Primates (2010) 51:337–351
DOI 10.1007/s10329-010-0208-4

ORIGINAL ARTICLE

Two-year-old children copy more reliably and more often than nonhuman great apes in multiple observational learning tasks

Claudio Tennie · Kathrin Greve · Heinz Gretscher · Josep Call

Received: 14 January 2010 / Accepted: 9 July 2010 / Published online: 5 August 2010

Abstract Individuals observing a proficient model can potentially benefit by copying at least one of the following three elements: motor movements (i.e., actions), goals, and results. Although several studies have investigated this issue in human infants, there are still very few studies that have systematically examined great apes' ability to spontaneously copy each of these three elements (particularly in comparison with human infants). We tested great apes and human children with eight two-target puzzle boxes—with varying levels of difficulty—to isolate the aspects that the various species may be more prone to copying. We found first trial evidence for observational learning of actions, goals, and results in children. Some copying was found for apes as well, but only if their performance was averaged across trials.

Keywords Imitation · Emulation · Observational learning · Two-action task

Introduction

Some forms of observational learning (e.g., imitative learning) are regarded as instrumental for the acquisition of language and the development of human culture (Tomasello et al. 1993; Tomasello 1999). Contrasting the observational learning abilities of humans and nonhuman great apes (henceforth apes) has allowed researchers to test some aspects of the cultural learning hypothesis (Tomasello et al. 1993; Tomasello 1999). In fact, the last few years have witnessed a flurry of comparative studies devoted to identify the similarities and differences between species in the information that they acquire from demonstrations (e.g., Nagell et al. 1993; Call and Tomasello 1995; Whiten et al. 1996; Stoinski et al. 2001; Herrmann et al. 2007). Carpenter and Call (2002) distinguished three types of information that observers can extract from demonstrations: body movements used by the demonstrator (actions), effects of those actions on the environment (results), and objective of the demonstrator in executing the actions (goals). It has been proposed that the propensity to copy the actions of others, especially when they are not part of the observer's repertoire, distinguishes humans from apes (Call and Carpenter 2002). Although children can copy both actions and results reliably (Huang et al. 2002; Huang and Charman 2005; Tennie et al. 2006), several studies have reported that children copy actions more readily than they copy results (object movements: Nagell et al. 1993; Meltzoff 1995; Tennie et al. 2006; static end-states: Bellagamba and Tomasello 1999; Call et al. 2005; but see also Thompson and Russell 2004). In contrast, action copying in apes is more elusive (Tomasello et al. 1987; Nagell et al. 1993; Call and Tomasello 1994; Myowa-Yamakoshi and Matsuzawa 1999; Call et al. 2005; Tennie et al. 2006). Some researchers have argued that for apes, result copying is the preferred way to extract information from demonstrations (Tomasello 1999; Myowa-Yamakoshi and Matsuzawa 2000; Call et al. 2005). In general, chimpanzees have been characterized as focusing more on results rather than the actions that produce those results, whereas children seem more focused on the actions rather

---

Electronic supplementary material The online version of this article (doi:10.1007/s10329-010-0208-4) contains supplementary material (ESM), which is available to authorized users.

C. Tennie (✉) · K. Greve · H. Gretscher · J. Call
Department of Developmental and Comparative Psychology,
Max Planck Institute for Evolutionary Anthropology,
Deutscher Platz 6, 04103 Leipzig, Germany
e-mail: tennie@eva.mpg.de

than the results. The contrast between apes and children in their preferred learning mechanism may be at the core of cumulative forms of culture (Tomasello 1999; Tennie et al. 2009), particularly when no objects or artifacts are involved (e.g., gestural communication). This is not to say that apes cannot ever learn to copy actions, including novel actions. Several studies have shown that chimpanzees and orangutans can be trained to imitate on command (Hayes and Hayes 1951; Custance et al. 1995; Miles et al. 1996), and enculturated apes are more proficient than mother-reared chimpanzees at reproducing actions on objects (Tomasello et al. 1993; Bering et al. 2000; Bjorklund et al. 2002; enculturated chimpanzees might also be better problem solvers, see Furlong et al. 2008). This could possibly mean that the underlying reason for the ape reluctancy to copy actions is more one of motivation than competence.

In recent years, the question of chimpanzee cultures has received renewed interest (e.g., Whiten et al. 1999) and a handful of experimental studies by Whiten and colleagues convincingly show the acquisition and transmission of behaviors within and between groups (Whiten et al. 2005, 2007; Horner and Whiten 2005; Horner et al. 2006; Bonnie et al. 2007; Hopper et al. 2007). Generally, these studies involved chimpanzees observing other chimpanzees or humans opening different types of puzzle boxes to get a reward, or placing objects in one of two locations. Most of these studies, however, have not distinguished between the different types of information that observers extract from the demonstrators. In some of these studies, chimpanzees could observe both the demonstrator's actions and the results produced by those actions (including the movement of subparts of the puzzle boxes), whereas in others, chimpanzees may have simply learned the place where to apply an action, not the action itself. There are three possible exceptions to this. (1) Horner and Whiten (2005) proposed that chimpanzees were more likely to copy the actions of a human demonstrator when their puzzle box was opaque than when their puzzle box was transparent. In the latter case, the chimpanzees used their own strategy, i.e., not the same one they observed from the model. However, the actions that were demonstrated (and that were later used by the chimpanzees) come naturally to chimpanzees. The task involved inserting a stick into holes, an action (or result) that does not need to be imitated by chimpanzees. We believe that Horner and Whiten have sufficiently shown instead that chimpanzees are sensitive to locations with regard to causal information—with the necessary actions "filled in" individually (i.e., not imitated). And so, we do not believe this study conclusively showed action copying in chimpanzees (see also Tennie et al. 2006). (2) More recently, Hopper et al. (2007) set out to tease apart the different contributions of information about results and actions in the acquisition of novel responses. They found no evidence for result copying if subjects only witnessed the objects moving on their own accord (ghost condition), but did find copying after subjects witnessed both action and result information (social demonstration condition). This is the first study to suggest that chimpanzees paid more attention to actions than results after the contribution of results was analyzed separately. In contrast, Hopper et al. (2008) found some evidence that the ghost condition did, in fact, elicit copying in chimpanzees, albeit less so than did a full demonstration condition. Also, Tennie et al. (2006) found that apes, unlike children, did not benefit differentially from watching a conspecific open a swing door to get a reward compared with seeing the swing door open on its own in a ghost condition (in fact, the apes did not copy in either condition). (3) Tennie et al. (2010) showed that chimpanzees required no target action demonstrations—the presentation of results information (brought about by completely different actions) was enough to enable subjects to copy. Thus, the results about action and result copying are currently mixed, with some studies finding evidence for action copying, result copying, result copying and action copying, or neither. One reason for this difference may be that particular apparatuses favor one of the two types of information available from demonstrations.

Our goal in this study was to resolve the discrepancy between studies by presenting apes and children with multiple tasks rather than a single task, and by manipulating the type of information (actions, goals, and results) that they could witness from demonstrations. The comparison with children was important to relate our findings to the developmental literature. The use of multiple tasks rather than a single task was important for three reasons. First, it offers us a better estimate of the observational learning propensities of individuals. Second, multiple tasks serve as controls for each other and render it unlikely that the obtained results (either positive or negative) are due only to a task-specific feature. Third, copying actions or results (or goals) likely is only effective within an optimal "copying zone": if the task is too easy, apes may ignore demonstrations because they can solve the task on their own; but if the task is too hard, apes may not be able to benefit from the information they witness. It is very hard to know a priori where this optimal zone lies and how large it may be. Several authors have argued that this zone may be particularly narrow in the great apes compared with humans (Call et al. 2005; Tennie et al. 2006; Hopper et al. 2007).

We compared apes and children in two sets of four tasks in which they had to manipulate a puzzle box to get a reward after witnessing a demonstration. Our aim was to present tasks that seemed likely to represent different levels of difficulty. We tried to achieve this by using a variety of different tasks and by taking into account the demands of former tasks (e.g., Whiten et al. 1996). Each task was based on the two-target method (compare Meltzoff and Moore 1977 and Dawson and Foss 1965); that is, each box was designed so that it could be opened in two different ways (henceforth "target methods"—which, again, differed between tasks), and different groups of subjects experienced one of the two target methods for each task. In the full demonstration condition, an ape or human demonstrator showed the solution to the task, thereby simultaneously producing information about actions, goals, and results. In the intention condition, demonstrators produced information about actions and goals (and for human children, this was reduced to information about goals in set 2). In the end-state condition, subjects only received information about end results (e.g., Bellagamba and Tomasello 1999). Finally, in the baseline condition, no information was provided about how to open the box.

Methods

Subjects

Humans

Seventy-four human infants [32 in set 1 (15 girls, 17 boys), 42 in set 2 (22 girls, 20 boys)] were recruited through the child laboratory of the Max Planck Institute for Evolutionary Anthropology in Leipzig, Germany. Infants' ages were 24 month ± 2 weeks [average age 2 years, standard deviation (SD) 8.46 days]. None of the children had previously participated in similar observational learning experiments. The children were tested individually in the presence of a parent in a silent room at the institute. Each child was tested with only one set of apparatuses. Stickers were used as rewards for the children (and were inserted into the apparatuses for demonstrations as well as trials).

Apes

Thirty-six great apes [five bonobos (two females, three males), five orangutans (four females, one male), five gorillas (four females, one male), and 21 chimpanzees (15 females, six males)] participated in the study. There were six juveniles (up to 6 years), nine subadults (6–9 years), and 21 adults (10 years and older) [mean age at the beginning of the study = 14.72 years, SD 9.26 years; see Tables 2 and 3 in Appendix (ESM)]. After being tested in baseline or end-state conditions, 20 subjects were trained to become demonstrators for their conspecifics. All apes were socially housed at the Wolfgang Köhler Primate Research Centre (WKPRC) at Leipzig Zoo, Germany. Testing took place in rooms designated for experimental studies. Most of the apes had already participated in several cognitive experiments before, some also concerning observational learning. However, they had neither been trained to copy actions nor did they have any experience with comparable apparatuses to those used in the study. Each ape was tested with both sets of apparatuses. Grapes were used as rewards in the same way as stickers were for children. All ape subjects were fed a diet of fruit, vegetables, pellets, and occasional eggs and were regularly supplied with leafy twigs of various tree and shrub species. None of the animals were deprived of food or water. For the gorillas, enclosure sizes were 2,300 m^2 (outdoor enclosure) and 264 m^2 (indoor enclosure); for the chimpanzees "A" group 4,000 m^2 (outdoor enclosure) and 430 m^2 (indoor enclosure); for the chimpanzees "B" group 1,400 m^2 (outdoor enclosure) and 175 m^2 (indoor enclosure); for the bonobos 2,300 m^2 (outdoor enclosure) and 256 m^2 (indoor enclosure); and for the orangutans 1,680 m^2 (outdoor enclosure) and 230 m^2 (indoor enclosure). Sizes of the sleeping cages for each species were between 32 and 47 m^2. There were various climbing structures and various enrichment devices, such as food that required processing or tool use (e.g., puzzle boxes) and ropes for all apes. The indoor temperature (including sleeping rooms) was set to at least 23°C. If the outside temperature dropped, active heating ensured that the inside temperature did not drop.

Dropouts

Due to motivational problems, three ape subjects (two chimpanzees and one gorilla) only participated in one of the two sets. An additional gorilla received no social demonstrations for one of the two sets, as she was relocated to another zoo during the study period. In 11 cases, we were unable to train a conspecific demonstrator (five cases in gorillas, six cases in chimpanzees; all in set 1) so no data could be collected in these particular cases. Additionally, 18 trials had to be discarded due to the following reasons: subjects refused to participate; demonstration failures; offspring intervention; and camera failure. Altogether, this resulted in 118 missing trials (from a total of 1,152). For the children, 16 trials (from a total of 296 trials) had to be omitted due to demonstration errors, parents' intervention, motivational problems, and premature contact with the apparatus.

Apparatuses

We used two sets of four puzzle boxes. Each puzzle box contained a visible reward locked inside. All eight puzzle boxes (see Fig. 1) differed in their outer appearance and the actions required to get access to the reward, but every apparatus' case could be opened in two different ways (i.e., using one of two target methods). Due to the differences in hand size and overall strength between children and apes, we made a child and an ape version of each apparatus that differed in dimension and resistance to wear and tear. Apes' apparatuses were larger and more resistant than the children's counterparts. One of the puzzle boxes from set 1 (Dumbbell) had to be replaced for the apes because it could not be properly attached to the substrate for presentation (which resulted in an overall total of nine puzzle boxes). Therefore, apes received an alternative apparatus (Moveable Window) that was similar in appearance and handling: both apparatuses consisted of a tube with a centrally located reward, and both required subjects either to pull in the direction of the tube axis or to manipulate the tube centrally (breaking for the Dumbbell, twisting for the Moveable Window; please note here that some children also twisted the Dumbbell in order to get the reward).

For training and demonstration purposes, we built a rigged version of each apparatus so we could surreptitiously block one (or both) of the two possible solutions. Rigged apparatuses ensured that only one target method could be demonstrated to the apes and allowed the human demonstrator of set 2 to convincingly provide intention demonstrations. We used the following apparatuses (see Fig. 1 for an overview):

Dumbbell

Two plastic tubes (length 8 cm, diameter 2.5 cm) that were glued together by three small stripes of clear adhesive tape. The ends of the resulting tube were closed by removable caps (length 1.4 cm, diameter 2.7 cm) (Fig. 1a). The reward was centered in the tube and could be obtained by either breaking the tube in its two halves or pulling one of the caps off (see Call et al. 2005; Meltzoff 1995). This apparatus was exclusively used with children.

Moveable Window

A horizontally mounted plastic bar (length 30 cm, diameter 4 cm) with a hole (4 × 4 cm) in the upper middle where the reward was placed (Fig. 1b). A transparent tube with a corresponding opening covered the plastic bar (tube: length 10 cm, diameter 5 cm; opening: 5 × 5 cm). Initially, the opening was orientated downward, thereby blocking access to the clearance (Fig. 1b). The rewards could be obtained by either twisting the tube around so that the opening and the bar's clearance aligned or by sliding the whole outer tube to the side. This apparatus was exclusively used with apes.

Bullet

A horizontally mounted Plexiglas tube and three wooden bolts. For children: tube: length 18.5 cm, diameter 3.5 cm;

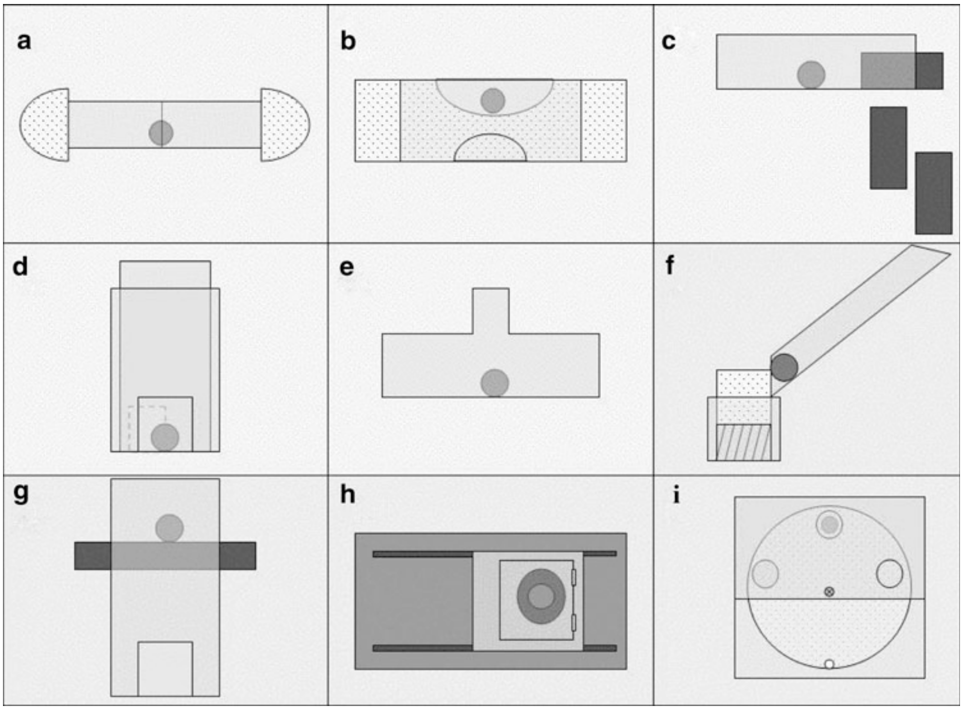

Fig. 1 Apparatuses: Set 1: **a** Dumbbell (*top view*); **b** Moveable Window (*front view*); **c** Bullet (*front view*); **d** Window-Tower (*front view*), **e** T-Pipe (*front view*). Set 2: **f** Sponge Cylinder (*side view*); **g** Rod Cylinder (*front view*); **h** Hood-Slide (*top view*); **i** Discman (*top view*)

bolts: length 6 cm, diameter 3 cm. For apes: tube: length 35 cm, diameter 5 cm; bolts: length 10 cm, diameter 4 cm (Fig. 1c). The reward rested in the middle of the tube and could be retrieved by either quickly propelling one bolt through the tube or by inserting all three bolts successively.

Window-Tower

A hollow, upright, Plexiglas tube (length 19 cm, diameter 7 cm) fixed to the ground, with an opening near the bottom (5 × 5 cm). A moveable tube (length 16 cm, diameter 5.5 cm) with a corresponding opening (4.5 × 4.5 cm) fitted into it, with its opening directed in the opposite direction to the original opening; the latter tube was sealed at the top. Both openings were marked to make the tubes more distinguishable (Fig. 1d). The reward was lying on the ground within the tubes and could be accessed by either twisting the inner tube until the openings aligned or by pulling out the inner tube.

T-Pipe

A large, hollow, horizontally mounted (at a height of about 15 cm) Plexiglas tube (length 30 cm, diameter 7 cm) with a hollow smaller tube (length 5 cm, diameter 5 cm) orthogonally embedded in its middle. This smaller tube was orientated upward in an inverted T-shape. The three openings of the T-shape were marked. The reward was placed in the middle of the bigger, horizontal tube directly below the center of the smaller tube (Fig. 1e). The reward could be obtained by either pulling the small tube toward oneself so that the large tube rotated and the reward fell down through the small tube, or by pushing down one side of the large horizontal tube to make the reward fall out from the side of the large tube.

Sponge Cylinder

A tilted, hollow, Plexiglas tube (length 10 cm, diameter 1.5 cm, slope 35°) obstructed at its base by a solid plastic cylinder (length 5 cm, diameter 3.8 cm). This cylinder rested on a piece of sponge inside another hollow Plexiglas tube (length 6 cm, diameter 4 cm) that did not obstruct the tilted Plexiglas tube (Fig. 1f). The reward was lying at the bottom end of the tilted tube and could be obtained by either pulling out the obstructing plastic cylinder, allowing the food reward to roll out over the sponge where it could be recovered, or by pressing the cylinder down into the sponge, allowing the food reward to roll out over the top of the cylinder.

Rod Cylinder

A vertically mounted Plexiglas tube (length: 19 cm, diameter 6 cm) with both ends closed. In the side, at the base, was an opening (5 × 5 cm). Two opposing holes in the sides of the upper part of the tube allowed a wooden rod (length 10 cm, diameter 2.5 cm) to be horizontally placed across the tube (Fig. 1g). The reward rested on the rod and was accessible only by moving the rod to the left or right, allowing it to drop down and be retrieved through to the opening.

Hood-Slide

A hinged Plexiglas door (9 × 10 cm) covering a hole that contained the reward. The door could be moved along a track or be opened upwards like a trap door (Fig. 1h). The reward could be obtained by either sliding the Plexiglas plate or by lifting it up.

Discman

Middle plate: We cut out a disc from the centre of a Plexiglas square (24 × 24 cm). The disc was reinserted so that it could be rotated. A hole was cut at 12 o'clock in the disc big enough for a reward to be placed. A handle was placed at 6 o'clock on the disc. Back plate: A Plexiglas sheet, the same dimensions as the middle square, was placed behind the middle plate. A small hole was cut into this plate corresponding to 10 o'clock on the disc. Front plate: This plate was slightly over half the size of the other plates and covered the top front half of the disc. A small hole was cut into the plate corresponding to 2 o'clock on the disc. The three plates were fastened together with a central screw that allowed the disc to spin on a central axis (Fig. 1i). A reward was placed in the disc's hole. At the 12 o'clock position, it was inaccessible. By using the handle, the disc could be rotated. If the reward was rotated clockwise to 2 o'clock, it could be retrieved through the corresponding hole in the front plate. If it was rotated anti-clockwise to 10 o'clock, the reward could be retrieved through the hole in the back plate. The whole apparatus was inclined (slope 45°) backward and fixed at this inclination to a horizontal Plexiglas base. This allowed any reward rotated toward the back hole to fall out, whereas the rewards rotated to the hole in the front plate could simply be picked out.

Procedure

The general procedure consisted of giving subjects a demonstration about how to open the apparatus (demonstration phase) and then allowing them to manipulate the apparatus to see if they could open it (test phase). The exact procedure used, however, differed between children and apes.

Child testing

Child testing took place at a table (80 × 80 cm) with the experimenter and the child (sitting on his or her parent's lap) facing each other. The apparatus was presented on a moveable platform (65 × 39 cm) outside of the subject's reach while the experimenter administered one of the conditions. The experimenter encouraged the child to watch by saying "Watch what I'm doing here" ("Guck mal, was ich mache") before the first demonstration and "Look" ("Guck mal") before each of the two subsequent demonstrations. During this phase, the parent gently restrained the child in case he or she attempted to grab the apparatus before the demonstration period was over. Once the child had witnessed a demonstration, the experimenter drew a curtain (65 × 53 cm), blocking the child's visual access to the apparatus and (if necessary) reset the apparatus to its initial (unopened) state. After all demonstrations had been given the experimenter pulled back the curtain, pushed the apparatus toward the child, and invited him or her to use the apparatus for 40 s in order to get the reward by stating: "Now it's your turn"("Und jetzt du"). If the child asked the parent or experimenter for help, the same promp words were given but no help was provided.

Ape testing

Ape testing took place in two adjacent cages (about 15 m^2 total) with a booth (approximately 1 m^2) between them (see Fig. 2). The apparatus was mounted inside a metal-framed Perspex cube (50 × 70 × 55 cm^3). Apes watched the demonstrations from cage A. Demonstrations that involved another subject were carried out by a conspecific belonging to the subject's social group. Subjects received one verbal prompt from the experimenter ("Come over and look"). If the subject remained in a position where its line-of-sight into the cube was obscured, it was again prompted with the same words. After the subject witnessed the demonstrations from cage A, the demonstrator left the testing area, and the subject was moved to room B where it could access the apparatus. If the apparatus needed resetting, the experimenter inserted occluders that blocked the subject's visual access to the apparatus. In two of the conditions in which no conspecific offered a demonstration (end-state, baseline), the subject accessed the apparatus from room A. Just as with children, apes were verbally encouraged to manipulate the apparatus after they had access to it.

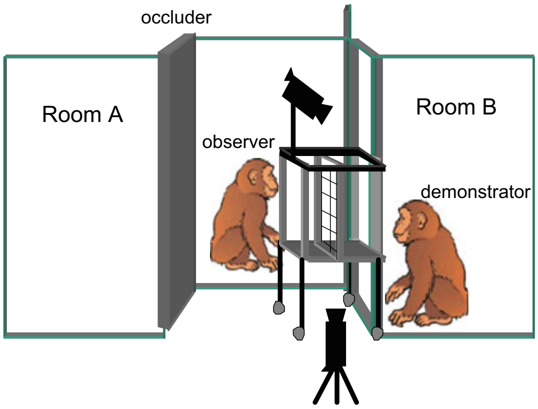

Fig. 2 Setup for great apes

Conditions

We administered one control (baseline) and three experimental conditions, as follows:

Full model

Subjects witnessed a conspecific manipulating and opening the apparatus using one of the two target methods and getting the reward after each demonstration, respectively. Thus, in this condition, information about actions, goals, and results was available.

Intention

Same as in the full-model condition except that demonstrators failed to produce the desired result to obtain the reward. Hence, subjects received no information about results (i.e., they did not observe any part of the apparatus moving).

End state

The apparatus was presented in the physical state in which it came to rest after one of the two target methods had been applied and the reward had been extracted. Therefore, this condition only revealed information about results (more specifically, information about static end states).

Baseline

The apparatus was presented in its initial state without any prior information about how to get the reward. Note that rewards were present and visible in all four conditions, though the full model condition was the only condition in which subjects observed the successful retrieval of the reward. Prior to demonstrations, in the end-state condition, the reward was placed to where it would have moved if the respective target method had been applied. Rewards in the intention and the baseline conditions remained in their initial position.

Test phase

After the demonstration phase, subjects having observed the minimum amount of demonstrations (see below) were allowed to manipulate the apparatus. The specific procedures for the presentation of the respective conditions and testing differed as follows.

Child testing

Social demonstrations (full model and intention) were given by an adult human experimenter (set 1: female; set 2: male). Subjects were required to watch at least one out of three demonstrations in full before getting access to the apparatus. For human children, there were two versions of the intention condition corresponding to each of the two sets of apparatuses. In set 1, the demonstrator gripped the relevant parts of the apparatus and executed the appropriate hand movements but her hands "accidentally" slipped off the manipulandum (e.g., Meltzoff 1995). Hence, information about both actions and goals were available. In set 2, the demonstrator also gripped the relevant parts of the apparatus but no slipping movement occurred. Thus, the solution was indicated by grip posture and direction of force application alone, thereby only revealing information about goals. In both sets, the experimenter used facial expressions and vocalizations as if applying real force. Intention demonstrations following the first one were initiated by the experimenter exclaiming "again" ("nochmal") (just before encouraging the child to watch). In the two nonsocial conditions (end state and baseline), subjects could survey the apparatus without touching it for 10 s before getting access. Children received one trial per condition in a single session (i.e., a total of four trials). Each trial lasted a maximum of 40 s.

Ape testing

Social demonstrations were given by a trained ape. Thirteen demonstrators were female and seven were male [one juvenile, six subadults, and 13 adults; mean age at the beginning of the study = 14.15 years, standard deviation (SD) 6.1 years]. Subjects had to watch at least three out of ten demonstrations in full before getting their first test trials on their first testing day. To ensure demonstration of only the desired target method, ape demonstrators received prior training in the performance of only one target method, and furthermore, they acted on modified apparatuses in which either the alternative target method (full model) or both target methods (intention) had been blocked. In the intention condition, ape demonstrators sometimes displayed methods other than the one desired after failing several times. The frequency with which this occurred was generally low (7.5% of cases in set 2, for which this data was available). Whenever this happened, demonstrations were interrupted and the demonstrator was retrained with a rigged apparatus out of sight of the observer, and then the demonstration phase resumed (overall, this introduced a delay of about 15 min to demonstrations). In the two nonsocial conditions (end state and baseline), presentations lasted 5 min (end state) and 60 s (baseline), respectively. Ape subjects received four trials per condition, divided into two blocks of two consecutive trials on different days (i.e., a total of 16 trials per set). Apes got an equal amount of demonstrations on both testing days, though on their second day, they were given access to the apparatus regardless of how many demonstrations they had observed (i.e., once subjects met the required minimum of three observed demonstration for the first testing day, the second day followed automatically on the subsequent session). A trial for the apes lasted a maximum of 5 min.

Design

Each subject received all four conditions in each set but each administered with a different apparatus. Hence, each set was composed by four apparatuses. We counterbalanced the apparatus type and condition across subjects.

Child testing

One group of children was tested with the apparatuses from set 1, whereas a second group was tested with the apparatuses from set 2. The time elapsed from the administration of set 1 and set 2 prevented us from testing the same children with both sets of apparatuses. Children received the apparatuses in the following order: (set 1: Dumbbell, Bullet, Window-Tower, T-Pipe; set 2: Rod Cylinder, Sponge Cylinder, Discman, Hood-Slide), and within this fixed order, the conditions and demonstrated target methods were then counterbalanced. We assigned the same number of male and female children to each condition.

Ape testing

Apes received both set 1 and set 2 (in this order). Thus, the condition was a within-subject factor, whereas set type differed between children (between-subject) and apes (within-subject). We also counterbalanced the two possible solutions to each puzzle box. For the apes, tasks and conditions were counterbalanced, and the order in which the subject received the condition was random. Due to the groups' composition, it was not possible to counterbalance age and sex for the apes, but we attempted to distribute subjects across conditions in the most balanced manner possible.

Data scoring and analysis

Testing was videotaped by two cameras from different angles, and trials were scored from the videotapes. We scored whether subjects got the reward (success) and their manipulations of the apparatus (target method 1, target method 2, other methods and exploratory acts). We only counted those manipulations as methods (including, but not restricted to, target methods) that were aimed at the reward: e.g., simply touching the apparatus would not be counted as applying any method, but poking into an opening, for example, would be. Target methods did not need to be fully performed to be recorded as such: e.g., manipulating the apparatus only half way in a target way (i.e., in one of the two ways that subjects may have observed) would count as performing this particular target method. Target methods were counted regardless of whether subjects had ever seen them demonstrated (since target methods could come about by chance as well). We paid particular attention to the first method employed by the subject (first attempt), which is considered to be most informative for observational learning studies (see Huang et al. 2002). A second observer scored 20% of the subjects to assess the interobserver reliability of performance. Interobserver reliability for children's first acts and first target method as well as for apes' first target method was excellent (Cohen's $\kappa = 0.82$, $\kappa = 0.95$ and $\kappa = 0.93$, respectively). Interobserver reliability for apes' first acts was good (Cohen's $\kappa = 0.72$).

We analyzed three dependent variables (success, first attempts, and matching score) as a function of condition and species. Additionally, we assessed the effect of task difficulty (as measured by baseline success rates) on the matching score. The main analyses are based on trial-one performances, but we also analyzed the performances for the apes when averaged across trials. We focused on the apes' initial performance also to compare human and ape performance. Thus, when comparing apes with children for success, first attempts, and task difficulty, the data always stems from subjects' first trials [and here apes' matching score could be based on any of their four test trials—depending on their actual timing of performance (see also below)].

The matching score indicated whether subjects used the same target method that they witnessed in the experimental conditions after taking into account the subjects' "natural" tendency displayed in the baseline [for illustration and formula, see Appendix, Table 1 (ESM)]. For the matching score analysis, only the first trial in which subjects applied one of the two target methods was considered (depending on the apes' performance, this could be any of their four test trials). If the subject matched the demonstrated target method, it was assigned a positive matching score corresponding to the expected occurrence probability of the alternative target method derived from baseline (i.e., the higher the baseline frequency of the demonstrated target method, the lower the score matching demonstrations). If subjects applied the alternative target method to the one demonstrated, they were assigned a negative matching score corresponding to the expected occurrence probability of the demonstrated target method (i.e., the higher the baseline frequency of the demonstrated target method, the higher the negative score subjects received for not matching demonstrations). For instance, suppose a subject received a full-model demonstration of target method 1 with the Window Tower and used target method 2 when given access to the apparatus. Further suppose that two out of eight apes and two out of eight children who received the Window Tower in the baseline condition applied target method 1 (subjects who never showed any target method were not included in the calculation of the baseline frequencies—and so, in this example, suppose that all eight apes and eight children eventually applied one of the target methods). This means that the expected occurrence frequency of target method 1 would be 0.25 ($[2 + 2]/[8 + 8]$). As the subject in our example did not match the demonstrated target method, he or she would receive a score of −0.25, whereas if he or she matched the observed target method, the score would be 0.75. In general, a positive matching score represents more matching than mismatching, whereas a negative score represents the opposite [see also Appendix, Table 1 (ESM)].

One of the key properties of the matching score is that it takes into account the subjects' natural tendency with each apparatus as measured during the baseline condition. Then it uses this information to assess whether subjects who witnessed additional information deviate from the pattern shown by subjects who did not witness anything (baseline). This is particularly important when multiple apparatuses are involved and two alternative target methods for opening each apparatus are not equiprobable. Severe biases constitute a major problem for a statistical approach when analyzing a two-target task in terms of matching or nonmatching responses, especially for small sample sizes. Our matching score solves this potential problem by taking into account potential natural biases. This means that a match for an apparatus with a strong bias toward the demonstrated target method has less influence on the matching score than a match for an apparatus with a strong bias toward the alternative target method. Consequently, this type of analysis constitutes an important refinement compared with the standard matching/nonmatching approach used in the field of observational learning research.

Note that for the seven apparatuses presented to apes as well as children, we found only one species difference in baseline frequencies for the respective target methods (for the Discman: Fisher's exact test, $P = 0.028$), which is within the limits of chance [see Appendix, Table 2 (ESM); Fisher's omnibus test, $\chi^2 = 12.9$, $df = 16$, $P = 0.68$]. Thus, we combined children's and apes' baseline data to increase accuracy of the respective target methods' estimated occurrence probabilities. The expectancy values used to determine the averaged apes' mean matching score was calculated in the same way, with the exception that ape baseline frequencies for the respective target methods were also averaged for all trials in which the subject applied a target method. All statistical tests were nonparametric and two-tailed.

Results

Success

Figure 3a presents the percentage of successful subjects (in the first trial) as a function of species and condition and task set.

Children

There were significant differences across conditions for children in set 1 (Cochran's $Q = 14.474$, $df = 3$, $n = 23$, $P = 0.002$) and set 2 (Cochran's $Q = 16.398$, $df = 3$, $n = 38$, $P = 0.001$). Pairwise comparisons indicated that children were significantly more successful in the full model than in the end state (McNemar; set 1: $n = 30$, $P = 0.006$; set 2: $n = 40$, $P = 0.002$) or the baseline condition (McNemar; set 1: $n = 30$, $P = 0.001$; set 2: $n = 41$, $P = 0.013$). Children in the full-model condition were also more likely to succeed than children in the intention condition for set 2 (McNemar; $n = 39$, $P = 0.0001$). Children's performance in sets 1 and 2 did not significantly differ for any condition (full model: $\chi^2 = 0.286$, $df = 1$, $n = 71$, $P = 0.733$; intention: $\chi^2 = 0.946$, $df = 1$, $n = 64$, $P = 0.438$; end state: $\chi^2 = 0.421$, $df = 1$, $n = 72$, $P = 0.635$; baseline: $\chi^2 = 2.211$, $df = 1$, $n = 73$, $P = 0.161$).

Apes

There were no significant differences across conditions in apes' first trial success in set 1 (Cochran's $Q = 0.6$, $df = 3$, $n = 24$, $P = 0.896$) or in set 2 (Cochran's $Q = 1.385$, $df = 3$, $n = 33$, $P = 0.709$). Additionally, apes performed significantly better in the full model and baseline condition of set 2 compared with set 1 (McNemar; full model: $n = 29$, $P = 0.039$; baseline $n = 32$, $P = 0.021$). No other significant differences were observed (McNemar; intention: $n = 23$, $P = 0.219$; end state: $n = 33$, $P = 0.125$). Regarding species differences, gorillas and chimpanzees seemed least proficient [(see Fig. 3b and Appendix, Table 3 (ESM)]. For both sets and all conditions, their level of success never exceeded that of the bonobos or orangutans, though species differences were found to be significant only in the baseline of set 1

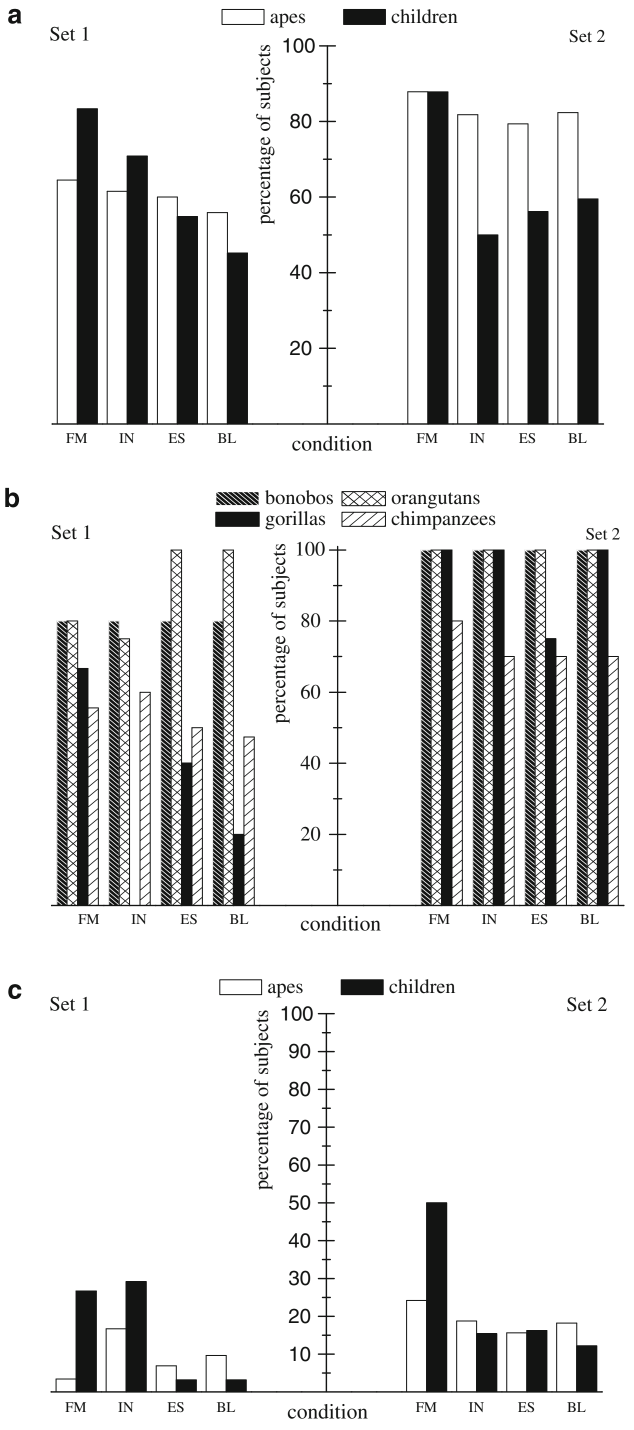

Fig. 3 **a** Percentage of apes and human children who solved the tasks in the first trial across conditions **b** split up by ape species; **c** percentage of apes and human children directly applying a target method in the first trial

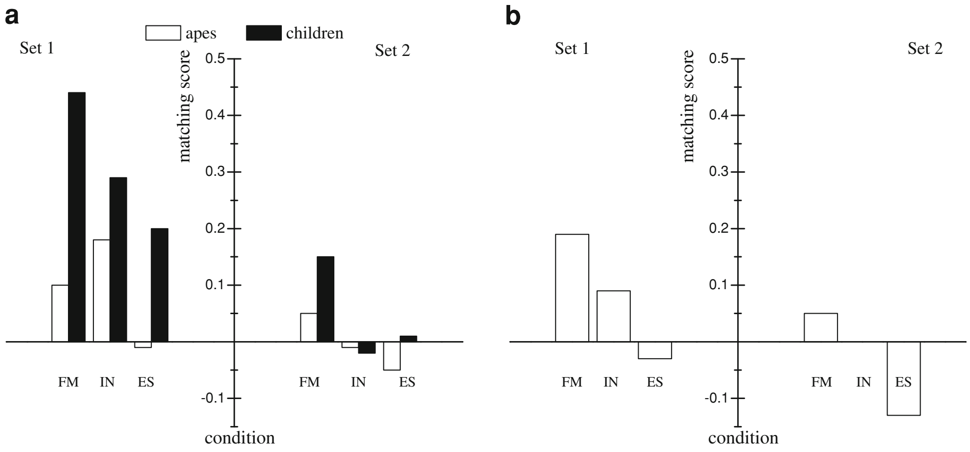

Fig. 4 Mean matching scores for each condition **a** for apes' and children's first trials and **b** averaged over all trials in that apes applied one of the two target methods

(Fisher's exact; $P = 0.034$). Post hoc tests revealed orangutans to be more successful in baseline than gorillas (Fisher's exact; $P = 0.048$). Using apes' average success over all four trials instead of just the first trial data also produced no significant differences across conditions (Friedman test; set 1: $\chi^2 = 0.877$, $df = 3$, $n = 24$, $P = 0.843$; set 2: $\chi^2 = 2.391$, $df = 3$, $n = 33$, $P = 0.522$).

First attempts

Figure 3c presents the percentage of subjects who immediately (i.e., without applying exploratory touches or other methods beforehand) used a target method (not necessarily the demonstrated one) as a function of condition, task set, and species.

Children

There were significant differences across conditions for children in set 1 (Cochran's $Q = 8.581$, $df = 3$, $n = 23$, $P = 0.031$) and set 2 (Cochran's $Q = 12.878$, $df = 3$, $n = 34$, $P = 0.004$). Pairwise tests for set 1 revealed that children used a target method more often in the full model than the end-state (McNemar; $n = 30$, $P = 0.039$) or baseline conditions (McNemar; $n = 30$, $P = 0.039$). Additionally, children also used a target method more often in the intention than the end-state condition (McNemar; $n = 24$, $P = 0.034$). Pairwise tests for set 2 revealed that children used a target method more often in the full model than in any other condition (McNemar; intention: $n = 37$, $P = 0.012$; end state: $n = 37$, $P = 0.017$; baseline: $n = 40$, $P = 0.001$). Comparing conditions across sets revealed that children in set 2 used a target method with the same frequency in all conditions as those in set 1 (full model demonstration: $\chi^2 = 3.889$, $df = 1$, $n = 70$, $P = 0.084$; intention condition: $\chi^2 = 1.723$, $df = 1$, $n = 63$, $P = 0.214$; end-state condition: $\chi^2 = 3.082$, $df = 1$, $n = 68$, $P = 0.116$; baseline: $\chi^2 = 1.859$, $df = 1$, $n = 72$, $P = 0.227$).

Apes

There were no significant differences across conditions for apes' first attempts in set 1 (Cochran's $Q = 5.526$, $df = 3$, $n = 17$, $P = 0.137$) or set 2 (Cochran's $Q = 0.387$, $df = 3$, $n = 31$, $P = 0.943$). Comparing conditions across sets also revealed no significant differences (McNemar; full model: $n = 28$, $P = 0.07$; intention: $n = 22$, $P = 1$; end state: $n = 29$, $P = 1$; baseline: $n = 29$, $P = 0.508$). We also found no differences with regard to species' first attempts (Fisher's exact, separately for both sets and each condition: all $P > 0.1$).

Matching scores

Figure 4a shows the mean matching scores as a function of condition, task set, and species.

Children

Children's matching scores in set 1 were significantly higher than expected from the baseline in each condition (Wilcoxon signed pair; full model: $Z = -4.245$, $n = 29$, $P < 0.001$; intention: $Z = -2.220$, $n = 22$, $P = 0.025$; end state: $Z = -2.351$, $n = 24$, $P = 0.017$). Moreover, there were no significant differences across conditions (Friedman test; $\chi^2 = 1.216$, $df = 2$, $n = 14$, $P = 0.564$). Children's matching score in set 2 was significantly above chance in the full-model condition (Wilcoxon signed pair; $Z = -2.007$, $n = 39$, $P = 0.045$) but not in the intention (Wilcoxon signed pair; $Z = -0.277$, $n = 27$, $P = 0.791$) and end-state (Wilcoxon signed pair; $Z = -0.143$, $n = 24$, $P = 0.898$) conditions. However, there were no significant differences across conditions (Friedman test; $\chi^2 = 2.517$, $df = 2$, $n = 15$, $P = 0.308$). Comparing conditions across sets, results showed that children in set 1 matched full-model demonstrations more frequently than children in set 2 (Mann–Whitney; $U = 375$, $n = 68$, $P = 0.017$). There were also significant differences between sets for intention (Mann–Whitney; $U = 199$, $n = 49$, $P = 0.048$) but no differences for the end-state demonstrations (Mann–Whitney; $U = 221$, $n = 48$, $P = 0.169$).

Apes

Apes did not differ significantly from baseline levels in any of the conditions of set 1 (Wilcoxon signed pair; full model: $Z = -0.974$, $n = 25$, $P = 0.344$; intention: $Z = -1.539$, $n = 19$, $P = 0.135$; end state: $Z = -0.388$, $n = 26$, $P = 0.709$) or set 2 (Wilcoxon signed pair; full model: $Z = -0.695$, $n = 30$, $P = 0.497$; intention: $Z = -0.152$, $n = 29$, $P = 0.881$; end state: $Z = -0.476$, $n = 30$, $P = 0.643$). Moreover, there were no significant differences across conditions in set 1 (Friedman test; $\chi^2 = 1.087$, $df = 2$, $n = 12$, $P = 0.614$) or set 2 (Friedman test; $\chi^2 = 0.144$, $df = 2$, $n = 26$, $P = 0.948$) or between set 1 and set 2 in any of the conditions (Wilcoxon signed ranks; full model: $Z = -0.759$, $n = 24$, $P = 0.459$; intention: $Z = -0.544$, $n = 16$, $P = 0.606$; end state: $Z = -0.803$, $n = 24$, $P = 0.433$). Similarly, we found no significant differences between species (Kruskal–Wallis, separately for both sets and each information condition: all $\chi^2 < 4.07$; all $n \geq 19$, all $df = 3$, all $P > 0.1$) or human- and mother-reared apes (Mann–Whitney, all $U \geq 28.0$, all $n \geq 19$, all $P > 0.1$). Furthermore, we found no indication that apes' age had an effect (Spearman's ρ range −0.30 to 0.11, all $P > 0.1$). No ape demonstrator-related characteristic had an effect on subjects' matching scores [see Appendix, Table 5 (ESM)]: subjects were not more likely to follow demonstrations by dominants than by subordinates (Mann–Whitney, separately for both sets and each social condition: all $U \geq 31.0$, all $n \geq 19$, all $P > 0.1$). They also did not copy older demonstrators' actions more than those of younger ones (Mann–Whitney, all $U \geq 31.0$, all $n \geq 19$, all $P > 0.1$), and they also did not match males' demonstrations more closely than females' (Mann–Whitney, all $U \geq 37.5$, all $n \geq 19$, all $P > 0.1$). The latter negative finding even holds if we additionally consider the subject's gender as a factor (leading to four subject–demonstrator gender combinations; Kruskal–Wallis; all $\chi^2 < 2.67$, all $df = 3$, all $P > 0.1$).

Figure 4b shows apes' mean matching scores averaged over all trials in which subjects applied one of the two target methods. Thus averaged, apes matched the demonstrated target methods above levels expected from baseline in the full-model condition of set 1 (Wilcoxon signed pair; $Z = -2.669$, $n = 25$, $P = 0.006$). In all other conditions of set 1 and in all conditions of set 2, apes' performances did not differ significantly from baseline expectations (Wilcoxon signed pair; set 1: $Z = -1.663$, $n = 19$, $P = 0.100$; end state: $Z = -0.432$, $n = 26$, $P = 0.675$; set 2: full model: $Z = -0.783$, $n = 30$, $P = 0.442$; intention: $Z = -0.022$, $n = 29$, $P = 0.987$; end state: $Z = -1.575$, $n = 30$, $P = 0.118$). Comparing across conditions revealed a significant difference for set 1 (Friedman test; $\chi^2 = 6.500$, $df = 2$, $n = 12$, $P = 0.038$), though post hoc pair-wise comparisons were not significant (Wilcoxon signed pair; full model–intention: $Z = -1.727$, $n = 14$, $P = 0.087$; full model–end state: $Z = -1.887$, $n = 22$, $P = 0.059$; intention–end state: $Z = -1.676$, $n = 15$, $P = 0.097$). Likewise, there was no significant difference between conditions in set 2 (Friedman test; $\chi^2 = 0.077$, $df = 2$, $n = 26$, $P = 0.990$).

Effect of task difficulty

Figure 5 plots subjects' baseline success against the matching score for each of the apparatuses as a function of experimental condition and species.

Children

For children, no significant correlations were found (Spearman correlation; full model: $\rho = -0.51$, $n = 8$, $P = 0.196$; intention: $\rho = -0.08$, $n = 8$, $P = 0.851$; end state: $\rho = -0.05$, $n = 8$, $P = 0.918$).

Apes

There was a significant negative correlation for apes between baseline success and matching score in the intention condition (Spearman correlation; intention: $\rho = -0.88$, $n = 8$, $P = 0.001$), whereas we found no significant correlation for apes in the full-model condition (Spearman correlation; $\rho = 0.06$, $n = 8$, $P = 0.898$) or in the end-state condition (Spearman correlation; $\rho = -0.31$, $n = 8$, $P = 0.45$).

Discussion

Children matched targets in all experimental conditions of set 1 and the full-model condition of set 2. Their level of copying during their very first attempt was especially marked in the full-model condition in both sets, as well as in the intention condition of set 1. In contrast, we found no evidence that the great apes copied the information provided in any of the experimental conditions and for any of the apparatuses in their first trial. However, we found evidence that the apes matched the demonstrated target methods in the full-model condition of set 1 when all trials were considered (i.e., not just the first one). Furthermore, we found a significant correlation between task difficulty and the apes' matching scores in the intention condition.

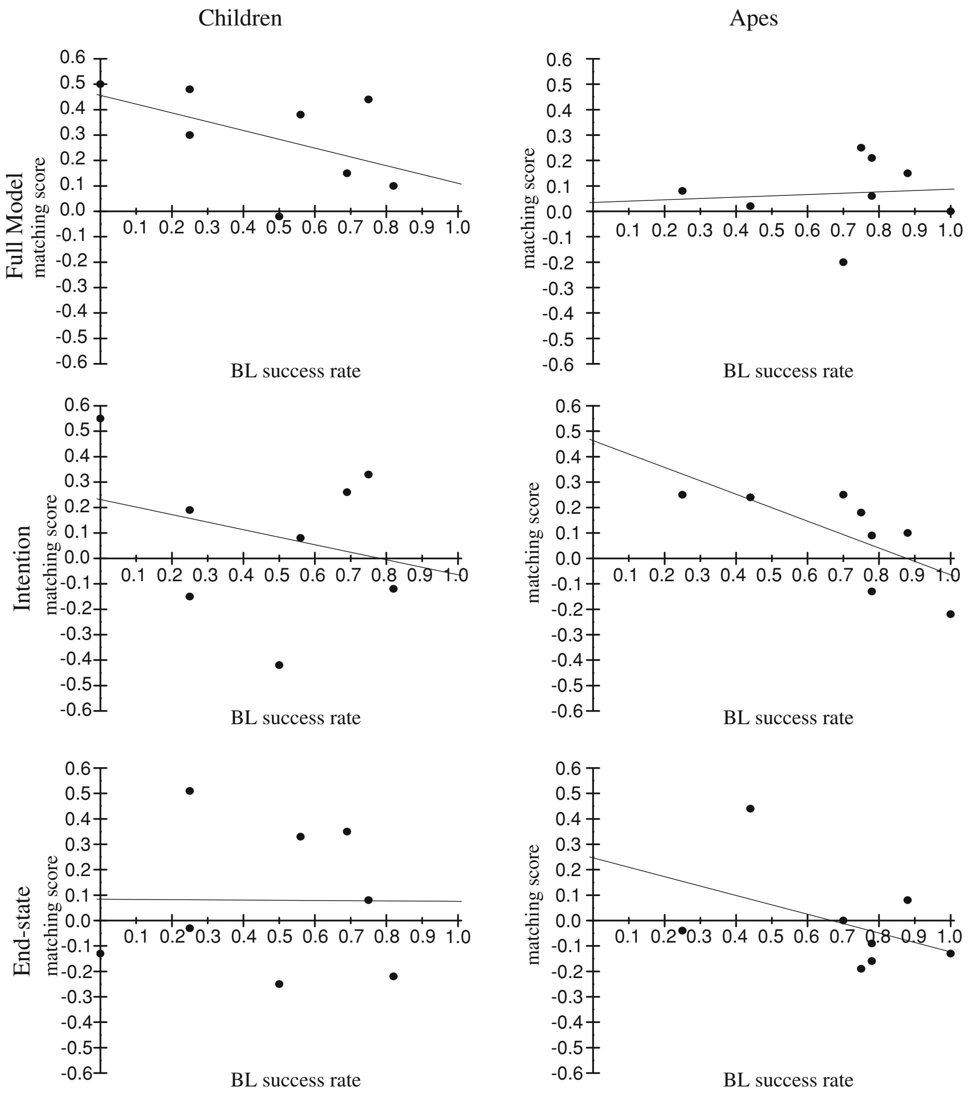

Fig. 5 Effect of difficulty level, as measured by baseline success rates, on matching scores (*left column* children, *right column* apes; *top row* full model, *middle row* intention, *bottom row* end state). Given that the analyses were nonparametric Spearman's correlations, the *lines of best fit* are included for illustrative purposes only

Children

For children, the full-model condition elicited instant copying in both sets of apparatuses. This is in concordance with the literature (see Call and Carpenter 2003, for a review; see also Call et al. 2005) and hence justifies our methods. The intention condition elicited copying in set 1, thus replicating previous results (e.g., Meltzoff 1995; Bellagamba and Tomasello 1999; Johnson et al. 2001) but not in set 2. This discrepancy between sets might be explained by the type of information provided in the two sets' intention condition. In the first set, the experimenter produced some information (albeit partial) about actions (e.g., hands slipping off the apparatus), whereas in the second set, the experimenter did not perform actions. This may mean that children largely focused on the demonstrator's actions to infer their goals. Removing the actions from the demonstration left children without their primary source of information, and this would explain the discrepancy in results between the first and the second set.

One could argue that these results call for a reinterpretation of Meltzoff's (1995) classic result in which infants completed the experimenter's goal based on witnessing a set of incomplete actions. In Meltzoff's study, children may not have copied the experimenter's goal but completed the actions they observed [for a critique of Meltzoff (1995) based on results copying, see Huang et al. (2002) and Huang and Charman (2005)]. However, this would not explain why children still produce the appropriate outcome (the goal) when the actions that have been observed differ from those produced due to size difference between the objects used by the experimenter and the objects used by the child (Meltzoff 2002). Moreover, there are now a number of studies showing that infants from the age of 12 months can reliably use goal information when copying others (Bellagamba and Tomasello 1999; Gergely et al. 2002;

Carpenter et al. 2005; Schwier et al. 2006). Thus, we suggest that although copying actions may be a primary route to infer goals, it is not the only one, and such results highlight that children are attuned to the actions of others.

Contrary to most previous studies (e.g., Bellagamba and Tomasello 1999; Want and Harris 2002; Call et al. 2005; see also some age classes in Huang et al. 2002; Tennie et al. 2006), we found some evidence that children copied results reliably in the absence of actions in the first set. This result is particularly important because our end-state condition incorporated neither the presentation of initial states before children witnessed the end state (e.g., Huang et al. 2002) nor displayed the apparatus transformation (from prior state to end state) by using a ghost condition (e.g. Tennie et al. 2006). It is unclear why we found no evidence of such copying in set 2. Set difficulty cannot account for this discrepancy, because both sets had a comparable level of difficulty. It is conceivable that the end states of the apparatuses in set 2 were less informative than those in set 1, thus making it harder for children to infer the transitional states and manipulations required to bring about the end result. However, we have no data to support this speculation.

The explanation we favor is based on the force required to operate the apparatuses in set 1 and set 2. In general, and just by chance (not by intentional design), it appears that the apparatuses in set 1 offered more resistance to change (i.e., required the application of greater force to produce a result) than those in set 2 (consistent with visual impression of Fig. 3c; albeit not significant). Although this did not influence success levels, it might have led subjects to sometimes try-first-and-then-stick-to the alternative target method, thereby diluting any present copying tendencies (i.e. "conservatism" of a path leading to success; compare Hrubesch et al. 2008). This view is reinforced by the fact that we found less matching in set 2's full model condition than in the corresponding condition of set 1. This means that it may have been easier to develop any target method, and not necessarily the demonstrated target method, in set 2 than in set 1. This may have had the strongest impact on the end-state condition, since children were prevented from witnessing the experimenter's actions that may have otherwise further helped to guide their responses.

Apes

Compared with the various positive effects observed in children after witnessing demonstrations, particularly demonstrations involving actions, we found no evidence that apes benefited from witnessing demonstrations of any kind in their first trial. Thus, ape copying was considerably weaker than children copying. However, we found a significant matching effect in the full-model condition of set 1 when averaging all trials. We take this to mean three things: First, as intended, our selection of tasks fell in the general copying zone of apes, even though it did not result in an immediate target matching. The effect of task difficulty on the intention matching scores of set 1 also points in this direction. Second, as expected, we found a difference between children and apes in success and matching performance. Children copied in the first trial and at high levels, whereas apes' matching was less robust and less pronounced. Third, it is unclear whether apes' matching was based on action copying, because the full-model condition provided observers with information about actions, goals and results (and this condition contained more results information than did the end-state condition). Additionally, the lack of significant matching in the first trial weakens the case for action copying in apes because information acquired through acting on the apparatus, not just observational information, may have at least contributed to the subjects' responses.

Comparisons of apes' and children's performance

One of our study's goals consisted of designing tasks with an appropriate level of difficulty. The observed positive correlation between task difficulty and the likelihood of matching in apes for the intention condition (together with the matching results mentioned above) partially fulfilled this goal by showing that the harder the task, the more likely the apes are to copy in the intention condition, which fits with some previous results (i.e. Horner and Whiten 2005). However, we found no such an effect for the full demonstration or the end-state condition in apes. For children, if there exists any correlation between task difficulty and likelihood of matching, it is likely to be in the full-model condition. Although this correlation was not statistically significant, it was in the same direction as the one of the apes: i.e., children copied more when the task was harder. Thus, we can speculate that apes and children may copy more when they do not know (as suggested by Horner and Whiten 2005). Moreover, this finding lends some support to the idea that humans evolved imitative and pedagogical skills that aid them in solving tasks that are too "opaque" to be solved individually (Gergely and Csibra 2005).

Our findings contribute to the growing body of evidence that, in comparison with human children, action copying is developed very little in great apes (e.g., Myowa-Yamakoshi and Matsuzawa 1999; Nagell et al. 1993; Call et al. 2005), but all our data also suggest that result copying is a more elusive phenomenon in great apes than previously thought (Tennie et al. 2006; Hopper et al. 2007, but see Hopper et al. 2008). If the great apes are not so skilled at copying new actions or results, how can one explain the recent experimental results on social transmission (see "Introduction")?

One possibility is that the main type of information that apes acquire from others concerns the locations where they are to apply their actions, or tool use. They may learn that inserting a stick in a particular hole produces a reward, but they may not learn about the action—or the result produced by the action. By design, all our tasks except for one (T-Pipe; note that this task was present in set 1—for which we found a significant copying effect) prevented apes from acquiring information about where to apply their actions, because the two alternative strategies to open each box were applied to the same location (local enhancement control). Note that other tasks that eliminated location as a source of information have also failed to produce positive results in great apes (e.g., Nagell et al. 1993; Call and Tomasello 1995; Tennie et al. 2006). Failing to control for this type of information might explain some earlier positive findings in great ape observational learning studies that have hitherto been interpreted as representing cases of true copying (e.g., Horner and Whiten 2005—though see also a subcondition that did control for local enhancement, similar to Horner et al. (2006)—both with more positive results; Whiten et al. 2005, 2007; Bonnie et al. 2007).

Finally, our results confirm previous research showing that children are more influenced by demonstration of actions than are apes (more so, in any case, than to results; see especially the performance differences in children between the intention condition of set 1 and set 2) and use them in various ways to learn to solve new problems very quickly. In contrast, apes showed matching only if all their trials were analyzed together, and they did not copy in the intention condition that lacked results information (but it should be noted that the ape intention condition was not as perfect as the children intention condition of set 1 because the ape intention condition contained some inaccurate demonstrations—due the difficulty of training apes to reliably perform target methods when their efforts met with no success). This finding is consistent with the view that apes do pay attention to results information first. The fact that apes did not copy in our end-state condition may be due to the fact that end states are only part of the whole results information transmitted in our full-model condition (in the latter, results included both initial and transformation states). It may also be due to the fact that our end-state condition lacked any goal information (whereas this was present in the full-model condition). Lastly, taken alone, our findings are at least consistent with the view that apes do, in fact, copy actions—but even if that were true, it seems likely that they only do so if action demonstrations are combined with other types of information (i.e., results, or goals and results). Which view is correct is an empirical question after all that awaits the right combination of methods. Tested under comparable conditions, children, unlike apes, copy actions more faithfully; they do so after little exposure (often one trial is sufficient, as in this study); with familiar or unfamiliar demonstrators; and for a variety of tasks. Therefore, we can say with confidence that there is a large difference—at least in degree (if not in kind)—between the observational learning mechanisms that humans and great apes normally use. Even a mere difference in degree alone may go a long way to explain the presence of cumulative culture in humans, as well as its virtual absence in apes (Tomasello 1999; Tennie et al. 2009).

Acknowledgments We thank Jenny Collard, Johannes Grossmann, Alenka Hribrar, Beate Kahl, Julia Keil, Roger Mundry, Raik Pieszek, Sebastian Schütte, and Hjalmar Turesson. We also thank the anonymous reviewers and the Max Planck Society. All experiments complied with current German laws.

References

Bellagamba F, Tomasello M (1999) Re-enacting intended acts: comparing 12- and 18-month olds. Infant Behav Dev 22:277–282

Bering JM, Bjorklund DF, Ragan P (2000) Deferred imitation of object-related actions in human-reared juvenile chimpanzees and orangutans. Dev Psychobiol 36:218–232

Bjorklund D, Yunger JL, Bering JM, Ragan P (2002) The generalization of deferred imitation in enculturated chimpanzees (*Pan troglodytes*). Anim Cogn 5:49–58

Bonnie KE, Horner V, Whiten A, de Waal FBM (2007) Spread of arbitrary conventions among chimpanzees: a controlled experiment. Proc R Soc Lond B Biol Sci 274:367–372

Call J, Carpenter M (2002) Three sources of information in social learning. In: Dautenhahn K, Nehaniv CL (eds) Imitation in animals and artifacts. MIT Press, Cambridge, pp 211–228

Call J, Carpenter M (2003) On imitation in apes and children. Infanc y Aprend 26:325–349

Call J, Tomasello M (1994) The social learning of tool use by orangutans (*Pongo pygmaeus*). Hum Evol 9:297–313

Call J, Tomasello M (1995) Use of social information in the problem solving of orangutans (*Pongo pygmaeus*) and human children (*Homo sapiens*). J Comp Psychol 109:308–320

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*). Anim Cogn 8:151–163

Carpenter M, Call J (2002) The chemistry of social learning. Dev Sci 5:22–24

Carpenter M, Call J, Tomasello M (2005) Twelve- and 18-month-olds copy actions in terms of goals. Dev Sci 8:F13–F20

Custance DM, Whiten A, Bard KA (1995) Can young chimpanzees (*Pan troglodytes*) imitate arbitrary actions? Hayes and Hayes (1952) revisited. Behaviour 132:837–859

Dawson BV, Foss BM (1965) Observational learning in budgerigars. Anim Behav 13:470–474

Furlong EE, Boose KJ, Boysen ST (2008) Raking it in: the impact of enculturation on chimpanzee tool use. Anim Cogn 11:83–97

Gergely G, Csibra G (2005) The social construction of the cultural mind: imitative learning as a mechanism of human pedagogy. Interact Stud 6:463–481

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

Hayes KJ, Hayes C (1951) The intellectual development of a home-raised chimpanzee. Proc Am Philos Soc 95:105–109

Herrmann E, Call J, Lloreda M, Hare B, Tomasello M (2007) Humans have evolved specialized skills of social cognition: the cultural intelligence hypothesis. Science 317:1360–1366

Hopper LM, Spiteri A, Lambeth SP, Schapiro SJ, Horner V, Whiten A (2007) Experimental studies of traditions and underlying transmission processes in chimpanzees. Anim Behav 73:1021–1032

Hopper LM, Lambeth SP, Schapiro SJ, Whiten A (2008) Observational learning in chimpanzees and children studied through "ghost" conditions. Proc R Soc Lond B Biol Sci 275:835–840

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

Horner V, Whiten A, Flynn E, de Waal FBM (2006) Faithful replication of foraging techniques along cultural transmission chains by chimpanzees and children. Proc Natl Acad Sci USA 103:13878–13883

Hrubesch C, Preuschoft S, van Schaik CP (2008) Skill mastery inhibits adoption of observed alternative solutions among chimpanzees (*Pan troglodytes*). Anim Cogn 12:209–216

Huang CT, Charman T (2005) Gradations of emulation learning in infants' imitation of actions on objects. J Exp Child Psychol 92:276–302

Huang CT, Heyes CM, Charman T (2002) Infants' behavioral reenactment of "failed attempts": exploring the roles of emulation learning, stimulus enhancement, and understanding of intentions. Dev Psychol 38:840–855

Johnson SC, Booth A, O'Hearn K (2001) Inferring the goals of a nonhuman agent. Cogn Dev 16:637–656

Meltzoff AN (1995) Understanding the intentions of others: re-enactment of intended acts by 18-month-old children. Dev Psychol 31:838–850

Meltzoff AN (2002) Elements of a developmental theory of imitation. In: Meltzoff AN, Prinz W (eds) The imitative mind: development, evolution, and brain bases. Cambridge University Press, New York, pp 19–41

Meltzoff AN, Moore MK (1977) Imitation of facial and manual gestures by human neonates. Science 198:75–78

Miles HL, Mitchell RW, Harper SE (1996) Simon says: the development of imitation in an enculturated orangutan. In: Russon AE, Bard KA, Parker ST (eds) Reaching into thought: the minds of the great apes. Cambridge University Press, New York, pp 278–299

Myowa-Yamakoshi M, Matsuzawa T (1999) Factors influencing imitation of manipulatory actions in chimpanzees (*Pan troglodytes*). J Comp Psychol 113:128–136

Myowa-Yamakoshi M, Matsuzawa T (2000) Imitation of intentional manipulatory actions in chimpanzees (*Pan troglodytes*). J Comp Psychol 114:381–391

Nagell K, Olguin R, Tomasello M (1993) Processes of social learning in the tool use of chimpanzees (*Pan troglodytes*) and human children (*Homo sapiens*). J Comp Psychol 107:174–186

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

Stoinski TS, Wrate JL, Ure N, Whiten A (2001) Imitative learning by captive western lowland gorillas (*Gorilla gorilla gorilla*) in a simulated food-processing task. J Comp Psychol 115:272–281

Tennie C, Call J, Tomasello M (2006) Push or pull: imitation vs. emulation in great apes and human children. Ethology 112:1159–1169

Tennie C, Call J, Tomasello M (2009) Ratcheting up the ratchet: on the evolution of cumulative culture. Philos Trans R Soc Lond B 364:2405–2415

Tennie C, Call J, Tomasello M (2010) Evidence for emulation in chimpanzees in social settings using the floating peanut task. PLoS ONE 5:e10544. doi:10.1371/journal.pone.0010544

Thompson DE, Russell J (2004) The ghost condition: imitation versus emulation in young children's observational learning. Dev Psychol 40:882–889

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

Tomasello M, Davis-Dasilva M, Camak L, Bard KA (1987) Observational learning of tool use by young chimpanzees. Hum Evol 2:175–183

Tomasello M, Savage-Rumbaugh S, Kruger AC (1993) Imitative learning of actions on objects by children, chimpanzees, and enculturated chimpanzees. Child Dev 64:1688–1705

Want SC, Harris PL (2002) How do children ape? Applying concepts from the study of non-human primates to the developmental study of "imitation" in children. Dev Sci 5:1–41

Whiten A, Custance DM, Gómez J-C, Teixidor P, Bard KA (1996) Imitative learning of artificial fruit processing in children (*Homo sapiens*) and chimpanzees (*Pan troglodytes*). J Comp Psychol 110:3–14

Whiten A, Goodall J, McGrew WC, Nishida T, Reynolds V, Sugiyama Y, Tutin CEG, Wrangham RW, Boesch C (1999) Cultures in chimpanzees. Nature 399:682–685

Whiten A, Horner V, de Waal FBM (2005) Conformity to cultural norms of tool use in chimpanzees. Nature 437:737–740

Whiten A, Spiteri A, Horner V, Bonnie KE, Lambeth SP, Schapiro SJ, de Waal FBM (2007) Transmission of multiple traditions within and between chimpanzee groups. Curr Biol 17:1038–1043