 80 trials in in-group and 1 out of 76 trials in out-group condition) were excluded from the analyses. Altogether the final sample contained 79 trials in case of native and 75 trials in case of foreign condition. An independent coder blind to the conditions also coded the videos, interrater reliability was excellent (Cohen's kappa: 0.96).

We also analyzed children's attentiveness during the demonstration phase for each object set from the moment the model named the tool until she finished the performance with the given object. A second coder also coded the looking times of participants for 30% of the videos. Agreement between the two coders was 99.45 %.

Results and discussion

Statistical analyses were executed with the help of SPSS 20 Software. We used Generalized Linear Mixed Models with binary regression to check the differences considering tool choices across conditions. Condition, sex, age, experimenter and their two-way interactions were included as fixed factors and participant as a random effect. The Corrected Akaike information criterion for the -2 log pseudo likelihood was 642.914. Results show that the corrected model was significant (F (6, 131) = 2.951, p = 0.01). Moreover, as expected, condition (native or foreign language model) had a significant main effect on tool choice (F (1,131) = 5.398; p = 0.022) (See Figure 2). Children in the native condition were more likely to act in a mutually exclusive way and choose the alternative tool for achieving the new target (63.3% of all trials) than children in the foreign condition (42.7% of all trials). Furthermore it is important to emphasize that the alternative tool choice rate in the native condition was significantly higher than predicted by chance (t = 2.435; p = 0.017). However, it did not differ from chance level in the foreign condition (t = -1.275; p = 0.2).

None of the control variables had any significant main effects on children's choices; however, we found a significant interaction between age and condition, F (1,131) = 5.892; p = 0.017, showing that the effect of condition was more pronounced for older children. To illustrate this interaction, we created two age groups (below and above 48 months of age) in both the native and the foreign condition. This revealed that children showed a similar pattern of tool choice in the native condition regardless of age reflecting mutual exclusivity. In the foreign condition older children were more likely to choose the demonstration tool than younger children (younger: 47.22%, older: 38.5%).

Additional analyses on children's looking times during the demonstration show no significant difference between the conditions (Mean percentages of looking: Native: 97.1 %; Foreign: 99.43%, t (10) = -0.797; p = 0.444).

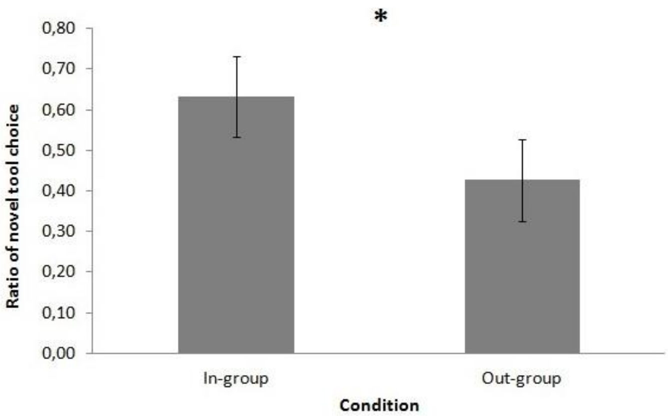


Figure 2:Avarage ration of alternative tool choices on the four test trials in the two conditions (Native vs. Foreign demonstrator). *The asterisk indicates a significant difference at the level of 0.05.*

These findings confirm our hypothesis, showing that children learn mutually exclusive functions from a linguistic in-group model only. The theory we put forward posits that the difference between in-, and out-group informants lies in infants expectation about the social, cultural validity of the information conveyed by those models. However, the random performance in the foreign condition is also compatible with the possibility that there is no learning of any kind in the foreign condition. For instance, hearing the foreign utterance may be a distraction that children cannot overcome and so fail to encode the information in the first place. The fact that children in the two conditions attended to the demonstration at equal length is not conclusive in this regard. To provide compelling evidence for our proposal we need to confirm that children in the foreign condition do not use the information as basis for creating mutually exclusive function representations *despite* being able to recall the model's actions.

Experiment 2

For this reason we conducted two control experiments (Experiment 2a and 2b). The aim of Experiment 2a was to replicate the lack of mutually exclusive tool choice (i.e. function learning) found in the foreign condition of Experiment 1 under slightly modified circumstances: the model was a different individual who spoke Russian as a foreign language. Experiment 2b aimed to find evidence that 4-year-old children can recall the information conveyed by a linguistic out-group model if they are explicitly prompted to do so. That is, the lack of ME does not originate from their failure to remember the model's actions, rather their expectation regarding the lack of validity of the information conveyed by a person who, being an out-group member, does not have access to the knowledge that the child's in-group shares.

Methods

Participants. 14-14 monolingual children took part in both Experiment 2a (M_{age} = 48.1 months, *SD* = 1.17 months, 3 boys and 11 girls) and Experiment 2b (M_{age} = 47.13 months, *SD* = 2 months, 7 boys and 7 girls). Additional 1 child was tested in Experiment 2b but excluded for refusing to engage in the task. All participants were tested at the Baby lab, parents provided informed consent.

Materials. The same object sets were used as in Experiment 1.

Procedure. The same female model served as the linguistic out-group model in both control Experiments, speaking in Russian. The experimenter was one of two female research assistants.

Experiment 2a followed the procedure of the foreign condition of Experiment 1. Experiment 2b (memory control) was entirely matched to Exp 2a in terms of the language based group-induction phase and the demonstration phase, but applied a different test procedure. Specifically, after the model left, the experimenter re-entered the room and said: “Do you remember that Olga was here just now? She played with some toys, right? Can you show me how she played?” The experimenter then placed the two goal objects belonging to the same set in front of the child at the two sides of the table, and placed the colour variant of the demonstration tool and the alternative tool into a cylindrical container in between the goal objects. If the child was reluctant to act on the objects, the experimenter emphasized that she would like to know how Olga played as she wasn’t in the room when it happened. The four object sets were introduced in the same order as in the demonstration which followed one of two predetermined orders. Both tools were affordant to achieve the outcome on both goal objects.

Coding. For Experiment 2a, the same coding scheme was used as in Experiment 1.

In Experiment 2b children could have recalled both the tool and the goal component of the function demonstration correctly or incorrectly, creating four action possibilities for each object set. Thus the probability of recalling the demonstrated function by chance was judged to be 25%. For each child we calculated the percent of *perfectly* recalled functions over the four sets. Both re-enacting the demonstrated function and pointing to the goal object and/or the tool were coded as answers. In those cases where the child explicitly stated that he/she could not remember the action or one of its components, that was coded as an incorrect answer. A second coder, blind to the hypothesis coded all videos. Interrater agreement was excellent in both control groups (Cohen's Kappa = .929, $p < .0001$, and Cohen's Kappa = .800, $p < .0001$ for Experiment 2a and 2b, respectively).

Results and discussion. Children chose the alternative tool in Experiment 2a in 46.4% of the trials, which does not significantly differ from chance, one sample t-test: t (13) = -0.806, p = .435. The findings thus replicate those of Experiment 1 with a different model, speaking a different foreign language: children did not form mutually exclusive tool-function mappings when the information came from a linguistic out-group model.

On the other hand, the rate of perfect re-enactment in Experiment 2b was 79.63%, SD = 20.31. A one sample t-test revealed that children recalled the demonstrated function above the established chance level – 25%, t (13) = 9.282, $p < .0001$. Memory performance for each object set separately was at 69.2%, 71.4%, 92.3% and 85.7% (for tentusz, irim, kavu and bólum respectively), showing that performance was high and relatively homogenous across the test objects. The findings of Experiment 2a and b thus rule out the possibility that children failed to remember the demonstration, and instead suggest that the lack of mutual exclusivity in the foreign condition came from children's refusal to treat the obtained information as socially binding, normative function.

General discussion

The aim of the current study was to expand our knowledge on pre-school children's learning strategies regarding tool functions, specifically whether they take into account the socially, culturally embedded nature of functions in their attempt to acquire relevant and valid knowledge.

The above findings provide evidence that children at the age of 4 years treat tool-function relations as mutually exclusive if information is provided by a native speaking, thereby knowledgeable source. After observing a single function demonstration of an unfamiliar tool, children in our study turned to an alternative tool when they were prompted to acquire a novel goal. That is, the function representation they created was specific enough to serve as a basis for *dissociating* non-demonstrated functions from the demonstration tool, which is the developmentally later emerging component of the mutual exclusivity assumption.

Many have shown before that brief exposure to tool function is sufficient for 2 to 3-year-olds to form enduring representations of the tool-function link, applying the same tool for the same function later on (Casler & Kelemen, 2005, 2007). The component still missing at that age is the expectation that the given tool needs to serve that particular function *only*. The earliest age in which the latter could be tapped relatively consistently is at 4-years, although the effect seems to be fragile even then (Casler & Kelemen, 2005, see the contrasting findings between Study 1 and 2). Furthermore, in previous tests participants were provided with rich informational context during demonstration, including both first person action experience and verbal description of the tools. We provide evidence for the existence of the mutual exclusivity assumption in the domain of functions at the age of 4 years by a study where the demonstration was purely observational and did not include verbal information about the tools apart from the single label provided for them.

Importantly, the data also point out that 4-year-olds restrict learning about functions to individuals, who speak their native language. In Experiment 1, children who observed a foreign speaker use the demonstration tool chose randomly between the demonstration tool and an alternative tool when it was their turn to reach a novel goal. The effect of condition was more pronounced in older children. While the pattern of tool choice for the native condition was similar in both age groups, participants in the higher range of the sample were more likely to choose the demonstration tool than children in the lower range. Experiment 2a replicated the foreign condition's results with a separate group of 4-year-olds and a different language as foreign language, while Experiment 2b demonstrated that the lack of mutually exclusive tool choice in the foreign conditions did not originate from children's failure to encode the presented information.

We propose that the difference between the native and foreign conditions stems from young learners' propensity to consider the reliability of various sources of information and adjust their learning strategy accordingly. Tool-use actions can be informative about many things: the observer can acquire information about the individual (how she likes to do things), about the action possibilities that the given object affords (what can be done if need be), as well as the culturally agreed, normative function of the artifact. Which of these should be learnt depends on context and our results provide evidence that at 4 years of age children can in fact adjust their learning strategy to the context. Function learning is restricted to situations in which the model's features indicate access to the shared knowledge of the child's social group. Crucially, it is not that children failed to encode and learn the information presented by an out-group: they could re-enact it faithfully when they were asked to do so. Rather, the interpretation and consequent use of the information differed: the actions of an out-group informant were not used to create mutually exclusive tool-function mappings.

There may be more than one explanation why specifically four-year olds do not form mutually exclusive tool-function mappings based on a linguistic out-group model's actions. For instance, it is possible that when observing out-group tool-use behavior children suspend the mutual exclusivity expectation, allowing the possibility that out-groups design tools with multiple functions. If so, the lack of mutual exclusivity should not only emerge in the child's own behavior, but also in her expectations about *fellow out-group members'* behavior. This prediction is yet to be verified.

Alternatively, children could also maintain the assumption that if a novel tool is competently used by an out-group person than that is evidence that the tool *belongs exclusively* to the out-group, and should not be utilized by the in-group. The basis for such an assumption may come from the fact that while different groups often create different tools to attain similar functions by (e.g. using forks or chopsticks to eat), it happens very rarely that different groups come up with *the same tool* as a solution for different functions. Although theoretically viable, overall the data do not support this possibility. The above strategy would lead children to refrain from using the modeled tool in the foreign condition entirely, leading them to choose the alternative tool instead of the random tool-choice apparent in our data.

On the other hand, the explanation we advocate posits that children show reservations to infer normative knowledge from the actions of an out-group. This account does not presume that children have a full-fledged theory about the out-group: e.g. whether or not out-groups create their own tools, how they are constructed and whether the model's action indicated normative function *to her own group*. It merely states that information coming from an out-group cannot be indicative of knowledge that, at least partially, depends on the social consensus of the in-group, like functions. Future research will have to clarify whether children expect that, similarly to their own group, other groups have a shared set of knowledge despite that the child herself does not have access to that knowledgebase.

Note that the effect we report cannot be explained by a native speaker's differential ability to induce a *preference* towards the model herself or the demonstration tool (Buttelmann et al., 2013). It has been shown that infants from as young as 9-months of age show preference for native speakers and items associated with a native speaker (Kinzler et al., 2007; Kinzler et al., 2012). If, in the current experiment, children of the native condition encoded the demonstration tool as generally more preferable than those in the foreign condition, this should have led them to choose the demonstration tool in the test phase more often than chance, and more often than children in the foreign condition. However, our data shows a *lower* rate of choosing the demonstration tool in the native condition compared to both the foreign condition and chance level. Hence, the difference between conditions cannot be explained by a tendency to view objects associated with a native speaker as more preferable.

As the only cue to group membership in our studies was language use, the question arises what exactly does language signal to the child? The fact of sharing knowledge is one of the key factors that define a group (Barth, 2002), and language specifically enables the transmission of that information, thereby signaling the borders of potentially meaningful social groups. We argue that if a person is perceived to belong to the child's in-group based on a linguistic cue, the child may safely enough expect that other pieces of that person's knowledge are also shared within the relevant group and should be learnt. This may not be a valid inference at all times, but as most heuristics, provides the basis for accurate enough predictions to guide selective social learning. This process may unfold without the child deliberately reasoning about the cultural applicability of the observed information, as heuristics are not necessarily deliberate or conscious (Cimpian & Salomon, 2014).

The findings we present complement Oláh and colleagues' (2016) study reporting that 3-year-olds commit more scale errors on novel tools if the function of the tool is presented by a linguistic in-group, as opposed to an out-group, model. Taken together, these two studies indicate that by the age of 4 years, the acquisition of function learning is aided by the mutual exclusivity principle, and that both its generalization (Oláh et al., 2016) and its dissociation component (present experiment) works in a flexible way, being selectively applied to information coming from culturally knowledgeable sources.

Building on the mutual exclusivity bias rather than selective imitation carries great advantage, as it allows us to conclude that the learning difference indeed concerns *functions*, rather than affordances. While cultural consensus defines specific and exclusive relations between tools and their intended function, there is no such prescriptive relation between objects and their affordances: there is an infinite array of actions that can, physically, be conducted on any given object. Copying a demonstrated tool choice can result from treating that action as the tool's function, or merely encoding that the tool affords that particular action. Mutual exclusivity, however, signals the acquisition of normative functions.

Although the mutual exclusivity heuristic is employed by young learners regarding man-made tools and their functions, the way it is used seems to differ in substantial features from the bias demonstrated in language acquisition. While it is supposedly a pervasive and early emerging feature of language acquisition (Markman & Wachtel, 1988; Liittschwager & Markman, 1994; Markman et al., 2003), the ME assumption in the domain of functions is obtained gradually over development. Two-year-olds refrain from applying the ME principle for functions even if those functions are stated linguistically, as facts, despite that they readily form exclusive relations between tools and labels or other types of facts (Casler, 2014). The source of this difference may reside in the different underlying structure of verbal symbols and functions.

As Kushnir et al (2013) phrase it, artifacts are similar to words: "just as words have conventionally known meanings, artifacts have conventionally known functions" (p. 447). While that is undoubtedly true, there is an important difference to be noted. The relation between linguistic labels (symbols) and their referents is *solely* determined by cultural traditions. Labels are only adept to convey the meaning that a social group accepts it to convey. In contrast, while tools, too, have a cultural layer, they can not only be interpreted as cultural products, but also as physical objects. Tools are objects with physical properties that enable a wide range of actions, only one of which is the function it was designed to serve. Learning about the action possibilities of an object is a valid aim in observational learning independently of the goal to acquire the shared cultural knowledge. Consequently, while the ME principle may facilitate quick acquisition of tools' culturally defined use, it would unnecessarily restrict learning about the causal, physical properties of tools and all those actions that a tool enables or affords. It is possible that the scope of social learning (as well as the mechanisms and heuristics serving it) shifts over time. While in early childhood children are eager to learn any physically valid information about artifacts, they become more specialized for culturally shared function information around the age of 3-4 years, exploiting the ME principle only then.

To summarize, the current study shows that at four years of age, when children start to demonstrate adult like mutual exclusivity in the domain of function learning, the principle is already applied selectively. Information about a tool’s possible use gained by observing the actions of a cultural out-group person is not used to establish mutually exclusive tool-function relations. This indicates that 4-year-olds could extract knowledgeability from one culturally defined behavior (language) and apply that information when learning about another culturally established kind of knowledge (function).

Acknowledgements

This work was supported by the Hungarian Scientific Research Fund (grant number OTKA K-109352) and the Momentum Grant (grant number: LP-2017-17/2017). We thank children and their caregivers for participating in the study as well as the host kindergartens and the teachers for providing support for our research. We are grateful for Iulia Savos and Júlia Baross for their help in data collection.

References

1. Barth, F. (2002). An Anthropology of Knowledge. *Current Anthropology*, 43(1), 1-18. https://doi.org/10.1086/324131
2. Birch, S. A., Akmal, N., & Frampton, K. L. (2010). Two-year-olds are vigilant of others' non-verbal cues to credibility. *Developmental science*, 13(2), 363-369.
3. Birch, S. A., Vauthier, S. A., & Bloom, P. (2008). Three-and four-year-olds spontaneously use others' past performance to guide their learning. *Cognition*, 107(3), 1018-1034.
4. Brooker, I., & Poulin-Dubois, D. (2013). Is a bird an apple? The effect of speaker labeling accuracy on infants' word learning, imitation, and helping behaviors. *Infancy*, 18(s1), E46-E68.
5. Buttelmann, D., Zmyj, N., Daum, M., & Carpenter, M. (2013). Selective imitation of in-group over out-group members in 14-month-old infants. *Child Development*, 84(2), 422-428.
6. Carpenter, M., Akhtar, N., & Tomasello, M. (1998). Fourteen-through 18-month-old infants differentially imitate intentional and accidental actions. *Infant behavior and development*, 21(2), 315-330.
7. Casler, K. (2014). New Tool, New Function? Toddlers' Use of Mutual Exclusivity When Mapping Information to Objects. *Infancy*, 19(2), 162-178.
8. Casler, K., Eshleman, A., Greene, K., & Terziyan, T. (2011). Children's scale errors with tools. *Developmental psychology*, 47(3), 857.
9. Casler, K., Hoffman, K., & Eshleman, A. (2014). Do adults make scale errors too? How function sometimes trumps size. *Journal of Experimental Psychology: General*, 143(4), 1690.

10. Casler, K., & Kelemen, D. (2005). Young children's rapid learning about artifacts. *Developmental science*, 8(6), 472-480.

11. Casler, K., & Kelemen, D. (2007). Reasoning about artifacts at 24 months: The developing teleo-functional stance. *Cognition*, 103(1), 120-130.

12. Cimpian, A., & Salomon, E. (2014). The inherence heuristic: An intuitive means of making sense of the world, and a potential precursor to psychological essentialism. *Behavioral and Brain Sciences*, 37(5), 461-480. https://doi.org/10.1017/S0140525X13002197

13. DeLoache, J. S., Uttal, D. H., & Rosengren, K. S. (2004). Scale errors offer evidence for a perception-action dissociation early in life. *Science*, 304(5673), 1027-1029.

14. Esseily, R., Somogyi, E., & Guellai, B. (2016). The Relative Importance of Language in Guiding Social Preferences Through Development. *Frontiers in Psychology*, 7.

15. Howard, L. H., Henderson, A. M., Carrazza, C., & Woodward, A. L. (2015). Infants' and Young Children's Imitation of Linguistic In-Group and Out-Group Informants. *Child development*, 86(1), 259-275.

16. Kinzler, K. D., Dupoux, E., & Spelke, E. S. (2007). The native language of social cognition. *Proceedings of the National Academy of Sciences*, 104(30), 12577-12580.

17. Kinzler, K. D., Dupoux, E., & Spelke, E. S. (2012). 'Native'Objects and Collaborators: Infants' Object Choices and Acts of Giving Reflect Favor for Native Over Foreign Speakers. *Journal of Cognition and Development*, 13(1), 67-81.

18. Krieger, A. A., Möller, C., Zmyj, N., & Aschersleben, G. (2016). Tom Is Not More Likely to Imitate Lisa Than Ying: The Influence of a Model's Race Indicated by Physical Appearance on Children's Imitation. *Frontiers in Psychology*, 7.

19. Kushnir, T., Vredenburgh, C., & Schneider, L. A. (2013). "Who can help me fix this toy?" The distinction between causal knowledge and word knowledge guides

preschoolers' selective requests for information. *Developmental psychology*, 49(3), 446.

20. Liittschwager, J. C., & Markman, E. M. (1994). Sixteen-and 24-month-olds' use of mutual exclusivity as a default assumption in second-label learning. *Developmental Psychology*, 30(6), 955.
21. Markman, E. M., & Wachtel, G. F. (1988). Children's use of mutual exclusivity to constrain the meanings of words. *Cognitive psychology*, 20(2), 121-157.
22. Markman, E. M., Wasow, J. L., & Hansen, M. B. (2003). Use of the mutual exclusivity assumption by young word learners. *Cognitive psychology*, 47(3), 241-275.
23. Marno, H., Guellai, B., Vidal, Y., Franzoi, J., Nespor, M., & Mehler, J. (2016). Infants' selectively pay attention to the information they receive from a native speaker of their language. *Frontiers in Psychology*, 7.
24. Mehler, J., Jusczyk, P., Lambertz, G., Halsted, N., Bertoncini, J., & Amiel-Tison, C. (1988). A precursor of language acquisition in young infants. *Cognition*, 29(2), 143-178.
25. Oláh, K., Elekes, F., Bródy, G., & Király, I. (2014). Social category formation is induced by cues of sharing knowledge in young children. *PloS one,* 9(7), e101680.
26. Oláh, K., Elekes, F., Pető, R., Peres, K., & Király, I. (2016). 3-Year-Old Children Selectively Generalize Object Functions Following a Demonstration from a Linguistic In-group Member: Evidence from the Phenomenon of Scale Error. *Frontiers in Psychology*, 7.
27. Perloff, R. M. (1982). Gender constancy and same-sex imitation: A developmental study. *The Journal of Psychology*, 111(1), 81-86.

28. Poulin-Dubois, D., Brooker, I., & Polonia, A. (2011). Infants prefer to imitate a reliable person. *Infant Behavior and Development*, 34(2), 303-309.

29. Seehagen, S., & Herbert, J. S. (2010). The role of demonstrator familiarity and language cues on infant imitation from television. *Infant Behavior and Development*, 33(2), 168-175.

30. Soley, G., & Spelke, E. S. (2016). Shared cultural knowledge: Effects of music on young children's social preferences. *Cognition*, 148, 106-116.

31. Taylor, M. G. (2013). Gender influences on children's selective trust of adult testimony. *Journal of experimental child psychology*, 115(4), 672-690. https://doi.org/10.1016/j.jecp.2013.04.003

32. VanderBorght, M., & Jaswal, V. K. (2009). Who knows best? Preschoolers sometimes prefer child informants over adult informants. *Infant and child development*, 18(1), 61-71.

33. Ware, E. A., Uttal, D. H., & DeLoache, J. S. (2010). Everyday scale errors. *Developmental science*, 13(1), 28-36.

34. Wood, L.A., Kendall, R., & Flynn, E.G. (2013). Whom do children copy? Model-based biases in learning. *Developmental Review*, 33(4), 341-356. https://doi.org/10.1016/j.dr.2013.08.002

35. Zmyj, N., Buttelmann, D., Carpenter, M., & Daum, M. M. (2010). The reliability of a model influences 14-month-olds' imitation. *Journal of Experimental Child Psychology*, 106(4), 208-220.

36. Zmyj, N., Daum, M. M., Prinz, W., Nielsen, M., & Aschersleben, G. (2012). Fourteen-Month-Olds' Imitation of Differently Aged Models. *Infant and Child Development*, 21(3), 250-266.

Figure caption

Figure 1. Materials. The upper row shows the object sets presented during the demonstration phase, while the lower row depicts the corresponding sets children encountered in the test phase. Original tools were named by pseudo-words as indicated at the top.

General Discussion

The overall aim of the dissertation was to investigate the availability of the so called *generative models* in infants and young children: we used experimental methods to study whether children are equipped with models, interpretative schemas that enable them to accomplish fast information mapping and inference based learning in different domains.

The collection of studies introduced in the theses allows us to conclude that children, already in their very first year of life utilize generative models - like the well-documented example, *Teleological Stance* that help them to understand *agents* around them - to adapt to their environment and maximize predictive learning with the help of information selection (see also the compatible, recent model of the *Naïve Utility Calculus,* Jara-Ettinger et al., in press).

However, the operation of these generative models is induced by the social context. We build on the assumption that the overall purpose of development not only covers the emergence of the mature set of cognitive capacities, but also the refinement of such capacities in relation to the most typical environment it is developing in. The most peculiar context of humans is being together with conspecifics. In this context, where the presence of partners is ubiquitous, knowledge on the partners is essential. That guides behavior and that makes it possible to successfully interact with each other. Indeed, social context represents a group of individuals who are knowledgeable experts with full-blown cognitive machinery. Based on this, we postulate that the most important catalyst of development is having a motivation to learn, and to become expert in this social context.

The fulfillment of this motivation is served by the ever presence of cooperative and communicative partners, and consequently by the possibility that partners could deliver information in two ways (see also III. line of research, *Navigation in the social world: Naïve Psychology*; *Theses 7-9.*): they can boost learning about the environment as they could act as knowledgeable partners modulating and gradually refining the training of the novice. Also, they are important part of the environment, so children ought to learn about them as well. The focal question for us was whether these special circumstances provide any beneficial pre-structuring for the developing mind: what were the generative models (if any) that facilitate, even speed up the learning process, and whether these models were dependent on the availability of partners and their active contribution or not. In other words, we tried to grasp the advantages of being 'social' in the emergence of cognitive functions. In order to answer these related questions we have developed four lines of investigation.

The first line of research***, I. Social Learning,*** (Theses 1-3.) introduced *imitation* as an outstanding tool for behavior acquisition. This copying behavior occurs in most of the cases for novel behavior, previously unknown to or out of the repertoire of the observer (cf. Want & Harris, 2002; and see also Call, Carpenter & Tomasello, 2005). It is seen as a tool for learning novel behavior, that is, as a *form of social learning* that can lead to the transfer of information between individuals, and also between groups of people. This form of information transfer allows the transmission of behaviors, such as tool use, customs and rituals as well – behavioral patterns that are generally considered as cultural knowledge (Boyd & Richerson, 1985). Although imitation as a process is accepted as an important tool of information transfer in social sciences in general, the nature of this skill is still disputed - whether it is a conscious cognitive mechanism or a basic low-level mechanism (Gergely, Bekkering & Király, 2002; Paulus, Hunnius, Vissers, & Bekkering, 2011a).

In this dispute, we have provided evidence that imitation is rather an inference-based process fueled by a generative model, *the Teleological Stance*. Respectively, observers, even very young infants apply the rationality principle in the process of social learning: they take into account the efficiency of observed actions to achieve a specific goal or outcome when deciding whether to reenact a specific behavior or not. In other words, action interpretation -guided by the rationality principle- allows the selection of the to-be-learned novel action sequences (Gergely, et al, 2002). The pattern of selective imitation was also shown for complex tool-use behavior (Király, 2009: *Thesis 1*.). In the experimental condition of this study, the model first demonstrated that the hand was not efficient to attain a given goal (to take a little box out of a bigger one), and then used a novel tool (a string with a magnet) successfully. Infants inhibited their prepotent response, the hand action, when the demonstration highlighted and justified the use and efficiency of the novel means.

The ability to apply the Teleological Stance is available for low-functioning, nonverbal children with autism as well in a restricted format: nonverbal children with autism attributed goals to the observed model, but did not show a sensitivity to apply the inferences in order to appreciate the model's prior intentions even in nonverbal contexts (Somogyi et al, 2013: *Thesis 2*).

The description of rational imitation emphasized that action interpretation emerges from the observation of the overt behavior within a given context: in the head-touch procedure, when the experimenter's hands were occupied, touching the light with her head was the only available means for turning on the light, so its use was justified. Consequently, based on the observation of behavior and context, infants could have interpreted her head-touch as not essential to accomplishing her goal. Infants thus used their hands to light up the lamp. But one might raise the question of, why the majority of infants reenacted the novel head action in the other, hands free condition? Here both the model's and infant's own hands were free, so infants could have (rationally) opted for accomplishing the goal by performing the more efficient ‘‘hand action’’. A possible interpretation would be that in this situation, where the experimenter's hands were free, and she could have used her hand to turn on the light, infants may have made the inference that using her head was indeed essential to accomplishing her goal.

However, this apparently paradoxical behavior highlighted the need for a broader explanation framework that could go beyond the account of teleological stance as an interpretation framework. This account builds on infants' inclination to interpret action demonstrations as social communication. Based on this assumption, we have recognized the possibility that young children, facing ambiguity, are able to invite other generative models as well and integrate them: concretely, in the absence of the possibility to apply to the Teleological Stance, children rely on the communicative signals of experienced others.

Because of the communicative signals that accompany the action demonstration, infants construe of it as a communicative action rather than a purely instrumental action (see Csibra & Gergely, 2011; Gergely & Jacob, 2012). They take the situation as a *teaching context* where the demonstrated action and the communicative context together make them learn a functional action. Let us consider the head touch paradigm again: in the hands occupied condition, Teleological Stance is sufficient to form a coherent interpretation of the situation; however, this is not the case in the hands free condition. Thus, observers will search for alternative explanations. We suggested that the communicative nature of the demonstration would activate the expectation that there were some relevance of the specific behavior presented. Because of this, besides construing of the final goal as “lighting the lamp”, infants will also construe of a *subgoal*, which is to achieve this by making contact between the lamp and their forehead. Rationality is interpreted in terms of this subgoal: if the goal is to touch the lamp with the forehead, then the demonstrated action no longer violates the rationality principle. Since the function of this sub-goal is not apparent, but the demonstrator has expressed the intention to pass this behavior on, the action itself will be interpreted as normative (Király et al., 2013: *Thesis 3*)

Our model of inference based selective imitation highlights that infants do not automatically produce a matching motor program (with high fidelity) but they encode the goal of the situation and additionally, they retrieve a behavior that is successful to attain it. Importantly, the pedagogical stance modulates what is further selected to be learnt in the situation: communication enrich the encoding of the concrete goal indicating that the manifestation of a particular means (or features of it) points to it as a culturally relevant way of goal-attainment.

The refined model we argue for (Király et al., in prep.: *Thesis 3*) elaborates the role of inferential processes beyond action analysis, and argues that ostensive communication enables infants to represent the goal structure (overall goal and subgoal of specific means) of novel actions even when the causal relations between means and end are cognitively opaque. The *presumption of relevance* guides the interpretation of the demonstration by segmenting the communicative context and the relevant content. Still, the principle of efficiency (rationality) is employed to compute and disregard action elements of the observed demonstration that are justifiable by situation-specific physical constraints.

The second line of research, - studies on the organization and development of memory (***II. Memory,*** *Theses 4-6.*) - allowed us to accept that the generative model of Teleological Stance (introduced in I. *Social learning* studies) play a significant role also in the formation of general memories. Event elements that are selected as necessary parts of efficient goal attainment at the time of encountering are filtered out and encoded as predictive, generic action representations. This format - which is rather semantic in essence - is retained after a delay, even after a single exposure to the original event (Király, 2009: *Thesis 4*). In situations, where successful problem solving would require the recall of event details (that were suboptimal, and so irrelevant with respect to efficient goal attainment in the original context), two years olds show that they can only retrieve those elements that were part of the efficient event structure, and unable to retain the suboptimal specifics. This noticeable inflexibility of early memory formation is treated as a byproduct of being able to filter the relevant and efficient event elements, the predictive components of the ongoing event in support of developing a trustworthy knowledge base. Nonetheless, this pattern of finding provides indirect support and also explanation for the lack of episodic memory competencies in early ages. In the theoretical frame we propose, episodic memory emerges to serve the function to reorganize and update knowledge based on mnemonic retrieval (Kampis et al., 2014: *Thesis 5*; see also Király et al., in prep.: *Thesis 8*).

Still, one of the important functions of memory organization undoubtedly is to grant access to shared cultural knowledge. The evidence provided by our investigation – preschoolers outperform adults on a text memory task and act as being members of an oral culture - support the idea that the actual mnemonic capacities and capacity constraints of children reflect different (and changing) memory organization strategies. These strategies, however, are regulated by the need to maximize the information that can be accessed as shared long-term knowledge (Király et al., 2017: *Thesis 6*).

The advancement of long-term knowledge in essence corresponds to the acquisition of shared cultural knowledge that we investigated in the fourth line of studies (***IV. Cultural learning and Naïve Sociology***, *Theses 10-12.*). We proposed that understanding the behavioral norms and cultural knowledge that guide behavior and make it possible to successfully interact with social partners is rooted in the capacity of representing fellow humans as belonging to certain groups. This is a model that develops as an add-on to Natural Pedagogy. In our view, social category formation – Naïve Sociology – serves an epistemic cognitive function: fast mapping of the availability of a relatively stable, shared knowledge base**.** One of the primary functions of this capacity is assumed to be to guide the acquisition of group-level knowledge, already in early childhood.

There are results showing that social categorization does not imply an “in-group” preference per se even in infancy, but a preference for people who share traits with those in their environment (e.g. Bar-Haim et al., 2006), i.e. those people they have to be able to interact with on a daily basis. This shows that categorization in humans serves more than their need to belong and affiliate with a group, *any* group (see minimal group paradigm, Tajfel et al, 1971). Rather, social categorization is driven by humans' motivation to map what others know which, in turn, can be used to identify those people with whom they should *strive* to reach a shared set of knowledge, i.e. learn from. This notion is supported by findings showing that language is a stronger categorization cue for infants than race is (Kinzler & Spelke, 2011). The language a person speaks carries more information about his/her knowledge (representations) than race does, as language marks the borders of a stable, shared representational space - culture. These shared representations make us, individuals, a group. However, people differ in how much of the group forming shared knowledge they possess, for instance due to their young age. Part of the culturally shared knowledge may not be shared with novices yet; they should nevertheless aim to acquire all of that knowledge eventually. The challenge is to identify those informants whose behavior reliably shows the shared knowledge of the group they want to belong to.

These ideas made us suggest that the human mind from the very beginning is especially sensitive to cues that indicate whether knowledge bases match between individuals. Understanding the boundaries of social groups defined by cultural knowledge guides behavior in at least two important ways. First, it helps individual members *define the limits of the validity of certain behavior patterns and ideas* and consequently make adjust the behavior in order to interact adequately with others. Second, this process will favor *selective learning processes* that ensure that one will be in possession of relevant cultural knowledge.

Our proposition was that identifying cultural in-group members have a special significance for children, namely to guide learning processes by endorsing information only from culturally knowledgeable individuals. A behavioral cue that entails that a person shares the knowledge space of the target group/culture lead infants to categorize that person as "in-group" which, in turn, necessarily induce an *epistemic trust* towards any information that person may manifest later on, even if that is not yet part of the perceivers knowledge base.

The results of our studies provide evidence that *Naïve Sociology* as a generative model serves the function to identify the borders of shared knowledge (Oláh et al., 2014: *Thesis 10.*). When receiving a novel piece of information from a representative of shared knowledge, that piece of information is treated as part of the same representational space and selectively learnt from individuals who have been categorized as culturally knowledgeable (Oláh et al., 2017; Pető, Elekes et al., in prep.: *Theses 11-12*). Thus, Naïve Sociology contributes to the selectivity of social learning.

In the third line of research, ***III. Navigation in the social world: Naïve Psychology*** (*Theses 7-9*), we proposed that the ubiquitous nature of sociality is rooted in the capability that humans can *learn from* (see also above *Natural Pedagogy* and *Naïve Sociology*) and *about others* as well. Importantly, communication provides the context for these learning processes. Namely, the communicative partner can play two separate roles: could be the main source of information, and also the target of observation and learning. The aim of our studies was to highlight that from very early on, children are able to exploit both roles of the partner in order to enrich their knowledge on different fields of the environment in an integrated format.

Specifically, object-directed behavior could provide two types of information: they can be used communicatively as *referential symbolic devices* to convey culturally shared knowledge about referents that can be generalized to other individuals, or they can convey the expressers' *person-specific*, subjective disposition toward objects. Our studies revealed that infants mostly rely on their so-called *'object-centered'* interpretation: they form expectations that all others will perform the same kind of object-directed actions that are adequate given the objective valence quality or generic representation of the referent object they have just learnt about (Gergely et al. 2007; Kampis et al., 2013: *Theses 7-8*). Furthermore, we demonstrated that children already in their second year of life could flexibly assign either a person-centered interpretation or an object-centered interpretation to referential emotion displays (Egyed et al., 2013: *Thesis 7*).

The possibility that children were able to apply both the object centered and the person centered interpretative schemas when interpreting another person's behavior, made us reconsider the findings on theory-of mind competencies in young infants. We could reveal that there is a primacy of the object-centered approach: infants generalize the content they acquired through the perspective of a specific person to others (Kampis et al., 2013: *Thesis 8*).

Relatedly, there is an ongoing debate on the availability, and nature of early theory of mind (Naïve Psychology) competences (Perner, 1991; Rakoczy, 2012). The so called implicit theory of mind tests rely on robust behavioral measures, like looking time or anticipatory looks applied in simple object choice context. In most of the cases, these simple object choice scenarios can be interpreted by both the object-centered and by the person-centered interpretative frames as well, and these possibilities cannot be disentangled in the classic, existing approaches. However, in the framework of mindreading, we suggest a potential hierarchy between the above two interpretative schemas (generative models): the person specific interpretation could be described as an object-centered content bound to a specific person, distinctively.

Furthermore, we have investigated the functional characteristics of the mechanism that allows tracking other people's knowledge and mental states on-line. We assume that the primary role of online mindreading is to refine our expectations about the mental contents (knowledge) of the specific individual, in the immediate and ever-changing present.

The traditional view on theory of mind (ToM, Naïve Psychology) equates it with the full-blown, off-line, reasoning like process human adults (but not infants or non-human animals), possess and which capacity serves post-hoc behavior prediction and interpretation (Perner, 1991; Rakoczy, 2012). In our framework, the function of ToM goes beyond this traditional view: monitoring the partner's online access to a common representational space. More precisely, we propose that humans, once having categorized an interactional partner as in-group, maintain the baseline assumption that they share a culture-based common ground, a matching representational space (or no such shared representational space in case of out-group members, see *Theses 10-12*, and also see *Recapturing the cognitive bases of human sociality* below). This primary, rough assumption is then supplemented with the help of online ToM that detects whether representations also match regarding the immediate environment, that is, on the level of situation and person specific beliefs. Online ToM is ready to detect discrepancies in knowledge base arising from differential episodic experiences. Information exchange between interactional partners can only be successful if the borders of shared representations are mapped on both time-scales. Theory of Mind, in other words, monitors whether the partner's communicative content can be trusted (Mascaro & Sperber, 2009). This function of

ToM not only let the perceiver to track the distributed knowledge to modify upcoming interactions, but also this is the way to open a novel organizational frame for shared representations. In other words, each interaction between partners allows the perceiver to construe the shared knowledge base of different groups (even if they may be classified as “out-group”), thereby creating a more sophisticated representational space in which future interactions may be carried out. Theory of mind mechanisms will then be dynamically adjusted accordingly on an episodic level.

We have provided evidence that both adults and children are able to access the content of an interactional partner’s level-2 perspective online and without instruction to do so (Elekes, et al., 2016; 2017: *Thesis 9*). These findings suggest that having prior information about the partner’s attentional focus plays a role in this online effect, probably through narrowing down the circle of to-be-represented perspective content. This ability may contribute to humans’ capacity to detect the dynamic aspects of knowledge and pinpoint any possible *representational differences* between interaction partners who, on a larger scale (e.g. through cultural group membership) may still share a representational space.

Overall, based on the empirical evidence introduced we propose that children who are curious learners apply generative models in order to get fast access to knowledge that is useful and predictive. These generative models are triggered by knowledgeable social partners: children can observe their behavior and also can engage in interaction with them, and these social practices allow children to follow the epistemic guidance of social partners and select information given their specific behavior and perspective (see Table 1 for an overview of the function of models).

Most importantly, it seems that from the very beginning children are equipped with a tool, a generative model to understand and detect agents around them: the *Teleological Stance* make them perceive the actions - that are specific to *agents* - in terms of efficient means to bring about certain *goal states in the world*.

Additionally, children are prepared to situations, where the analysis of means end structure could fail. *Natural Pedagogy*, a special sensitivity to communicative referential behavior in them triggers the expectation of receiving *relevant* and therefore *generic information*. Such ostensive communication makes them attend to and accept information as to be learnt that is opaque for the individual learning machinery, yet constitute a stable frame for interactions over generations.

Table 1. : Overview of the function of interpretative, generative models in development

Generative model	**Description**	**Triggering cues**	**Initial task**	**Integration - extension**	**Reference to theses**
Teleological Stance- Learning from Actions	Relates three aspects of action interpretation in a systematic manner by the assumption of rationality: *goals, actions, and situational constraints.* Teleological reasoning operates with the help of two main assumptions (a) actions serve to bring about future goal states and (b) goal states are realized by the most efficient action available to the actor within the constraints of the situation.	*Agency, equifinal behavior, change of state in the context* Given information about any two of the three elements, one can infer and predict what the third element ought to be if the agent behaves in a rational manner	Understanding, representing actions as efficient means to bring about specific goal states in the world Selection and learning of predictive information	In integration with *Natural Pedagogy*: flexible social learning, selection of predictive, and generic information, even for opaque, social instrumental content: in the absence of the possibility to apply the rationality principle, children rely on the communicative signals of experienced others.	I. 1-3; II. 4, 5
Natural Pedagogy - Learning from others	Children are prepared to learn behaviors of which the function is not apparent at first glance. Communicative cues induce (a) the expectation that the information is relevant, and (b) also a readiness to learn that information	*Ostensive communicative cues,* *Direct demonstration,* *Sensitivity to and expectation of referential behavior*	Make children attend to and transmit relevant, - but for the individual learning machinery opaque –information: Learning generic knowledge		I. 3.; II.5.; III. 7.,8.; IV. 12
Naïve Psychology- Learning about others	Reasoning like process that serves (a) forward prediction of behavior based on attributed mental states and (b) interpretation of behavior with backward inferences on beliefs and desires as causes. Even in the ever changing present, this capacity allows to detect discrepancies in knowledge base arising from differential episodic experiences.	*Sensitivity to the partner's attentional focus* *Attribution of curiosity to the partner*	Forming expectations about the mental contents (knowledge) of the specific individual, the interactive partner in the ever changing present	Naïve Psychology and Naïve sociology are in continuous interplay: humans' capacity to think about others in terms of their mental states, i.e. Theory of Mind, and their tendency to form social category judgments about fellow individuals, i.e. Naïve Sociology, both serve the epistemic goal of identifying the borders of shared knowledge – setting the common ground for the present and for a long term period as well.	II. 5 III. 7-9. General Discussion
Naïve Sociology_ Learning from others as part of a group	(a) Expectation that information is especially relevant in the presence of shared knowledge cues that induce (b) a readiness to *selectively* learn that information. Belonging to certain groups or categories allows understanding the behavioral norms and knowledge that guide behavior.	*Cues that are direct indications of whether knowledge bases match between individuals*	Detection of people manifesting knowledge specific to a group; Selection of reliable sources in the sense that they will transmit information that is useful to survive and function as part of a group.		II. 6.; IV. 10-12; General Discussion

Relatedly, adapted social learning mechanisms should ensure that shared knowledge is preserved over time and that learners can assume both that the obtained knowledge is indeed shared by others and also that validity of the knowledge is nonetheless restricted. *Naïve Sociology* as a generative model contributes to the selection of reliable sources in the sense that they will transmit information that is useful to survive and function *as part of a group*.

Meanwhile, children should be able to realize that social partners are special agents around them, sharing not only the information seeking motivation but the basic cognitive mechanisms as well. With the help of *Naïve Psychology*, children very early on show the ability of forming expectations about the mental contents (knowledge) of the specific individual, the interactive partner around them – they *learn about others*.

Although we have yielded mostly indirect evidence, we presume that the integrated interplay of generative models give rise to qualitative changes with respect to performance in different cognitive domains. As an illustration, some possible changes and some ways of integration are described in the following section.

Teleological Stance and Natural Pedagogy work in tandem for the sake of helping children to become flexible social learners: in cases when the application of the rationality principle results in ambiguous outcome, children rely on the communicative signals of experienced others (*Thesis 2*). In accordance with this finding, early memory performance also can be best described as primarily collecting semantic or generic memories (*Thesis 5)*. This finding fits well into the model that young novices first need to build the bases of shared, generic knowledge. However, when children become able to detect the boundaries of shared knowledge with the help of Naïve Sociology (Thesis 10), and thus recognize the possibility of upholding different knowledge bases, they become sensitive to the fact the information is acquired from different sources that could be different in trustworthiness. We have discussed in details that partners recognized as knowledgeable are selectively trusted as teachers (*Thesis 11*). Nonetheless, children ought to become sensitive to monitor the sources of their own beliefs as well, mainly for being able to update the content of first person beliefs and also for the sake of tracking the overlaps of knowledge bases with others dynamically. Consequently, the purpose of retaining the causal history of beliefs necessitates the emergence of episodic memories (*Thesis 5 and 8*; and see also Mahr & Csibra, in press).

With the help of successful construction of episodic memories, children become able to track the changes in their own beliefs and supposedly also to detect the potential representational changes in their partners' beliefs. So, there is a suggested interdependence between episodic memory and Naïve Psychology, which opens novel perspectives with regard to the developmental trajectory of both domains. Namely, the emergence of episodic retrieval would be bootstrapped by communicative situations, especially when mindreading (Naïve Psychology) is involved; and relatedly, the emergence of episodic memory would allow the mindreading system to update previously attributed beliefs according to relevant new information (see *Thesis 8*).

In this sense, Naïve Psychology contributes to the monitoring of the history of beliefs and function for tracking the changes in knowledge bases. The idea that Naïve Sociology serves the function of mapping the borders of shared knowledge (*Thesis 10*) implies that this process has to be linked to the above mechanism used to track the epistemic states of others. Thus, Naïve Psychology and Naïve Sociology should be intertwined processes, as they both serve to gain information about the knowledge states of fellow humans. While Naïve Sociology might provide a quick access to the stable, cultural knowledge base of others, Naïve Psychology would constitute the dynamic element of the process for mental state attribution that can only be computed within the frame of the given interaction (*Thesis 9*). In other words, we propose that the above capacities are in a continuous interplay, so humans' capacity to think about others in terms of their mental states, i.e. Naïve Psychology or Theory of Mind, and their tendency to form social category judgments about fellow individuals, i.e. Naïve Sociology, both serve the epistemic goal of identifying the borders of shared knowledge. This theory would predict that mentalizing (Naïve Psychology in use) would be more intense when no information about background knowledge can be accessed – that is, for out-group members (this theory is introduced in more detail below in section *Recapturing the cognitive bases of human sociality*).

All in all, the findings introduced in the dissertation speak for the early availability of generative models that are induced by i) a general motivation to learn; and by ii) the always present and active contribution of social partners. These generative models make possible fast mapping of both the objective and the social world for novices in them.

The integrated use of the generative models give rise to elaborated, higher-order - supposedly human-specific - processes, like rich cultural knowledge, autobiographical memory and flexible mindreading. All these capacities together, in our opinion, serve the creation and maintenance of a continuously refined and changing shared representational space, the placeholder of human sociality.

Recapturing the cognitive bases of human sociality: an integrated model for future studies

Our studies so far have provided some insight how humans develop into becoming ultra-social beings: how they start to live their lives in a thick web of interactions. In line with the great socio-cognitive demand such a deeply social environment imposes on them, humans become experts at reasoning about others as intentional agents, to convey and receive messages through verbal and non-verbal communication, and to quickly form expectations about others based on their various category/group memberships. These modes of cooperative interactions are claimed to enable the accumulation of the fundamentals of culture. The idea, that the inheritance of socially set organizations shape the way interactions take place in social groups, and vice versa, made us turn to try to investigate the dynamics of human sociality. This research field has been recently introduced as an interdisciplinary target of research in order to grasp the uniqueness of social interactions in constituting the cultural (Enfield & Levinson, 2006)

Despite this initiation, thus far these skills have mostly been investigated separately, implying that they are functionally distinct capacities. The only assumed commonality was their evident role in navigating the social world. Taking it a step further, in our view, these two capacities should be studied together as we posit that Naïve Psychology and Naïve Sociology need to develop and work in tandem from the very beginning in order to allow fast learning of opaque information as shared knowledge, identify more competent social partners for the sake of effective social learning and establish common ground for ongoing and prospective interactions.

Humans exhibit a fundamental motivation to establish a shared representational space with their social partners, that is, to form a mutual understanding regarding the subset of representations (knowledge) that each of them possess: a precondition which allows them to understand others as well as to make themselves understood by their partners in collaborative situations (Tomasello, Carpenter, Call, Behne, & Moll, 2005). This motivation leads social partners to mutually track what the other knows, thinks, or believes about the world. It should be noted, however, that the resulting understanding (knowing how one's partner represents the world), even if acquired by both parties of an interaction does not necessarily imply that representation of 'shared knowledge' is achieved. Obtaining a genuinely shared representational space calls for the subsequent coordination of knowledge between individuals, the process that, in communication research, is termed 'grounding'. This phase can be realized via the medium of language (Clark and & Brennan, 1991). When communicating verbally, people aim to understand others and make themselves understood by asking for and providing feedback of each other's understanding of certain utterances, i.e. by trying to gain evidence that all members of the conversation "are on the same page". Pragmatic theories hold that this process requires reliance on mental representations of intentions that exist in the mind prior to the actual communication process (Clark, 1996). Alternative cognitive approaches conceptualize shared knowledge representations to be the by-product of dynamic and on-line cognitive processing and consider common ground to be an emergent property resulting from ordinary memory processes (Barr & Keysar, 2005). A recent model proposed by Kecskes (2008) argues that the reason why neither the pragmatic view nor the cognitive view have been able to adequately capture the real nature of representing common ground is because both approaches have ignored the fact that communication is necessarily a result of an interplay of intention and attention, where both of these processes are motivated by socio-cultural factors. In other words, it is constructed both from the knowledge that speakers derive from their prior experience (core common ground assumed as shared knowledge) and current experience (emergent common ground; emergent participant resources; see Kecskes & Zhang, 2010).

Likewise, applying the notion of the 'common ground' to a broader phenomenon, we argue that humans' motivation to acquire information about and from fellow humans (and their expectation that others would do the same) is a process that serves the formation of a similar, shared representational space, which, both depends on and provides the foundation of human sociality. This shared knowledge space provides the ground for making interpretations about the meaning of each other's actions (either verbal or behavioral) and for planning actions that will successfully lead to the intended outcome within the given social context.

The model we outline targets the first phase of this process, the features and functioning of those mechanisms that enable humans to rapidly and flexibly tap how others represent the world instead of focusing on how they ground their knowledge.

We argue that the two pillars of this capacity are Naïve Psychology or Theory of Mind (ToM) (introduced in section III.) and Naïve Sociology (introduced in section IV.), which are also in interplay with each other: the knowledge they provide about interactional partners is used to modulate our expectations about those people, which information feeds back to subsequent mentalizing efforts. Information in this system is both the output and the input of processing. To put it differently, the ultra-social nature of the human kind creates a motivation to learn about others, while the acquired knowledge enables people to be even more social by modulating the mechanism that led to the knowledge in the first place. This modulatory role is realized through influencing expectations about others' knowledge about the world.

In our view, Naïve sociology (*Theses 10-12.*) and spontaneous mindreading, Naïve Psychology (*Theses 8-9*) complement each other in guiding social interactions. Both forms of reasoning can be used to form expectations about a fellow individual. Naïve sociology creates the comprehensive background for social interactions reaching beyond the direct present, by providing cues to the presence or lack of shared, group level knowledge. Naïve Psychology is the process that fine-tunes our understanding about other humans in the social here-and-now, and thus can provide insight into the differences of knowledge or can realize emergent common ground. The standard view of human social interactions claims that success in communication, collaboration and even in social learning requires tracking of 'common ground' – the above introduced common history and shared knowledge base between social partners – argued to be realized by Naïve Psychology and memory processes (Sperber & Wilson, 1986, Tomasello & Carpenter, 2007, Barr & Keysar, 2005). Alternatively, we propose that social interactions exploit the capacity of Naive Sociology for detecting 'common ground' for sharing a common representational framework, or the lack of it on the one hand, and utilize Naïve Psychology for detecting 'distributed ground' for differentiating and organizing alternate representational spaces, or realize emergent common ground on the other hand. As such, while early available, efficient mentalizing capacities (Naïve Psychology) can be responsible for real-time tracking of the content of interactional partner's knowledge access (see Kovács et al., 2010; Elekes et al, 2016), an additional filtering system, Naïve Sociology would support the identification of representative of long term, stable shared knowledge (see also Cosmides, Tooby & Kurzban, 2003, Hi Hirschfeld, 1996, Oláh et al, 2016). Thus, these two capacities are in interplay to optimize learning from and learning about fellow humans as part of a larger social group.

As a next building block of a model on the development of cognitive capacities, we plan to show that, depending on the situational demands, the attributed more stable knowledge base influences on-line belief monitoring, and also that on-line tracking of the interactional partner's mental content can result in modifying the attributed shared representational framework. Specifically, we argue that Naïve Sociology allows the identification of a shared knowledge frame, triggering a default attributed common knowledge base in interactions. Indeed, one characteristic of humans' social categorization processes is its remarkable flexibility (differentiating social categorization from other forms of category representations). People have the ability to switch dynamically between mindsets related to different category distinctions (subcultures). Naïve Sociology eases this switching process, and as it invites epistemic trust as introduced above, it also allows the organization of knowledge into (sub)culture-relevant packages. Moreover, in specific cases, most importantly for cultural novices, it can happen that Naïve Sociology identifies that there is a different reference frame. We suppose that this invites epistemic vigilance (as generalizing one's own knowledge to the other person is not valid) and Naïve Psychology takes over the primacy, and it initiates the opening of a novel organizational frame for shared representations. We claim that this is the solution in psychological systems to dynamically handle multiple 'cultural settings' at the same time. By the simultaneous availability of the two processes, it is possible thus to both detect slight mismatches in the common representational space online and refine it (keeping the common grounds and learning for the long term as part of the framework), or to open a novel cultural package and set it as a separate representational space (learning from others and keeping it separate from own cultural knowledge).

Summary

The *Theses* were aimed to deepen our knowledge on the generative model-based nature and social embeddedness of the early available cognitive processes. In contrast to current dominant explanatory theories emphasizing that the influence of social context on cognitive development is simply rooted in humans' need to belong and affiliate with a group, we believe to provide evidence that children (and supposedly adults as well) use cues of shared knowledge to sort partners as sources of relevant generic knowledge. As such, the generative models introduced jointly serve a cognitive, epistemic function of acquiring and exchanging knowledge rapidly, and the establishment of shared, cultural representational space. Future investigations on the interplay of these generative models will allow us to understand better the core mechanisms of human social interactions, and human sociality more generally.

References

Apperly, I., A., Riggs, K., J., Simpson, A., Chiavarino, C., & Samson, D. (2006). Is belief reasoning automatic? *Psychological Science, 17,* 841–844.

Atance, C. M. & Sommerville, J. A. (2014). Assessing the role of memory in preschoolers' performance on episodic foresight tasks. *Memory, 22*(1), 118–128. DOI:10.1080/09658211.2013.820324

Barr, D.J. and Keysar, B. (2005). "Mindreading in an exotic case: The normal adult human". In Malle, B.F. and Hodges, S.D. (eds). *Other Minds: How Humans Bridge the Divide between Self and Other.* New York: Guilford Press, 271–283

Bahrick, L. E., Netto, D., Hernandez-Keif, M. (1998). Intermodal Perception of Adult and Child Faces and Voices by Infants. *Child Development*, 69: 1263-1275.

Bar-Haim, Y., Ziv, T., Lamy, D., Hodes, R.M. (2006). Nature and Nurture in Own-Race Face Processing. *Psychological Science,* vol. 17, no. 2, 159-163.

Bauer, P. J. & Leventon, J. S. (2013). Memory for one-time experiences in the second year of life: implications for the status of episodic memory. *Infancy, 18*(5), 755-781.

Bauer, P.J., Wenner, J.A., Dropik, P.L., & Wewerka, S.S. (2000). Parameters of remembering and forgetting in the transition from infancy to early childhood. *Monographs of the Society for Research in Child Developmen*t*, 65*, Serial No. 263.

Boyd R., & Richerson P. J. (1996). Why culture is common, but cultural evolution is rare. *Proceedings of the British Academy, 88*, 77–93.

Boyd, R., & Richerson, P. (1985). *Culture and the evolutionary process*. Chicago: University of Chicago Press.

Bruner, J. (2008). Culture and Mind: Their Fruitful Incommensurability. *Ethos*, Vol. 36, Issue 1., 29–45.

Bruner, J, (1996). *The Culture of Education*. Cambridge, Mass: Harvard University Press

Buttelmann, D., Zmyj, N., Daum., M., Carpenter, M. (2013). Selective imitation of ingroup over outgroup member in 14-month-old infants. *Child Development 84*, 422–428.

Butterfill, S., & Apperly, I. (2013). How to construct a minimal theory of mind. *Mind & Language, 28:* 606– 637.

Call, J., Carpenter, M., Tomasello, M. (2005). Copying results and copying actions in the process of social learning: chimpanzees (Pan troglodytes) and human children (Homo sapiens). *Animal Cognition*, 8 (3), 151–163

Caporael, L. R. (2007). Evolutionary theory for social and cultural psychology. In E. T. Higgins & A. Kruglanski (Eds.), *Social psychology: Handbook of basic principles* (pp. 3–18). New York: Guildford Press

Carey, S. (2011). The Origin of Concepts: A Precis. *Behavioral and Brain Sciences*, 34(3), 113–162. http://doi.org/10.1017/S0140525X10000919

Cimpian, A., Salomon, E. (2014). The inherence heuristic: An intuitive means of making sense of the world, and a potential precursor to psychological essentialism. *Behavioral and Brain Sciences* (2014) 37, 461–527 doi:10.1017/S0140525X13002197

Clark, A. & Karmiloff-Smith, A. (1994). The cognizer's innards: A psychological and philosophical perspective on the development of thought. Mind and Language 8 (4):487-519.

Clark, H. (1996). *Using Languages*. Cambridge: Cambridge University Press.

Clark, H. and Brennan, S. (1991). Grounding in communication. In L. Resnick, J. Levine, and S. Teasley (eds). *Perspectives on Socially Shared Cognition*.Washington, DC: American Psychological Association, 127–149.

Clayton, N.S., Bussey, T.J, & Dickinson, A. (2003). Can animals recall the past and plan for the future? *Nat Rev Neurosci.* 2003 Aug;4(8):685-91.

Cosmides, L., Tooby, J., Kurzban, R. (2003). Perceptions of race. *Trends in Cognitive Sciences*, Vol.7, Issue 4, 173-179.

Csibra, G. & Gergely, G. (2011). Natural pedagogy as evolutionary adaptation. *Philosophical Transactions of the Royal Society B, 366*, 1149-1157.

Csibra, G., Gergely, G. (2006). Social learning and social cognition: The case for pedagogy. In Munakata, Y. and Johnson, M. H. (eds.) *Processes of Change in Brain and Cognitive Development. Attention and Performance* XXI, 249-274. Oxford, Oxford University Press

Csibra. G., Bíró, S., Koós, O., & Gergely, G. (2003). One-year-old infants use teleological representations of actions productively. *Cognitive Science, 27*, 111-133.

Donald, M., (2010). The Exographic Revolution: Neuropsychological Sequelae. In: Malafouris L. & C. Renfrew (Eds.) The Cognitive Life of Things: Recasting the boundaries of the mind. (pp.71-79). Cambridge, UK: McDonald Institute Monographs.

Eacott, M.J., Crawley, R.A. (1998). The offset of childhood amnesia: memory for events that occurred before age 3. *Journal of Experimental Psychology: General*. 127(1). 22-33.

Enfield, N. J., & Levinson, S. C. (2006). Introduction: Human sociality as a new interdisciplinary field. In N. J. Enfield, & S. C. Levinson (Eds.). *Roots of human sociality: Culture, cognition and interaction* (pp. 1-35). Oxford: Berg

Ferry, A. L., Hespos, S. J., & Waxman, S. R. (2010). Categorization in 3- and 4-month-old infants: an advantage of words over tones. Child Development,81(2), 472–479.

Fivush, R. & Nelson, K. (2004). The emergence of autobiographical memory: A social cultural developmental theory. *Psychological Review, 111*(2), 486–511. DOI:10.1037/0033-295X.111.2.486

Fodor, J. (1983). *Modularity of Mind: An Essay on Faculty Psychology*. Cambridge, Mass.: MIT Press

Gergely, Gy., Nádasdy, Z., Csibra, G., Bíró, S. (1995). Taking the intentional stance at 12 months of age. *Cognition56(2),* 165-93.

Gergely, G. and Csibra, G. (2003). Teleological reasoning in infancy: The naive theory of rational action. *Trends in Cognitive Sciences*, 7 287-292.

Gergely, G., & Jacob, P. (2012). Reasoning about instrumental and communicative agency in human infancy. In: J. B. Benson, F. Xu & T, Kushnir, (Eds.), *Rational Constructivism in Cognitive Development, Advances in Child Development and Behavior*. pp. 59–94.

Gergely, G., Bekkering, H., & Király, I. (2002). Rational imitation in preverbal infants. *Nature*, 415, 755.

Gil-White, F. J. (2001) Are ethnic groups biological 'species' to the human brain?: Essentialism in our cognition of some social categories. *Current Anthropology* 42(4): 515-554

Haslam, N. (2006). Dehumanization: An integrative review. *Personality and Social Psychology Review*, 10(3), 252-264.

Hayne, H (2004). Infant memory development: Implications for childhood amnesia.

Developmental Review. 24: 33–73. doi:10.1016/j.dr.2003.09.007
Hermann, E., Call., J., Lloreda, M., Hare, B., Tomasello, M. (2007). Humans have evolved specialized skills of social cognition: the cultural intelligence hypothesis. *Science*, 317, 1360.
Hewlett, B.S., Roulette, C. J. (2016). Teaching in Hunter-Gatherer Infancy. *Royal Society Open Science*. 3: 150403. http://dx.doi.org/10.1098/rsos.150403
Heyes, C. M. (2012) Grist and mills: cultural inheritance of cultural learning. *Philosophical Transactions of the Royal Society B*, 367, 2181-2191.
Hirschfeld, L. (1996). *Race in the Making*, MIT Press
Howard, L.H., Henderson, A. M., Carrazza, C., Woodward, A.L. (2015). Infants' and young children's imitation of linguistic ingroup informants. *Child Development 86*, 259–275.
Howe M.L és Courage, M.L.(1997). The emergence and early development of autobiographical memory. *Psychological Review*. 1997 Jul;104(3):499-523.
Jara-Ettinger, J. et al. (2015a) Children's understanding of the costs and rewards underlying rational action. *Cognition 140, 14*–23.
Jara-Ettinger, J. et al. (2015b) Not so innocent: toddlers' inferences about costs and culpability. *Psychological Science* 26, 633–640
Jara-Ettinger, J., Gweon, H., Schulz, L.E., Tenenbaum, J.B. (in press): The Naïve Utility Calculus: Computational principles underlying commonsense psychology. *Trends in Cognitive Sciences*, http://dx.doi.org/10.1016/j.tics.2016.05.011
Karmiloff-Smith, A. (1995). Beyond Modularity: A Developmental Perspective on Cognitive Science. The MIT Press
Kecskes, I. (2008). Dueling contexts: A dynamic model of meaning. *Journal of Pragmatics 40*: 385–406.
Kecskes, I., Zhang, F. (2009). Activating, seeking, and creating common ground. *Pragmatics & Cognition* 17:2 . 331–355. doi 10.1075/p&c.17.2.06kec
Keil, F. C., Stein, C., Webb, L., Billings, V D & Rozenblit, L (2008). Discerning the Division of Cognitive Labor: An Emerging Understanding of How Knowledge Is Clustered in Other Minds, *Cognitive Science*, 32, 259 – 300
Kinzler, K.D, Corriveau, K.H., Harris, P.L. (2011). Childrens' selective trust in native accented speakers. *Developmental Science,* 14(1), 106-111.
Kinzler, K.D., Dupoux, E. and Spelke, E.S. (2007). The native language of social cognition. *The Proceedings of the National Academy of Sciences of the United States of America*, 104, 12577-12580.
Kinzler, K.D., Dupoux, E., Spelke, E.S. (2012). 'Native' Objects and Collaborators: Infants' Object Choices and Acts of Giving Reflect Favor for Native over Foreign Speakers. *Journal of Cognition and Development*, 13(1):1- 15.
Kinzler, K.D., Spelke, E.S. (2011). Do infants show social preferences for people differing in race? *Cognition*, 119, 1-9.
Klein, S. B. Cosmides, L., Gangi, C.E., Jackson, B., Tooby, J., Costabile, K.A. (2009). Evolution and episodic memory: An analysis and demonstration of a social function of episodic recollection. Social Cognition . 2009 April ; 27(2): 283–319.
Kovács, Á. M. (2015). Belief files in theory of mind reasoning. *Review of Philosophy and Psychology, Special Issue on Mental Files*, DOI 10.1007/s13164-015-0236-5
Kovács, Á. M., Téglás, E., & Endress, A.D. (2010). The social sense: susceptibility to others' beliefs in human infants and adults. *Science*, 330(6012), 1830-1834.

LeDoux JE (1996) *The Emotional Brain*. New York: Simon and Schuster.
Leinbach, M.D., Fagot, B.I. (1993). Categorical habituation to male and female faces: Gender schematic processing in infancy. *Infant Behavior and Development,* Volume 16, Issue 3, 317–332.
Mahr, J., Csibra, G. (2018): Why do we remember? The communicative function of episodic memory, *Behavioral and Brain Sciences,* 41 doi: 10.1017/S0140525X17000012
Mascaro, O. & Sperber, D. (2009). The moral, epistemic and mindreading components of children's vigilance towards deception. *Cognition*, *112*, 367-380
McCormack, T. & Hoerl, C. (2001). The child in time: Temporal concepts and self-consciousness in the development of episodic memory, In: C. Moore and K. Lemmon (eds.). *The Self in Time: Developmental Perspectives*, Lawrence Erlbaum, pp. 203-227.
Meltzoff, A.N (1990). Towards a developmental cognitive science: The implications of cross-modal matching and imitation for the development of representation and memory in infancy. Annals of the New York Academy of Sciences; 608:1–31.
Miller, C. L. (1983). Developmental changes in male/female voice classification by infants. *Infant Behavior and Development*, 6, 313–330.
Mullaly, S. L. & Maguire, E. A. (2014). Learning to remember: the early ontogeny of episodic memory. *Developmental Cognitive Neuroscience*, 9, 12-29.
Nadel, L. Zola-Morgan, S. (1984). Infantile amnesia: a neurobiologicalperspective. In In- fant Memory, ed. M. Moscovitch, pp. 145-72. NewYork: Plenum
Onishi, K. H. & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science*, 308, 255.
Paulus, M., Hunnius, S., Vissers, M., & Bekkering, H. (2011a). Imitation in infancy: Rational or motor resonance? *Child Development, 82,* 1047-1057.
Paulus, M., Hunnius, S., Vissers, M., & Bekkering, H. (2011b). Bridging the gap between the other and me: The functional role of motor resonance and action effects in infants' imitation. *Developmental Science, 14*, 901-910.
Perner, J. (1991). *Understanding the representational mind.* Cambridge, MA. MIT Press
Perner, J., and T. Ruffman. 2005. Infants' insight into the mind: how deep? *Science 308:* 214–216.
Piaget J. (1962). *Play, Dreams and Imitation in Childhood*. Gattegno, S., Hodgson F.M. Eds.. Routledge & Kegan Paul: London
Piaget, J. (1985). *The Equilibration of Cognitive Structures: The Central Problem of Intellectual Development*. Chicago: University of Chicago Press
Pléh, Cs. (2010) *A lélektan története*. Budapest:Osiris
Povinelli, D. J., Landry, A. M., Theall, L. A., Clark, B. R., & Castille, C. M. (1999). Development of young children's understanding that the recent past is causally bound to the present. *Developmental Psychology*, 35(6), 1426-1439.
Prinz, W (1997). Perception and action planning. *European Journal of Cognitive Psychology*. 9: 129–154. doi:10.1080/713752551
Quinn, P.C., Yahr, J., Kuhn, A., Slater, A.M., Pascalis, O. (2002). Representation of the gender of human faces by infants: A preference for female. *Perception*, 31:1109–1121.
Rakoczy, H. (2012). Do infants have a theory of mind? *British Journal of Developmental Psychology* 30: 59–74.
Russel, J., Hanna, R. (2012), A Minimalist Approach to the Development of Episodic

Memory. *Mind & Language*, 27: 29–54. doi:10.1111/j.1468-0017.2011.01434.x
Scarf, D., Gross, J., Colombo, M. & Hayne, H. 2013. To Have and to Hold: Episodic Memory in 3- and 4-Year-Old Children. *Developmental Psychobiology. 55* (2) 125–132. DOI 10.1002/dev.21004
Schacter, D.L., Guerin, S.A., & St. Jacques, P. (2011). Memory distortion: an adaptive perspective. *Trends in Cognitive Sciences, 15*(10), 467-474.
Senju A, Csibra G (2008). Gaze following in human infants depends on communicative signals. *Current Biology*. 2008;18(9):668-71.
Sherif, M, Harvey, O. J., White, J. B., Hood, W. R., Sherif, C. W. (1954/1961) Intergroup Conflict and Cooperation: The Robbers Cave Experiment, *Classics in the History of Psychology*
Shteynberg, G. (2010). A Silent Emergence of Culture: The Social Tuning Effect. *Journal of Personality and Social Psychology*, Vol. 99, No. 4, 683–689
Sidanius, J., Pratto, F. (1993). The inevitability of oppression and the dynamics of social dominance. In P. Sniderman, P. Tetlock (Eds.). *Prejudice, politics, and the American dilemma* (pp. 173-211). Stanford, CA: Stanford University Press.
Silvia, P.J. (2012) Curiosity and Motivation. Ryan, R.M Ed. *The Oxford Handbook of Human Motivation*. DOI: 10.1093/oxfordhb/9780195399820.013.001
Simcock, G, & Hayne, H. (2003). Age-related changes in verbal and non-verbal memory during early childhood. Developmental Psychology, 39, 805–814.
Southgate, V., C. Chevallier, and G. Csibra. 2010. Seventeen-month-olds appeal to false beliefs to interpret others' referential communication. *Developmental Science 13*: 907–912.
Southgate, V., Senju, A. & Csibra, G. (2007). Action anticipation through attribution of false belief by two-year-olds. *Psychological Science,* 18, 587-592.
Spelke, E (2000). Core Knowledge. American Psychologist. 55 (11): 1233–1243. doi:10.1037/0003-066X.55.11.1233
Spelke, E.S., Kinzler, K.D. (2007). Core knowledge. *Developmental Science,* Vol. 10, Issue 1, 89-96.
Sperber, D., Hirschfeld, L.A. (2004). The cognitive foundations of cultural stability and diversity. *Trends in Cognitive Sciences*, Vol.8, No.1, 40-46.
Sperber, D., & Wilson, D. (1986). Relevance: Communication and cognition. Cambridge, MA: Harvard University Press.
Sperber, D., Wilson, D. 2002. Pragmatics, modularity, and mind-reading. *Mind and Language* 17., 3-23.
Suddendorf, T., Nielsen, M., & von Gehlen, R. (2011). Children's capacity to remember a novel problem and to secure its future solution. *Developmental Science*, 14(1), 26–33. doi:10.1111/j.1467-7687.2010.00950.x
Surian, L., Caldi, S. & Sperber, D. (2007). Attribution of beliefs by 13-month-old infants. *Psychological Science,* 18, 580-586.
Tajfel, H., & Turner, J. C. (1986). The social identity theory of intergroup behavior. In S. Worchel & L. W. Austin (Eds.), *Psychology of intergroup relations* (pp. 2-24). Chicago: Nelson-Hall.
Tajfel, H., Billig, M., Bundy, R. P. and Flament, C. (1971). Social categorization and intergroup behaviour. *European Journal of Social Psychology*, Vol. 1, Issue 2, 149-178
Tomasello, M. & Carpenter, M. (2007). Shared intentionality. *Developmental Psychology*, 10, 121-125.
Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain*

Sciences, 28, 675 - 691.
Tulving, E. (1972). Episodic and semantic memory. In E. Tulving and W. Donaldson (Eds.), *Organization of Memory* (pp. 381–402). New York: Academic Press.
Tulving, E. (2005). Episodic memory and autonoesis: Uniquely human. In H. Terrace & J. Metcalfe (Eds.), *The Missing Link in Cognition* (pp. 4–56). New York: Oxford University Press.
Walden, T.A., Ogan, T.A. (1988) The development of social referencing. Child Development, 59(5):1230-40.
Want, S.C., Harris, P.L. (2002) How do children ape? Applying concepts from the study of non-human primates to the developmental study of 'imitation'in children. *Developmental Science, 5 (1)*. 1–14. doi/10.1111/1467-7687.00194
Westermann, G.; Mareschal, D.; Johnson, M. H.; Sirois, S.; Spratling, M. W.; Thomas, M. S. C. (2007). "Neuroconstructivism". *Developmental Science, 10 (1*): 75–83. doi:10.1111/j.1467-7687.2007.00567
Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development, The truth about false belief. *Child Development, 72,* 655–684.
Wimmer H; Perner J (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition. 13* (1): 103–128. doi:10.1016/0010-0277(83)90004-5