Journal of Comparative Psychology
2003, Vol. 117, No. 3, 325–336

0735-7036/03/$12.00 DOI: 10.1037/0735-7036.117.3.325

Do Humans Ape? Or Do Apes Human? Imitation and Intention in Humans (*Homo sapiens*) and Other Animals

Alexandra C. Horowitz
University of California, San Diego

A. Whiten, D. M. Custance, J.-C. Gomez, P. Teixidor, and K. A. Bard (1996) tested chimpanzees' (*Pan troglodytes*) and human children's (*Homo sapiens*) skills at imitation with a 2-action test on an "artificial fruit." Chimpanzees imitated to a restricted degree; children were more thoroughly imitative. Such results prompted some to assert that the difference in imitation indicates a difference in the subjects' understanding of the intentions of the demonstrator (M. Tomasello, 1996). In this experiment, 37 adult human subjects were tested with the artificial fruit. Far from being perfect imitators, the adults were less imitative than the children. These results cast doubt on the inference from imitative performance to an ability to understand others' intentions. The results also demonstrate how any test of imitation requires a control group and attention to the level of behavioral analysis.

Recent investigation into the extent of nonhuman animals' imitative abilities—itself part of the greater field of comparative human and animal cognition—has provided a testing ground for extrapolating from anecdotal reports to experimental investigation. Although many animals, especially primates, are said to perform "true" imitative actions (Heyes, 1996) in the wild, experimental studies have produced varying results. In laboratory investigation of imitation, many researchers have come to look more closely at what the animals do and to characterize seemingly imitative behavior as, instead, "emulation" or "stimulus enhancement" (Tomasello & Call, 1997; Whiten & Ham, 1992). Some have questioned whether any true imitation is observed in wild or otherwise unenculturated animals at all (Tomasello & Call, 1997).

Whiten, Custance, Gomez, Teixidor, and Bard (1996) performed a series of experiments designed to address both the challenges to claims of imitation in nonhuman animals and the ambiguities about what type of imitation, if any, is seen. Chimpanzees (mean age = 4.5 years) and human preschoolers (2-, 3-, and 4-year-olds) were tested on their ability to learn by observation to open a specially designed box. The manipulable box was termed an *artificial fruit* because its construction encourages natural primate motor movements in investigative handling. In the wild, it is often in the processing of food—getting to or cleaning food—that seemingly imitative behavior has been reported; the box apparatus was designed with such ecological considerations in mind. Additionally, the design of the fruit box provides an opportunity to follow novel and discrete actions (Whiten, 1998; Whiten et al., 1996). Its design allows for clear distinction between the levels of imitation that have been suggested.

In the presence of the subject, the experimenter demonstrated the unlocking and opening of this box, with one of two actions on latches that lock the box (see Table 1). The two-action design of the test controls for other interpretations of the behavior, such as stimulus enhancement and emulation (Dawson & Foss, 1965; Whiten, 1998), by limiting the analysis of the subject's behavior to a determination of which of two alternative actions the subject performed. After opening the box, the experimenter reached in and removed the reward for successful action: a treat (fruit, for chimpanzees; candy, for the children) that was secreted inside. The variable investigated, then, was whether the subject imitated the experimenter's behavior in opening the artificial fruit or whether he or she performed the contrasting action.

The experimenters found evidence of imitation in both chimpanzees and children (Whiten et al., 1996). The children's actions showed notably more conformity to the demonstrations than did the chimpanzees' actions. Specifically, a trend toward more faithful imitation could be seen from chimpanzees to the 2-year-olds, 3-year-olds, and 4-year-olds. Overall, Whiten et al. (1996) showed a significant effect of having seen a demonstration in each subject group on at least one component of the lock apparatuses.

Besides testing for observational learning, this experiment also allows tentative conclusions to be drawn about the metacognitive skill of its subjects. As defined in the literature, true imitation would require taking what Dennett (1987) has called the "intentional stance": understanding others' behavior as directed, as instantiation of beliefs and desires. Imitation studies have been pursued in the shadow of a search for understanding of intentions. The experimenters do not make claims about the intentional understanding of their subjects; it is the interpreters of their experiment who have. The stronger imitation by Whiten et al.'s (1996) human subjects led Tomasello (1996) to conclude that for chimpanzees, "the intentional states of the demonstrator . . . [are] either not perceived or irrelevant," whereas for humans, "the goal or intention of the demonstrator is a central part of what they perceive" in others' behavior (p. 331).

In the present study, I tested adult human subjects on Whiten et al.'s (1996) artificial fruit design. The subjects were normal, high-

Alexandra C. Horowitz, Department of Cognitive Science, University of California, San Diego.

This study was supported by a National Science Foundation training grant. I am grateful to Andrew Whiten for the use of the test equipment and for comments on earlier versions of this article.

Correspondence concerning this article should be addressed to Alexandra C. Horowitz, c/o 55 Central Park West #13 D, New York, New York 10023. E-mail: ahorowitz@crl.ucsd.edu

Table 1
Operations on the Locks

Bolts
The procedure follows that of Whiten et al. (1996), contrasting two unlocking actions on the bolts lock.
Twist technique. The bolts are removed one after the other from their rings by grasping each bolt with one hand and twisting it clockwise while pulling forward (the pulling is synchronized with the twisting action to give the impression that the turning allows removal of the bolts). Once the bolts are removed, the rings on the box lid are grasped and the lid opened. Each bolt is twisted approximately 6 times. (The bolts are not threaded, so the twist action is entirely unnecessary for removal from the rings.)
Poke technique. The bolts are removed sequentially by poking each bolt in turn away from the actor, with the thumb or index finger of one hand. The bolts are then gathered in one hand and the lid opened as before.
T lock
On this lock, contrasting actions are performed on each rod. On each trial, one action on the pin is matched with one action on the handle.
Pin: Turn technique. The perpendicular bar of the pin is grasped with the fingers of one hand, and the pin is turned clockwise an average of 4 times. (Rotation is not necessary to remove the pin.) The pin is removed by closing the fingers around the T of the pin and pulling.
Pin: Spin technique. With the side of the second or third finger, the T bar is rotated an average of 4 times. The pin is then removed as before.
Handle: Turn technique. By use of the perpendicular member of the rod, the rod is turned 90° to the right, so that the projecting lip of the handle is no longer over the box lid. The box is opened by grasping the rings and pulling up on the lid.
Handle: Pull technique. By use of the perpendicular member, the rod is lifted straight up and out of the barrel. The box is opened as before.

functioning humans who were not expected to have any motor or conceptual difficulties with the task. The human ability to imitate is unquestioned, and adult humans should demonstrate the highest level of imitation of all tested groups. As a result, this experiment serves as a potentially interesting test of any inference from experimental behavior to the subjects' minds: Will the subjects perform at a level that we have prior reason to expect that they can, or will they not succeed at a task that they should be able to easily perform? These subjects ought to define what the most imitative—and most aware of the demonstrator's intention—performance at the task could be.

Method

Subjects

The subjects were 37 adult humans (*Homo sapiens*) from the subject pool run by the Psychology and Cognitive Science Departments at the University of California, San Diego. Participants earned experimental credit for an undergraduate class in the Psychology or Cognitive Science Departments. Age range was from 18 to 25 years old ($M = 21.5$ years), and all subjects had 1 or more years of higher education. Seventeen subjects had participated in at least one other psychology experiment in the previous year. All but 2 subjects were right-handed. They were selected to contrast markedly in age with the human subjects in the original experiment by Whiten et al. (1996)—2-, 3-, and 4- year-olds—who were themselves chosen as comparisons to the chimpanzee subjects.

Task and Apparatus

The artificial fruit apparatus was acquired from the original experimenters and was used in the same manner as it was used with Whiten et al.'s (1996) subject groups (see *Procedure*, below). The artificial fruit is a Plexiglas box (21 × 17 × 14 cm) with a hinged lid and two lock mechanisms (a *T lock*—originally called a *barrel latch*—and a *bolts lock*) securing the lid (see Whiten et al., 1996, for an image of the artificial fruit). In most parts of the experiment, the locks are manipulated separately: One mechanism is dismantled and removed by the experimenter, and the remaining mechanism serves as the focal element.

The bolts lock consists of two smooth plastic rods, each 11.5 cm long. Each rod bridges two sets of barrel rings: one on the lid of the box and the other on the lip of the box. Both bolts need be removed to allow the lid to swing freely open. The T lock consists of a barrel attached to the side of the box and has two subcomponent rods, which need to be removed in sequence. Each rod has a perpendicular member on the business end for grasping. The first rod in order of removal is a 6.5-cm-long screw pin that fits horizontally into a hole in the barrel of the lock and into a hole in the second rod. The latter smooth rod (11.5 cm long and thicker than the first rod) is a handle that fits vertically into the barrel and has a perpendicular extension near its head that sits over the lip of the box lid. It is this element that keeps the box closed.

Procedure

The procedure used with the adult human subjects closely mirrored that used by Whiten et al. (1996) with the chimpanzees and the children. Subjects entered a small experimental room and were seated at a long table immediately next to the experimenter. The instruction given to subjects was minimal; this matched the procedure in Whiten et al. With the chimpanzee group, the experimenter was silent throughout. With the child subjects, the prompt "Do you see the sweet inside here?" before the experimenter opened the box was the only substantive verbal exchange. The adult subjects were told only that they were participating in a three-part behavior task: In the first part of the experiment, there would be no verbal instruction on the task; if they had questions, they were to hold off on asking them until after the experiment had concluded and to simply proceed however came naturally.

The experimenter produced the box with one lock apparatus in place and a Hershey's chocolate candy (taking the place of a fruit for the chimpanzees) planted inside. In plain view of the subject, the experimenter opened the box by one of the methods described in Table 1, removed the chocolate, and showed the subject the candy. The box was reconstructed out of view of the subject. Subsequently, no more than 30 s later, the box was replaced, relocked, in front of the subject. Although the subjects were not instructed as to the nature of the task, they were under a 2-min time limit to act on the box. At that time, when the subject had removed the candy or when the subject indicated to the experimenter that he or she was finished (whichever happened first), the box was taken away and reconstructed using the other lock apparatus. The demonstration condition was then repeated using the remaining lock and one of the opening methods for that lock (see Table 1). The entire procedure was videotaped and blind coded by 2 observers, as described more fully below. After the experiment but before being debriefed, subjects were given a questionnaire that asked them their impressions of the content of the task they had just completed.

Baseline Data

Diverging from Whiten et al.'s (1996) procedure, I also tested a separate group of baseline subjects to ascertain what actions a subject would perform on the artificial fruit in the absence of any demonstration (Stoinski, Wrate, Ure, & Whiten, 2001, also used a control group with gorilla subjects). Eleven naive subjects (mean age = 21.1 years; range = 19–27 years) were presented with the box without any demonstration. A videotape recorder was turned on before the subject entered the room. On entry, these subjects were given the same instruction and were presented with the artificial fruit, with the comment "You can have whatever's in the box." After this trial, these subjects saw two demonstration trials—one on each lock, in the manner described above—and were given the artificial fruit to act on after each one.

Analysis

The data are first considered in a manner directly comparable with the data from the experiment by Whiten et al. (1996). Then, additional analyses are described, which bear on the interpretation of the results from these imitation experiments.

Whiten Analysis

Independent observer (IO) scores. In the experiment by Whiten et al. (1996), the two main types of analysis were IO scores and action frequency counts. The IO scores were gathered by 2 individuals not involved with the experimentation, who were familiarized with the demonstration options on each lock apparatus. They viewed the videotape of the subjects' actions and made a judgment as to which of the two demonstrations the subject had seen. This same procedure was followed with the adult human data from the current study. After viewing a subject's performance, the observer chose one of the two actions he or she thought was seen by the subject and then rated his or her confidence about this on a scale that was calibrated from 1 to 7 (a rating of 1 represents highest confidence that Technique A was seen, 7 represents highest confidence that Technique B was seen, and 4 represents no confidence on the scorers' part either way). The scores given by the 2 IOs in this experiment were highly similar: On the bolts lock, there was 94% agreement as to which action was seen; on the T lock, 89%. Similar figures (ranging from 84% to 100%, across subjects and conditions) were attained in the original experiment. The median score and interquartile ranges for all subjects in each group (chimpanzees, 2-year-olds, 3-year-olds, 4-year-olds, and adults) are presented graphically later in the article.

Given the limitations of their small sample, Whiten et al. (1996) chose to represent the data with a comparison of each group's intragroup difference using Mann–Whitney U tests to determine the p value (one-tailed). That is, within each group, was there a difference in performance for having seen Technique A as opposed to Technique B? Intragroup adult data are also tabulated in the current study. In addition, the larger sample size of the present experiment enables intergroup comparisons by using an unpaired t and two-tailed p.

Action frequencies. Whiten et al. (1996) tallied the number of various actions performed on each lock. Results from the present study are shown in figures that appear later in the article for the main actions demonstrated on the box: the number of pokes and twists on the bolts, spins and turns of the pin, and turns and pulls of the handle. There is an additional figure for the action of pulling the bolt out.

Latency to open. Whiten et al. (1996) measured the time taken to open the box for each group. Comparable time figures are presented in this article. Adult subjects often continued acting on the artificial fruit after opening it, by, for instance, closing the box or restocking it with chocolate. The latency to finish the task is thus also noted in the current study.

Extended Analysis

Whiten et al.'s (1996) analysis is a measure of the degree to which the subjects' behavior matches the demonstrated actions. I conducted a complementary analysis of the baseline subjects' actions that did not restrict description to the demonstration categories (see Voelkl & Huber, 2000, for an example of the usefulness of this approach). These subjects were recorded acting on the artificial fruit without seeing a demonstration and then again after seeing a demonstration. Between-group and within-group comparisons are done.

As these subjects did not (initially) witness a demonstration, an IO score of imitation is inappropriate. Instead, the parameters in the Appendix

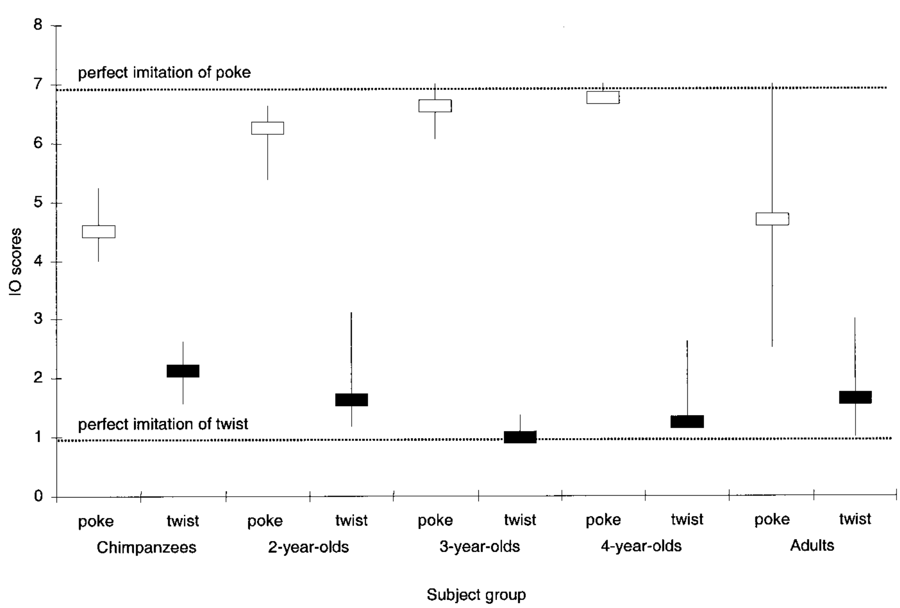


Figure 1. Independent observer (IO) scores of the subjects' actions on the bolts locks. Observers assessed whether the subject's behavior looked more like he or she had seen a demonstration of the poke technique (open bars) or the twist technique (solid bars). The technique that was demonstrated (poke or twist) is shown on the x-axis. Dashed lines indicate the score expected if the subjects imitated the demonstration perfectly. Chimpanzee and child data are from Whiten et al. (1996). Interquartile ranges are shown.

Table 2
Intergroup Comparisons of Independent Observer Scores on the Bolts and T Locks

	Subject groups compared			
Lock and action	Chimpanzees vs. adults	2-year-olds vs. adults	3-year-olds vs. adults	4-year-olds vs. adults
Bolts				
Poke	*ns*	*ns*	*ns* ($p > .15$)	*ns* ($p < .065$)
Twist	*ns*	*ns* ($p > .15$)	*ns* ($p > .15$)	*ns*
T				
Handle pull	*ns*	*ns*	*ns* ($p < .18$)	*ns*
Handle turn	*ns* ($p > .054$)	*ns*	*ns* ($p < .17$)	*ns* ($p < .17$)

Note. None of the comparisons were significant ($p > .05$, Mann–Whitney U test). Near-significant ($< .20$) p values are shown in parentheses. Nonadult data are from Whiten et al. (1996). Data for the pin techniques used with the T lock were not available from the original experiment.

are used to identify and score actions. The percentage of subjects who do each action can be compared across groups by using an unpaired t and two-tailed p.

Results

Bolts Lock

IO Scores

Figure 1 is a box-and-whiskers plot of the distribution of IO scores for all sample groups (from Whiten et al., 1996, and the current study) and includes comparison with data that would represent "perfect" imitative data. Although the tendency of the original subjects to imitate steadily increases from chimpanzee to child and through development, adult humans showed a reversion in imitative performance from the oldest children in terms of central tendency. In addition, the adult human data were much more variable than any other subject group's data.

Table 2 reconfigures the Figure 1 data as comparisons across subject groups of the subjects' allegiance to, or imitation of, the demonstrated action. There were no significant differences. The poke trials provided the only nearly significant difference between groups, $t(22) = 1.95$, $p < .065$, between the adult group and the 4-year-olds.

Table 3 compares intragroup performance on the locks for all subject groups. Considering behavior as scored by the IOs, there was a significant effect of the bolts lock demonstration on the adult human subjects' behavior ($p < .001$). In fact, all subject groups were significantly ($p < .05$) more likely to do the action demonstrated than the opposing action on the bolts lock: All subjects imitated.

Action Frequencies

The original experiment included action frequency tallies to complement the IO scores and to represent the subjects' actions directly. Figure 2A shows the number of pokes made on the bolt lock for each subject group for each condition (either the poke or twist technique demonstrated); Figure 2B shows the number of twists. A nondemonstrated move, pulling the bolt out (pull), is also included (see Figure 2C). Chimpanzee and child data show that these subject groups often far exceeded the number of moves of any type of action that were shown by the demonstration (see Figures 3 and 4 in Whiten et al., 1996). Whereas the most bolt twists that any subject saw in a demonstration was 16, 1 chimpanzee made 48 twists, one 4-year-old turned the bolts 86 times, and one 3-year-old turned the bolts 161 times. Adults never exceeded the number of pokes or twists demonstrated. However, a novel move, "pull," was seen in over half the adult subjects ($n = 20$). Although a twist of the bolts involves pulling the bolt out, a straight pull, which these subjects did, was never demonstrated. There was no significant difference (in either direction) in the likelihood of pull being performed, $t(33) = 0.54$, $p > .50$, whether the twist technique, which incorporates a pulling motion, or the poke technique had been demonstrated.

Table 3
Intragroup Comparisons of Independent Observer Scores on the Bolts and T Locks

	Subject group				
Lock and action	Chimpanzees	2-year-olds	3-year-olds	4-year-olds	Adults
Bolts	$p < .05$	$p < .05$	$p < .02$	$p < .02$	$p < .001$
T					
Handle	*ns*	$p < .05$	$p < .02$	$p < .02$	$p < .001$
Pin	*ns*	*ns*	*ns*	*ns*	*ns*

Note. Nonadult data are from Whiten et al. (1996).

Latency to Open

The final analysis by Whiten et al. (1996) on the bolts lock came in the calculation of the latency to open the box: Chimpanzees took 38 s, 2-year-olds took 77 s, 3-year-olds took 17 s, and 4-year-olds took 23 s (median times presented). Adults took 14 s on average.

The latency figure defines the task as "opening the box." The experiment with adults provided an extra parameter: Overall, 62% (23 out of 37) of the subjects did not define the task that way. They reclosed and relocked the box. If the task is instead defined by the subject's indication that he or she has finished acting on the box (by, for instance, sitting back in the chair, sliding the artificial fruit to the experimenter, or pausing significantly after opening the artificial fruit), the mean latency to complete the task rises to 19 s for adults.

T Lock

IO Scores

The IO scores for the adult subjects' action on the pin part of the T lock are shown in Figure 3 (no figure or raw data were given in the original experiment for comparison). Adult subjects performed similarly in both conditions: They turned the pin, regardless of the demonstration. Whiten et al. (1996) also reported that the spin technique was rarely used by any subject. Their results showed a nonsignificant difference for the technique observed to be performed in each of their subject groups (*p*s ranged from .17 to .44; Whiten et al., 1996). Adults also did not show an effect of demonstration (see Table 3).

IO scores for the handle element of the T lock for all subject groups are displayed in Figure 4. Whereas the children tended closer to perfect imitation with age, the adults were more highly variable and tended less toward imitation. Intergroup (adults vs. each other group, in turn) comparisons reveal no significant differences (see Table 2). There are some notable trends toward differences in imitative behavior between the adult and other subject groups: in the pulls trial, between adults and the highly imitative 3-year-olds, $t(19) = 1.41$, $p < .18$, and in the turn trials, between the adults and both 3- and 4-year-olds ($p < .17$) and chimpanzees ($p < .06$). The chimpanzees tended to pull, regardless of the demonstration.

Table 3 lists the intragroup significant differences for the handle and the pin trials. In these cases, all subject groups but the chimpanzees were imitative on the handle; none were imitative on the pin.

Action Frequencies

The action frequency tally is applicable to the pin component of the T lock. The spin technique was quite rare among all subject groups; however, general rotation of the pin was frequent. Human adults also often pulled the pin directly out. The total number of rotations by the chimpanzees varied from 3 to 12; for the children, the range was 16 to 209 (Whiten et al., 1996); and the adults rotated 0 to 8 times, with one outlier at 35. The human adults and chimpanzees performed quite similarly on this metric: The number of rotations the chimpanzees made was always below the number of rotations they saw demonstrated (4 trials of 4 = 16 rotations), and the number of rotations the adults made was below the demonstration number (1 trial of 4 = 4 rotations) 86% of the time. Frequency scores on the handle are not more informative than a simple tally of which acts are produced because the handle was unlikely to be either turned or pulled more than once.

Latency to Open

The median times to open the box using the T lock were reported by Whiten et al. (1996) as follows: chimpanzees, 47 s; 2-year-olds, 70 s; 3-year-olds, 124 s; and 4-year-olds, 91 s. The mean time for adult humans in the present study was 15 s. The extended latency for adults—as introduced above—was 19 s.

Extended Analysis Results

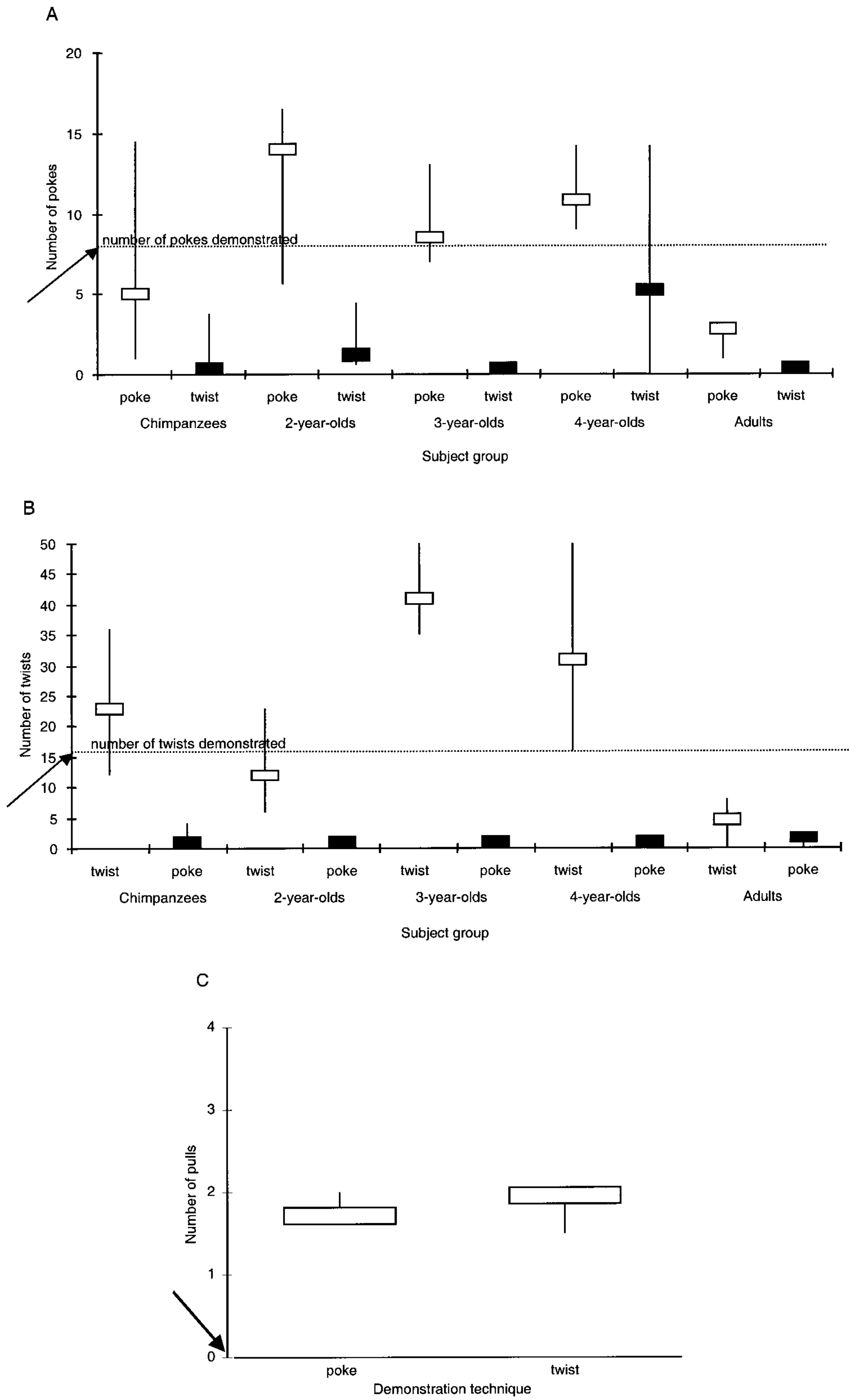
In the baseline condition, subjects were given neither a demonstration nor instruction but were handed a box with one or two locks fastened. In every case, the subject opted to open the box. Each subject moved within the first 15 s to attempt to undo one of the locks. On the bolts lock, the great majority of subjects (10 out of 11, 90%) pulled the rods out of the rings (the other subject pushed the bolts through). On the T lock, 4 out of 6 (67%) subjects operated the pin by turning it one or two times; the other subjects took the pin directly out. Actions on the handle were equally distributed between pulling it up and out ($n = 2$), turning the handle to the left ($n = 2$), and pulling it up but not out ($n = 2$).

Table 4 compares the performance of the baseline and the experimental groups (nonparametric Mann–Whitney *U* test). Did the group that saw a demonstration perform the demonstrated actions more often than the naive group? On the bolts lock, the poke technique was done significantly more often by the group that had seen the poke technique in demonstration ($p < .003$). The twist technique, however, was not ($p > .27$). On the T lock, the spin pin technique was done significantly more often by the demonstration group ($p < .024$), but the turn pin technique was not. Nor was there a significant difference between the baseline and experimental subjects' action on the handle.

Table 5 compares the baseline group's actions when simply handed the box with their actions in the second half of their trials after subsequent demonstrations in the manner of the original experimental group. Only one technique (spin pin) was seen significantly more often after seeing a demonstration than before. (However, in this case, this statistic is revealed to be wanting because the 1 subject who did spin the pin had been in a group that saw the turn technique demonstrated, not the spin technique.) Otherwise, the subjects' behavior was statistically unchanged.

Discussion

By the IO scores, there was a visible difference for each action between all subject groups. On the bolts lock, for instance, the chimpanzees and the children both exhibited some imitation. There is a clear increase in imitation from chimpanzee to human child to older child, but the trend is discontinuous when adults are included (see Figure 1). The 3-, 4-, and sometimes 2-year-olds tended toward perfect imitation. The adult data do not represent a natural extension to this trend. The ratings for those who saw the poke technique demonstrated were quite variable: The subjects' actions


A
20
15
10
5
0
Number of pokes
number of pokes demonstrated
poke
twist
poke
twist
poke
twist
poke
twist
poke
twist
Chimpanzees
2-year-olds
3-year-olds
4-year-olds
Adults
Subject group
B
50
45
40
35
30
25
20
15
10
5
0
Number of twists
number of twists demonstrated
twist
poke
twist
poke
twist
poke
twist
poke
twist
poke
Chimpanzees
2-year-olds
3-year-olds
4-year-olds
Adults
Subject group
C
4
3
2
1
0
Number of pulls
poke
twist
Demonstration technique

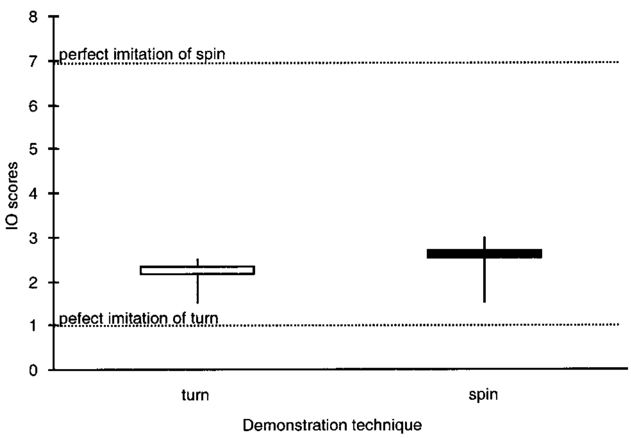


Figure 3. Independent observer (IO) scores of the subjects' actions on the pin component of the T lock for adult subjects only. Observers assessed whether the subject's behavior looked more like he or she had seen the turn technique (open bar) or the spin technique (solid bar). The technique that was demonstrated is shown on the *x*-axis. Dashed lines indicate the score expected if the subjects imitated the demonstration perfectly. Interquartile ranges are shown.

were equivocal indicators of which demonstration they saw. The subjects who saw the twist technique demonstrated were less dispersed—none was mistaken for a subject who saw the poke technique—but the data look more like the chimpanzee data than like the children's data (this difference was not statistically significant).

The IO scores on the pin part of the T lock (see Figure 3 for the adults) were virtually identically distributed whether the subject saw the spin or turn technique demonstrated. In fact, all subject groups performed similarly in both conditions. Turning the pin appeared to be the preferred option generally, as the significance comparisons show, regardless of the demonstration and across subject group. That anyone spun the pin at all does imply that a few (3 adult) subjects did follow the specifics of the demonstration, although the IO scoring does not reveal this trend. This lack of salience of the spin technique prompted later experimenters with the artificial fruit to contrast demonstrations of turning the pin with a straight pull of the pin instead (Stoinski et al., 2001): These authors found imitation in their subjects (gorillas).

On the handle apparatus, a significant effect was observed in all children between the two demonstration groups: Those who saw the handle pulled generally pulled, and those who saw the handle turned generally turned (see Figure 4). The chimpanzee evidence was not significantly biased toward imitation: Instead, the chimpanzees were scored as though they had seen pull demonstrated nearly every time. This implies that the chimpanzees were generally pulling. The adult evidence falls somewhere between the two groups. About half of each group looked imitative—similar to the discrete data of the children—but the other half was spread over both actions—similar to the chimpanzee data.

The intragroup significance numbers (see Table 3) seem to smooth out any potential differences between the groups. All groups showed intragroup significance between actions on the bolts, all but the chimpanzees showed significance on the handle, and no group showed significant intragroup differences on the pin. In other words, this measure essentially takes away enough information to almost equate all subjects' performances. Questions are justified about what the cross-group comparison of intragroup significance can rightfully show. It is saved from being a dangerously misleading metric by additional intergroup comparisons (see Table 2). They confirm the similarities across groups. In addition, they show a near-significant difference between adults and 4-year-olds in one case. This would be a surprising result if the adults were mere perfecters of childhood behavior.

Although action frequency figures can be compared (see Figure 2), this information has more to say about general characteristics of the subject groups as actors on objects than about any group's particular skill at imitation. The figures do suggest that children might copy an action to a high degree (twisting the bolts 161 times or turning the pin 209 times). This trend of overdoing the action is seen in all child subject groups. By contrast, the human adults and chimpanzees generally did not imitate exaggeratedly. (One adult subject did reach a group high of 35 pin turns; 1 chimpanzee did twist the bolts 48 times.) In fact, pulling the bolt is the only measure on which the performance number (for adults) exceeds that of the demonstration because it was not specifically demonstrated.

Latency figures were included because of the ostensive connection between imitation and speed of action: Some suggest that quick action indicates learning by imitation whereas slow response indicates a problem-solving technique other than imitation (Galef, 1990). Adult subjects' latency to open the box by either lock was shorter than any of the other subject groups. The longest latency was for the 2- and 3-year-olds. There was no obvious trend either within humans or cross-primate groups. The figures may indicate

Figure 2 (opposite). Action frequency tallies. A: The median number of pokes on bolts locks performed. The open bars indicate that subjects saw the poke technique demonstrated; the solid bars indicate that subjects saw the twist technique demonstrated. The arrow and the dashed line indicate the actual number of pokes demonstrated (normalized for the adult subjects; in Whiten et al., 1996, each subject saw each demonstration four times, whereas adults in this study had only one trial). Nonadult data are from Whiten et al. (1996). B: The median number of twists on bolts locks performed. The open bars indicate that subjects saw the twist technique demonstrated; the solid bars indicate that subjects saw the poke technique demonstrated. The arrow and the dashed line indicate the actual number of twists demonstrated (normalized for the adult subjects; in Whiten et al., 1996, each subject saw each demonstration four times, whereas adults in this study had only one trial). Nonadult data are from Whiten et al. (1996). C: Number of pulls of bolts locks for adults only. The arrow indicates that no demonstration of pull was given. Interquartile ranges are shown.

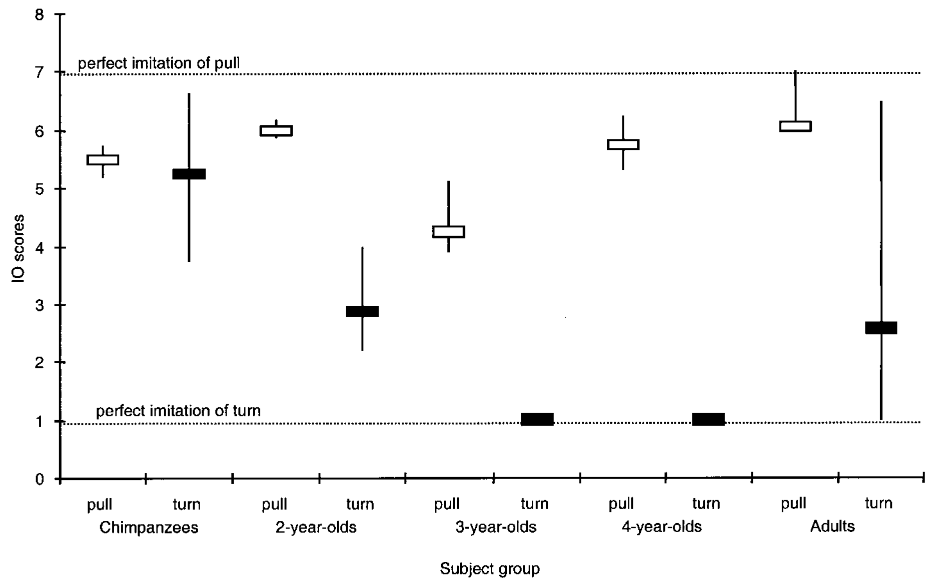


Figure 4. Independent observer (IO) scores of the subjects' actions on the handle component of the T lock. Observers assessed whether the subject's behavior looked more like he or she had seen the pull technique (open bars) or the turn technique (solid bars). The technique demonstrated is shown on the *x*-axis. Dashed lines indicate the score expected if the subjects imitated the demonstration perfectly. Nonadult data are from Whiten et al. (1996). Interquartile ranges are shown.

more about the acting styles of each subject group than about imitative intent.

The different latency-to-finish statistic reveals an ambiguity for the human adults in what the task was defined as being. A postexperiment questionnaire found that 32% of the subjects, when asked what they thought they were supposed to do, believed that the task was something other than "imitate the experimenter," such as "open the box (efficiently)" or "retrieve the chocolate." Many ($n = 23$) adult subjects did not finish until they reclosed the box—even though they had not witnessed the box being reconstructed—and a handful ($n = 5$) also replaced the demonstration chocolate with a new chocolate. This raises the question whether the desideratum as well as the scope of the demonstration was apparent to the child or chimpanzee subjects. This will be addressed in the discussion of varying levels of analysis below.

Baseline Results of Note

If the two-action design is a compelling test to determine imitation, one should expect a difference in behavior between subjects in a group that saw a demonstration and subjects in a naive group. The baseline adult human subjects pulled, pushed, and twisted the bolts; they turned the pin and pulled the handle out or turned it to the side. Many of these actions were demonstration actions. In overall performance, some baseline subjects acted on the locks very similarly to subjects who had seen a demonstration. As a result, there is good reason to be skeptical that what the adults in the demonstration conditions were doing is best described as imitation.

When the baseline group was given a demonstration, they did not then imitate faithfully. No subject imitated the demonstrated

Table 4
Comparison of Baseline Subjects' and Experimental Subjects' Performance

Lock and action	Baseline subjects[a]	Experimental subjects[a]	Significant
Bolts			
Any pulls	10/11	20/37	no
Any twists[b]	1/11	12/37	no
Any pokes[b]	0/11	9/37	yes[c]
T			
Turn pin[b]	4/6	27/37	no
Spin pin[b]	0/6	3/37	yes[d]
Pull pin straight out	2/6	11/37	no
Pull handle[b]	2/6	17/37	no
Turn handle[b]	2/6	11/37	no

[a] Values shown are the number of subjects who performed the action/total number of subjects. [b] Demonstrated action. [c] $p < .003$. [d] $p < .024$.

Table 5
Comparison of Baseline Subjects' Performance Before and After Seeing Demonstration

Lock and action	Before[a]	After[a]	Significant
Bolts			
Any pulls	10/11	5/9	no
Any twists[b]	1/11	3/9	no
Any pokes[b]	0/11	0/9	no
T			
Turn pin[b]	4/6	7/9	no
Spin pin[b]	0/6	1/9	yes[c]
Pull pin straight out	2/6	2/9	no
Pull handle[b]	2/6	4/9	no
Turn handle[b]	2/6	4/9	no

[a] Values shown are the number of subjects who performed the action/total number of subjects. [b] Demonstrated action. [c] $p < .034$.

technique perfectly. There was some imitation: The poke bolt and spin pin techniques were only seen after the demonstration (given the small baseline subject group, intragroup comparisons do not have the power to detect statistical significance). However, the other four demonstrated actions were not performed more frequently by the baseline group when they were given a demonstration. Seeing an example of the artificial fruit being opened did not persuade many subjects to abandon their naive fruit-opening strategies to imitate the experimenter.

This is particularly interesting because the majority of the baseline subjects nonetheless indicated on the postexperiment questionnaire that they believed that they were supposed to (do some version of) "follow what the experimenter did" in the latter part of the trial. But experimenter observations reveal the oddness of this claim: In a few cases, the subject did not pay attention to the experimenter's deliberate demonstration of the unlocking procedure. Subjects looked away, took the opportunity to clean their glasses, and so on. Evidently, the affordances of the box were so easily fully explored in their own handling of it that they did not feel the need to attend closely to a demonstration. (In casual conversation during the debriefing period, 1 subject remarked after hearing that the experiment gauged her level of imitation "Oh, you mean when I saw you messing with the box, if I imitate *that*?")

Levels of Analysis

As Whiten et al. (1996) comment in their Discussion, even after delineation of distinct mimetic processes, what counts as imitation—copying the form of an action—remains fundamentally vague. Which part of an act should be copied? The variation of adult performances indicates that subjects—and, likely, species—answer this differently. As a complement to the two-action method used by Whiten et al., one might consider the subjects' actions at different levels of specificity. In reality, the number of levels of consideration is impossibly large: Each action can be subdivided infinitely many times (Whiten & Custance, 1996) or conjoined with other actions to make series of actions, to make recognizable functional actions (such as "head scratch" in lieu of the series "raise arm," "cock finger," "contact head with finger," etc.), or even to speculate about mental levels ("scratched head in puzzlement").

Looking at the coders' interpretation of behavior is a somewhat indirect approach. Comparison of the behavior of the demonstrator to the behavior of the subject might be revealing. I conducted a preliminary analysis of the subjects' actions at four levels of specificity.

Table 6 lists examples of items on each level of analysis. At Level 1, actions are examined on a gross scale; a high score on this level would be emulative behavior. Level 2 is the level of the size and character of actions examined by Whiten et al. (1996); it weights evenly the IO scores (each averaged score a percentage), the action frequency count, and the cases of intragroup significance. Level 3 is intended to capture actions salient to adult humans: What gross motor action is being performed, how many times, and with what hand? Level 4 examines actions at a relatively microscopic scale, with precision beyond that of the functionality of the acts. For each behavioral component in a level, a score is assigned: 1 for imitation and 0 for no imitation, with graded actions in between.

Actions get more specific at higher levels: Instead of considering the box as the main unit of manipulation (Level 1), one could focus on the actions on the lock (Level 3), on the particulars of the actions on the lock (Levels 3 and 4), or on the particulars of each hand movement while acting on the lock (Level 4). The levels also differ on the time scale considered to be important: Level 4 even unwraps the video image into frames (one-tenth of a second each) to compare the action in each one, whereas Levels 1–3 do not put time into the equation at all. In other words, what the action to be imitated is differs across levels. Under this coding scheme, overturning of the pin or bolt (as 1 child's turning a bolt 200 times after a demonstration of 4 turns) would count as expressly not imitating: Although the form of the action is followed, the duration or quantity of the action is not.

Table 6
Levels of Analysis: Increasing Specificity

Open box	Remove treat
Level 1: Emulation	
Yes or no	Yes or no
Level 2: Whiten et al. (1996)	
Independent observer scores	
Action frequency	
Level 3: Salient action	
Hand or finger shape	Yes or no
Handgrip	
Sequence of actions	
Number of times actions done	
Level 4: Microscopic	
Level 3 plus	
Hand or finger used; grip type	Yes or no
Extraneous movements	Show chocolate
Direction of rotation	Keep treat or share
Time spent . . . , etc.[a]	Reclose box

[a] Many more actions than can fit in this table were considered in tabulating this score.

The scores for all elements, weighted evenly, are then added and the final score—as a percentage of total imitation points possible—is calculated. The parameters used to translate the video data into a record of actions are listed in the Appendix.

The results, considered on the various levels of analysis, are shown in Table 7. When the adults' actions are scored in this way, the same behavior that looks imitative on one level is not imitative on another level. At Level 1, all subjects would score 100% imitation on both locks (everyone opened the box), and at Level 4, imitation (scored for only 1 high-performing subject) was less than 5%. On the level of consideration of actions that Whiten et al. (1996) used (Level 2), the adults imitated, on average, 52% of the actions on the T lock and 75% of the actions on the bolts lock; mean imitation at Level 3 was similar on both lock apparatuses, at around 60% (bolts lock range = 31.25%–100%; T lock range = 31.25%–91.25%).

These levels of analysis capture the ambiguity of "imitation" to the subjects, two-thirds of whom believed, according to their answers on the postexperiment questionnaire, that they were in fact imitating the experimenter. Despite the subjects' claims, only when action is considered at the very top level (Level 1) did this majority actually act in a way that would count as imitating. On every other measure, every subject failed to completely imitate. I propose that the reason for this inconsistency is that human subjects are able to determine, through observation or through quick manipulation, what is actually required to open this apparatus. Even if they thought to imitate all the demonstrator's actions, the ease of opening the box some other, more efficient way is too compelling for many of the subjects to ignore. Ultimately, functionality overwhelmed their strategy.

The interpretation of the action of the subjects in the levels of analysis is more direct than the IO method of coding. These findings, with the baseline group report, indicate that the adult human performance is less imitative than the IO scores indicate.

Inferences From Imitation

Under the original experiment's interpretation of subject behavior (Whiten et al., 1996), my conclusion must be that although human adults show some imitative tendency, they do not generally perform true imitation, unlike children. On almost every measure of success or failure at imitation, the adults' performance was dramatically less than faithful. Although the adult behavior was sometimes (as in the poke bolt and spin pin trials) influenced by a demonstration, it often (as the common choice of the nondemonstrated pull bolt) was not. This is supported by the baseline subject data, which showed a strong similarity to the experimental groups' data, despite the baseline subjects' lack of a demonstration.

The differences between the chimpanzee data and the children's data have led to the characterization by some that the children are true imitators, whereas the chimpanzees are doing something like emulation. That is, the children are capable of copying a sequence of behaviors designed to lead to a goal, and the chimpanzees are interested only in the results of the behavior—a change in the state of the world—not in the behavior itself (Tomasello, 1996; Zentall, 1996). Tomasello and Call's (1997) comments epitomize this interpretation: Chimpanzees do not imitate, the authors suggested, "because they do not perceive or understand the original behavior as goal-directed . . . they do not understand the other as an intentional agent who is similar to themselves as an intentional agent" (p. 387).

Tomasello and Call (1997) are not alone in drawing a connection between imitation and sophisticated cognitive or even metacognitive ability. In humans, imitation is thought to be instrumental not only in understanding intentions but also in the development of self-awareness, the acquisition and transmission of culture, the learning of language, and even theory of mind (Meltzoff, 1996; Williams, Whiten, Suddendorf, & Perrett, 2001). Piaget (1962) considered imitation to be one of the necessary precursors of the ability to form mental representations; Hart and Fegley (1994) wrote that imitation has been agreed to be "central to the development of the sense of self" (p. 149). Imitation is widely considered to be a behavior that leads almost directly to those metacognitive skills that make us humans.

Within this human population, however, adults spontaneously imitated less faithfully than children (their comparative skill at imitating was not tested). Given that imitation has been implicated in such a diversity of learning stratagems, why might this be so? Perhaps imitation becomes less useful as humans start to acquire other problem-solving and learning algorithms. Although it is almost the only tool in the infant's toolbox, imitation is no longer a generally applicable tool for adults. As such, it is used less readily, perhaps only when learning a new, complex skill (A. Whiten, personal communication, February 2002). As development progresses, humans shed some of the strategies that were useful in youth. Further, imitation is common in the play of young children (Piaget, 1962), and adults who are knowingly in an experimental setting are presumably less likely to be playful and more likely to perform on the basis of their perception of the demand characteristics of the experiment (Orne, 1962). Subjects' expectations of what the task requires and their level of motivation may play no small role in their behavior (de Waal, 1998). It is interesting to note that, as mentioned above, two-thirds of the subjects reported in a postexperiment questionnaire that they believed that they were supposed to "imitate the experimenter;" precisely what they were to imitate becomes the crux of the concern.

The strong conclusion that must be drawn from the adult humans' limited imitation, by the inferences made by Tomasello (1996), is that human adults do not have an understanding of the intentions of others. I believe that, instead, the most powerful conclusion to be drawn from the experimental results is that this inference is wanting. Imitation alone cannot reveal anything about the mental states of chimpanzees or of humans. The psychological state that underlies the behavior "imitate" or "do not imitate" is

Table 7
Mean Percentage of Actions Imitated by Adult Subjects on Various Levels of Analysis

	Level of analysis			
Lock	1	2	3	4
Bolts	100	75	59	< 5
T	100	52	62	< 5

Note. Level 1 was the most general level of analysis, and Level 4 was a microscopic level of analysis.

underdetermined. Clearly, the ability to imitate and actually imitating are different skills. A number of experimental studies that have looked at chimpanzees' ability to imitate have found that chimpanzees can imitate novel actions on the prompt "Do this!" to a significant degree (Custance, Whiten, & Bard, 1995; Hayes & Hayes, 1952). Whiten et al. (1996) designed the artificial fruit task to ask the question "Do they imitate?" The answer is that these subjects do imitate, sometimes. That the chimpanzees and the adults do not imitate much of the time is not grounds to make conclusions about the subjects' understanding of other minds.

The adult human performance highlights a few points. First, the issue of what counts as an action to be imitated is not completely solved by the two-action test. Further, what the task is defined as being may be unclear, even to the most astute subjects. This bears significantly on the results generated and, by extension, on the conclusions drawn. Especially when assessing the mind of animals, this should cause no small amount of pause: A small change in what is perceived as the task could result in a large change in attribution. Second, because of the small sample sizes obliged by experimenting with animals, the power of the statistical tests used for comparison is reduced. However, the IO tests do allow for trends to be seen and assessed. Large human subject groups will add to the reliability of the results. In this way, adult humans, who are easier to sample than nonhuman primates, may be a useful and informative comparison group for behavioral tests of the capacities of nonhuman animals.

References

Custance, D. M., Whiten, A., & Bard, K. A. (1995). Can young chimpanzees (*Pan troglodytes*) imitate arbitrary actions? Hayes & Hayes (1952) revisited. *Behaviour, 132,* 837–859.

Dawson, B. V., & Foss, B. M. (1965). Observational learning in budgerigars. *Animal Behaviour, 13,* 470–474.

Dennett, D. C. (1987). *The intentional stance.* Cambridge, MA: MIT Press.

de Waal, F. (1998). No imitation without identification. *Behavioural and Brain Sciences, 21,* 687.

Galef, B. G., Jr. (1990). Tradition in animals: Field observations and laboratory analyses. In M. Bekoff & D. Jamieson (Eds.), *Interpretation and explanation of behavior: Comparative perspectives* (pp. 74–95). Boulder, CO: Westview Press.

Hart, D., & Fegley, S. (1994). Social imitation and the emergence of a mental model of self. In S. T. Parker, R. W. Mitchell, & M. L. Boccia (Eds.), *Self-awareness in animals and humans: Developmental perspectives* (pp. 149–165). New York: Cambridge University Press.

Hayes, K. J., & Hayes, C. (1952). Imitation in a home-raised chimpanzee. *Journal of Comparative and Physiological Psychology, 45,* 450–459.

Heyes, C. M. (1996). Genuine imitation? In B. G. Galef Jr. & C. M. Heyes (Eds.), *Social learning in animals: The roots of culture* (pp. 371–389). San Diego, CA: Academic Press.

Meltzoff, A. N. (1996). The human infant as imitative generalist. In B. G. Galef Jr. & C. M. Heyes (Eds.), *Social learning in animals: The roots of culture* (pp. 347–370). San Diego, CA: Academic Press.

Orne, M. T. (1962). On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist, 17,* 776–783.

Piaget, J. (1962). *Play, dreams, and imitation in childhood.* New York: Norton.

Stoinski, T. S., Wrate, J. L., Ure, N., & Whiten, A. (2001). Imitative learning by captive western lowland gorillas (*Gorilla gorilla gorilla*) in a simulated food-processing task. *Journal of Comparative Psychology, 115,* 272–281.

Tomasello, M. (1996). Do apes ape? In B. G. Galef Jr. & C. M. Heyes (Eds.), *Social learning in animals: The roots of culture* (pp. 319–346). San Diego, CA: Academic Press.

Tomasello, M., & Call, J. (1997). *Primate cognition.* New York: Oxford University Press.

Voelkl, B., & Huber, L. (2000). True imitation in marmosets. *Animal Behaviour, 60,* 195–202.

Whiten, A. (1998). Imitation of the sequential structure of actions by chimpanzees (*Pan troglodytes*). *Journal of Comparative Psychology, 112,* 270–281.

Whiten, A., & Custance, D. (1996). Studies of imitation in chimpanzees and children. In B. G. Galef Jr. & C. M. Heyes (Eds.), *Social learning in animals: The roots of culture* (pp. 291–317). San Diego, CA: Academic Press.

Whiten, A., Custance, D. M., Gomez, J.-C., Teixidor, P., & Bard, K. A. (1996). Imitative learning of artificial fruit processing in children (*Homo sapiens*) and chimpanzees (*Pan troglodytes*). *Journal of Comparative Psychology, 110,* 3–14.

Whiten, A., & Ham, R. (1992). On the nature and evolution of imitation in the animal kingdom: Reappraisal of a century of research. In P. J. B. Slater, J. S. Rosenblatt, C. Beer, & M. Milinski (Eds.), *Advances in the study of behavior* (pp. 239–283). San Diego, CA: Academic Press.

Williams, J. H. G., Whiten, A., Suddendorf, T., & Perrett, D. I. (2001). Imitation, mirror neurons and autism. *Neuroscience & Biobehavioral Review, 25,* 287–295.

Zentall, T. R. (1996). An analysis of imitative learning in animals. In B. G. Galef Jr. & C. M. Heyes (Eds.), *Social learning in animals: The roots of culture* (pp. 221–243). San Diego, CA: Academic Press.

(*Appendix follows*)

Appendix

Action Parameters

The parameters used to encode the baseline group's raw data into actions are shown below.

Bolts Lock

1. Twist versus any one-step movement (such as pull)*
2. Pull versus push versus poke
3. Number of times turned (expressed as a ratio: demonstration number to the subject number)*
4. Direction of push

5a. Finger used to poke, if poke

5b. Hand used*

6. Rod order*
7. Bolts brought boxward versus subjectward*
8. Rod in versus rod out (for imitation sequence)
9. Direction of pin twist: Only if IM1 = twist
10. Box open*
11. Chocolate out*

T Lock

1a. Pin turn versus pin spin (no score only if pull)

1b. Pin turn or pin spin versus pin straight out*

2. Number of times turned or spun (expressed as a ratio: demonstration number to the subject number)*
3. Rotate clockwise versus rotate counterclockwise
4. Handle out versus handle in*
5. Handle left versus handle right versus handle straight*
6. Scissor out pin versus grasp out*
7. Hand and finger used*
8. Order of operations (pin, handle)*
9. Box open*
10. Chocolate out*

In addition, these items then formed a subset of items necessary for the scoring on the levels of analysis (see Discussion). Subjects are assigned IM for imitation, NO for no imitation, or AMBIG for ambiguous or mixed actions for each item. Asterisks indicate that items are assigned in every case; the other items are assigned as appropriate. For instance, one subject's score for a T lock trial was: IM1a, IM1b, NO2 (4:1.5), NO3, IM4, NO5, NO6, AMBIG7, IM8, IM9, IM10.

Received April 15, 2002
Revision received December 16, 2002
Accepted December 17, 2002 ■