Identification of the Social and Cognitive Processes Underlying Human Cumulative Culture

L. G. Dean,[1] R. L. Kendal,[2]* S. J. Schapiro,[3] B. Thierry,[4] K. N. Laland[1]*

The remarkable ecological and demographic success of humanity is largely attributed to our capacity for cumulative culture, with knowledge and technology accumulating over time, yet the social and cognitive capabilities that have enabled cumulative culture remain unclear. In a comparative study of sequential problem solving, we provided groups of capuchin monkeys, chimpanzees, and children with an experimental puzzlebox that could be solved in three stages to retrieve rewards of increasing desirability. The success of the children, but not of the chimpanzees or capuchins, in reaching higher-level solutions was strongly associated with a package of sociocognitive processes—including teaching through verbal instruction, imitation, and prosociality—that were observed only in the children and covaried with performance.

The success of humanity in colonizing virtually every terrestrial habitat on the planet and resolving countless ecological, social, and technological challenges is widely attributed to our species' unique capability for "cumulative culture"—the extensive accumulation of knowledge, and iterative improvements in technology, over time (*1*, *2*). Although many animals—especially mammals, birds, and fishes—acquire knowledge and skills from others (often manifest in behavioral traditions), in no instance have these unambiguously exhibited "ratcheting" in complexity (*2*). Given that the adaptive value of cumulative learning is well established (*1*, *3*, *4*), the question as to why social learning is so much more widespread than cumulative culture constitutes a major evolutionary puzzle (*1*, *4*–*7*).

Although claims have been made for cumulative culture in other species (*8*–*10*), the evidence is circumstantial and contested (*2*, *5*, *6*, *11*). The resulting debate has spawned a large number of distinct hypotheses concerning the cognitive capabilities, or social conditions, thought to be necessary for cumulative culture. These explanations include a hypothesized critical dependency of cumulative culture on aspects of social cognition deemed to be exclusive to (or substantially enhanced in) humans, including teaching (henceforth hypothesis 1, or *H1*), language (*H2*), imitation (*H3*), and prosociality (*H4*) (*1*, *2*, *4*, *5*, *11*–*14*). Other explanations stress features of social structure that mitigate against the spread of superior solutions in animals other than humans, including scrounging (kleptoparasitism; *H5*), which can hinder social learning and demotivate resource production (*15*); the tendency of dominant individuals to monopolize resources, thereby preventing subordinates from learning (*H6*) (*16*); and a lack of attention to low-status inventors (*H7*) (*17*, *18*). A further (non-social cognition) hypothesis is that satisficing, or conservative behavior, hinders ratcheting in nonhumans (*H8*) (*8*, *19*). Large social networks (*20*, *21*) may enhance cultural diversity and promote cumulative culture, but we do not consider this hypothesis because it presupposes the existence of the necessary cognitive capabilities.

Cumulative culture has been investigated through historical analysis (*22*), in the psychological laboratory (*23*), and through experimentation in chimpanzees (*14*). However, until now, there has been no extensive and rigorous experimental investigation of the capacity for cumulative cultural learning that simultaneously tests humans and other animals using the same apparatus and that is capable of evaluating all of the aforementioned hypotheses. Here, we present such an investigation.

We designed a puzzlebox (Fig. 1) that could be solved at three stages of difficulty, with success at stage 2 building on stage 1 and success at stage 3 building on stage 2. We presented appropriately scaled versions, under a variety of conditions, to groups of children ($N = 35$, eight groups of 3- to 4-year-olds from three nurseries in Fife, UK), chimpanzees (*Pan troglodytes*; $N = 74$, eight mixed juvenile and adult groups at the Michale E. Keeling Center for Comparative Medicine and Research, University of Texas), and capuchin monkeys (*Cebus apella*; one group over 2 years, year 1 $N = 22$, year 2 $N = 18$, at the Centre de Primatologie, Strasbourg) (see supporting online material). All stages could be completed through two parallel options (Fig. 1), allowing us to investigate cooperation, tolerance, and social learning at the task; presentation in social groups allowed solutions to each level to spread among individuals. Experiment 1 included two conditions: an "open" condition, where groups could gain access to all stages and a "scaffolded" condition, where guards prevented access to the manipulandi associated with higher stages until performance at the lower stage reached criterion. In experiment 2, conducted only with chimpanzees, one female from each of four additional groups was isolated from her group and trained to use the puzzlebox to stage 3. The use of trained females of differing status as demonstrators allowed investigation of how social rank affected the spread of solutions.

Chimpanzees and capuchins were selected because the evidence for cultural traditions is as strong in these species as in any nonhuman (*24*, *25*), maximizing the chances of observing cumulative cultural learning. Moreover, chimpanzees, as our closest relative, provide an appropriate comparator to humans, with the performance of capuchins aiding interpretation of any chimpanzee-human differences. Children are widely used in comparative studies [e.g., (*26*)] to help tease out the effects of culture, as adults have been greatly enculturated by society.

We anticipated that children, but not chimpanzees or capuchins, would exhibit evidence of cumulative cultural learning, and the study was designed to sort between alternative explanations

---

[1] Centre for Social Learning and Cognitive Evolution, School of Biology, University of St. Andrews, Queen's Terrace, St. Andrews, Fife KY16 9TS, UK. [2] Centre for the Coevolution of Biology and Culture, Department of Anthropology, University of Durham, Durham DH1 3LE, UK. [3] Michale E. Keeling Center for Comparative Medicine and Research, University of Texas M.D. Anderson Cancer Center, Bastrop, TX 78602, USA. [4] Département Ecologie, Physiologie et Ethologie, Institut Pluridisciplinaire Hubert Curien, Centre National de la Recherche Scientifique, Université de Strasbourg, F-67087 Strasbourg, France.

*To whom correspondence should be addressed. E-mail: rachel.kendal@durham.ac.uk (R.L.K.); knl1@st-andrews.ac.uk (K.N.L.)

for this. For a hypothesis to be deemed supported (i.e., not falsified), we specified that it should satisfy two criteria: (i) Differences in the relevant predictor variable should covary with differences in species' mean performance in the cumulative task, in the predicted direction; and (ii) within-species variation in the relevant predictor should covary with variation in individual performance within species, in the predicted direction. Thus, we used analyses of covariation, both between and within species, to reject causal hypotheses that were inconsistent with the data. We further assumed that the manner in which social or cognitive processes currently operate is qualitatively similar to the manner in which they operated in ancestral environments. Hence, although our study sought to identify the proximate processes underlying cumulative cultural learning in contemporary populations, the same processes were potentially ancestral sources of selection.

Performance with the puzzlebox is summarized in Fig. 2A. After 30 hours of presentation of the task to each of four chimpanzee groups, only 1 of 33 individuals reached stage 3, with a further 4 having reached stage 2, and with each group having witnessed multiple solvers at stage 1 (experiment 1). Chimpanzee performance was not greatly enhanced by trained demonstrators (experiment 2), who performed stages 1 to 3 proficiently. A similar pattern was observed in the capuchins: After 53 hours, no individual reached stage 3 and only two individuals reached stage 2. Thus, the experiments provide no evidence for cumulative cultural learning in chimpanzees or capuchins. These findings stand in stark contrast to those of the children, where despite a far shorter exposure to the apparatus (2.5 hours), five of the eight groups had at least two individuals (out of a maximum of five) who reached stage 3, with multiple solvers at stages 2 or 3 in all but two groups (see supporting online material).

Analyses revealed support for four of the eight hypotheses (Fig. 2, B to E), suggesting that teaching, communication, observational learning, and prosociality all played important roles in human cultural learning but were absent (or played an impoverished role) in the learning of chimpanzees and capuchins.

A total of 23 unambiguous instances of teaching by direct instruction (i.e., referencing part of the puzzlebox) were observed ($\bar{x} = 0.69 \pm 0.32$), exclusively in the children (*H1*) (Fig. 2B), of which all involved task-relevant communication (e.g., "push that button there") and approximately one-third involved gesture. A strong positive relationship was observed between the amount of instruction received and the stage reached by a child [Spearman's $\rho = 0.598$, $P = 0.0001$ (*27*)]. Such an analysis fails to consider teaching precursors, or subtle processes similar to teaching, such as "pedagogical cuing" (*28*) or "scaffolding" (*29*). To explore whether chimpanzees or capuchins might facilitate learning in others (e.g., through enlisting offspring's interest in the task), we examined rates of both provisioning and food-calling by "knowledgeable" individuals. However, we observed substantially greater rates of tolerated theft of extracted food by mothers from offspring than vice versa in chimpanzees (Wilcoxon $W = 16$, $P = 0.026$) and no tolerated theft in mother-infant pairs of capuchins. Moreover, neither chimpanzees (Wilcoxon $W = 6.5$, $P = 0.77$) nor capuchins ($W = 9$, $P = 0.45$) exhibited any difference in the rate of recruitment of others to the puzzlebox before, versus immediately after, a food call, and low rates of calling were observed; in contrast, children who received verbal instruction outperformed those who did not (*H2*) (Mann-Whitney $U = 41$, $P = 0.002$; Fig. 2C).

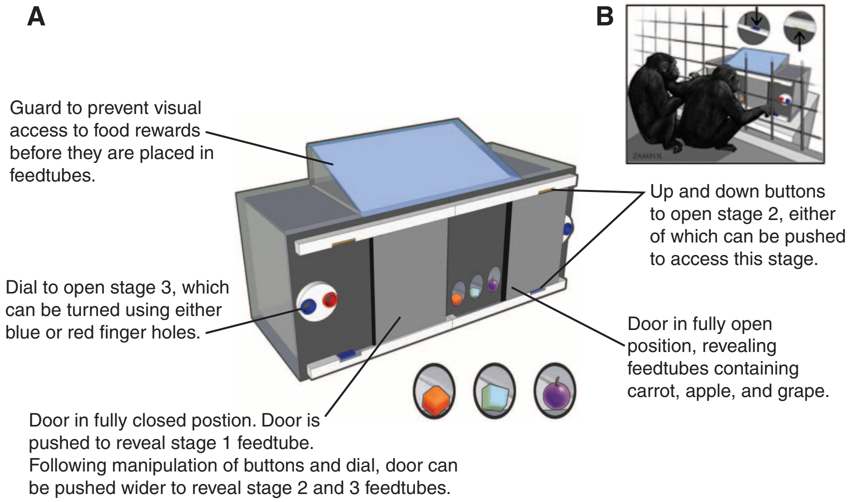


Fig. 1. (**A**). The cumulative culture puzzlebox, which could be solved at three sequential stages, each building on the preceding stage. (**B**) Illustration of puzzlebox use. Stage 1 required individuals to push a door in the horizontal plane to reveal a chute through which a low-grade reward was delivered. Stage 2 required individuals to depress a button and slide the door further to reveal a second chute for a medium-grade reward. Stage 3 required the solver to rotate a dial, releasing the door to slide still further to reveal a third chute containing a high-grade reward. All stages could be completed through two parallel options (alternative doors could be slid left or right at stage 1, alternative buttons at the top or bottom could be depressed at stage 2, and alternative colored finger holes enabled rotation of the dial at stage 3), with sets of three chutes on both left and right sides. This two-action, two-option design aided evaluation of alternative social learning mechanisms and allowed two individuals to operate the puzzlebox simultaneously. Replenishment of the chutes by the experimenter allowed the apparatus to be continuously used for long periods. Pilot work established an unambiguous ascendancy in the desirability of reward with stage (food stage 1 = carrot, 2 = apple, 3 = grapes for chimpanzees and capuchins; stickers of increasing size and attractiveness for children).

We also compared the rate at which individuals from each species, in their first response or during the subsequent minute, performed a matching manipulation (e.g., copy push down button on left) to that observed being performed by another individual departing the box (*H3*). Matching (table S6) could constitute copying the actions of others (i.e., imitation) or making the same manipulandi move in the same way (i.e., emulation). Children alone performed more matching than nonmatching manipulations (Wilcoxon $W = 163$, $P = 0.003$), they produced a significantly greater proportion of matching actions than both chimpanzees and capuchins (Kruskal-Wallis $\chi^2 = 18.13$, df = 2, $P = 0.001$; Fig. 2D), and the degree of matching they exhibited correlated positively with performance (Spearman's $\rho = 0.41$, $P = 0.01$). We observed chimpanzee social learning at stage 1 (option-bias analysis, $\chi^2 = 941.6$, $P = 0.021$) (*30*) but not at higher stages.

Regarding prosociality (*H4*), we hypothesized that if individuals voluntarily give rewards to others, this signifies an understanding that others share the motivation of achieving the goal that they had achieved. We observed 215 altruistic events where a child spontaneously gave another child a retrieved reward ($\bar{x} = 6.14 \pm 2.32$; 47% of children exhibited altruism), but not a single instance of the voluntary donation of food in either the chimpanzees or capuchins (Fig. 2E). The number of prosocial acts received covaried strongly with the stage that a child reached (Spearman's $\rho = 0.54$, $P = 0.001$). Moreover, the proportion of manipulations that children performed at the same time that another individual was in proximity was significantly greater than in either chimpanzees or capuchins, indicating greater tolerance of others, cooperation, and shared motivation among children.

The other four hypotheses failed to satisfy our criteria, providing little evidence that the capability for cumulative culture is affected by either social structure or nonsocial cognition. There was a positive, rather than the predicted negative, correlation between the amount of

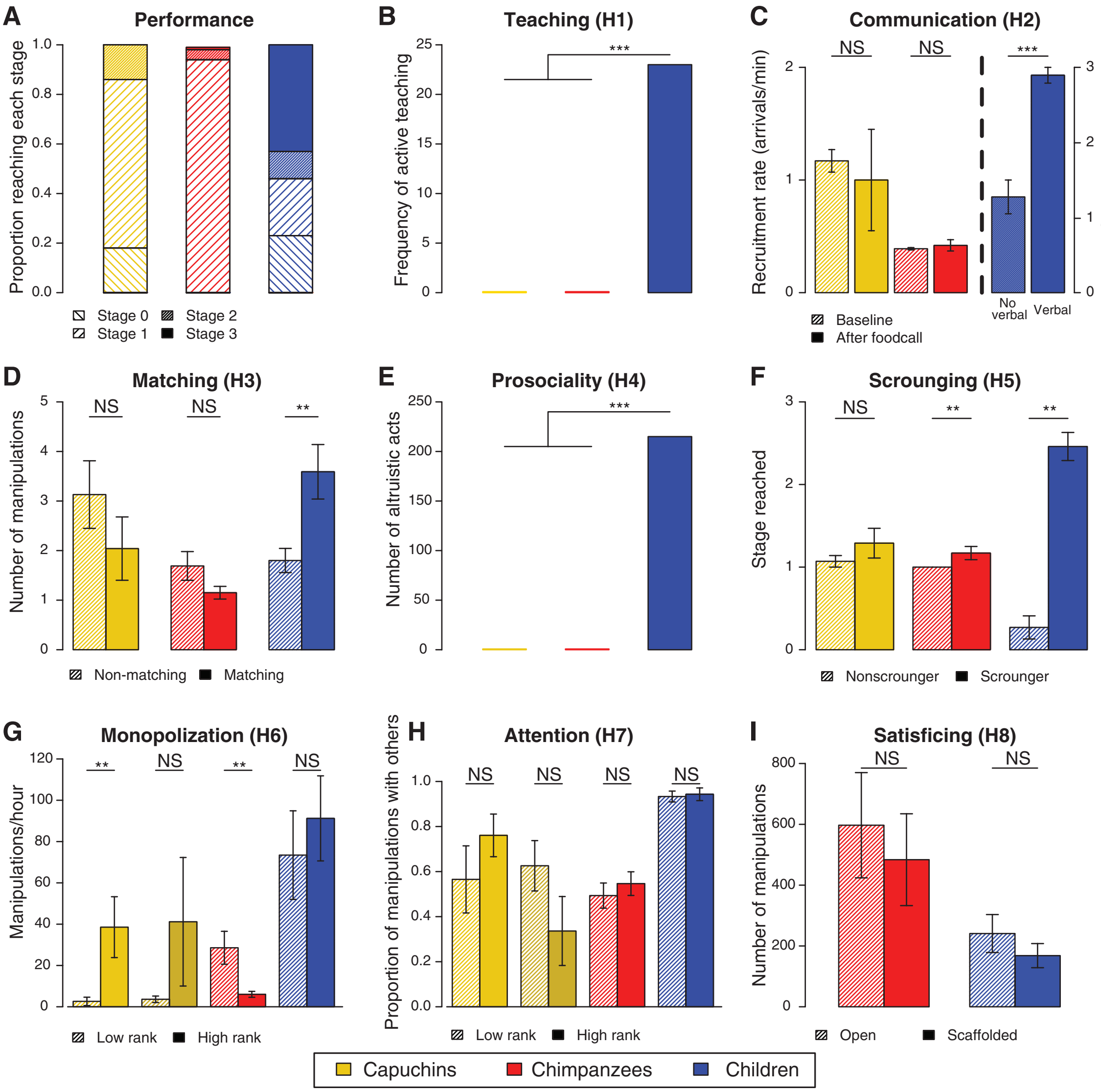


Fig. 2. (**A**) Attainment of stages 2 and 3 was exceptionally rare or absent in capuchins and chimpanzees but common in human children. (**B**) We observed 23 unambiguous instances of teaching, by direct instruction, exclusively in children. (**C**) In neither chimpanzees nor capuchins was there greater recruitment to the task after, versus before, a food call. Conversely, children who received verbal instruction outperformed those who did not. (**D**) Children alone performed more matching than nonmatching manipulations, and they produced a greater proportion of matching actions than did either chimpanzees or capuchins. (**E**) We observed 215 altruistic events (giving an extracted reward to others), exclusively in children. (**F**) There was no evidence that scrounging hindered performance in any species; children who were victims of scrounging outperformed children who were not. (**G**) Dominant children and chimpanzees did not monopolize the task, and high-ranking capuchins monopolized the puzzlebox in 2007 but not 2008. (**H**) Low-rankers did not receive less attention than high-rankers when manipulating the task. (**I**) In the open condition, where they received rewards at all stages, neither chimpanzees nor children manipulated the puzzlebox less than individuals in the scaffolded condition. In (G) and (H), capuchins (2007 and 2008 pale and dark yellow, respectively) and chimpanzees were split into three (high-, mid-, low-) rank categories, although for clarity and comparability with the child data, we present only analysis of high- versus low-ranked individuals. $^{**}P \leq 0.05$; $^{***}P \leq 0.01$; NS, not significant.

Table 1. Numbers (and percentages) of children reaching each stage, together with the amount of teaching, matching (e.g., imitation), and prosociality (altruism) exhibited.

Stage reached	Number of individuals (from *N* groups)	Number that received teaching	Number with more matching than nonmatching manipulations	Number that received altruism	Number that received no social support
0	8 (5)	0	1 (12.5%)	0	7 (87.5%)
1	8 (3)	0	4 (50%)	4 (50%)	3 (37.5%)
2	4 (3)	1 (25%)	1 (25%)	2 (50%)	1 (25%)
3	15 (5)	9 (60%)	11 (73%)	11 (73%)	0

scrounging an individual falls victim to and performance in capuchins (Spearman's $\rho = 0.71$, $P = 0.0002$), chimpanzees ($\rho = 0.312$, $P = 0.008$), and children ($\rho = 0.8$, $P = 6.87 \times 10^{-9}$), and no sign that scrounging hindered performance (*H5*) (Fig. 2F). Dominant children (Wilcoxon $W = 186$, $P = 0.15$) and chimpanzees [analysis of variance (ANOVA) $F_{2,72} = 3.49$, $P = 0.036$] did not monopolize the puzzlebox (*H6*), and although there was a positive correlation between rank and puzzlebox use among capuchins in 2007 (Kruskal-Wallis $\chi^2 = 8.23$, df = 2, $P = 0.016$), this was not repeated in 2008 ($\chi^2 = 0.13$, df = 2, $P = 0.93$) (Fig. 2G). When manipulating the box, low-rankers did not receive less attention (defined as having others within 1.5 m of the task) than high-rankers (*H7*) (Fig. 2H; capuchins 2007, Kruskal-Wallis $\chi^2 = 2.49$, df = 2, $P = 0.29$; capuchins 2008, $\chi^2 = 2.08$, df = 2, $P = 0.35$; chimpanzees, ANOVA $F_{2,72} = 1.22$, $P = 0.3$; children, Wilcoxon $W = 100$, $P = 0.61$), nor was there any evidence for satisficing or conservatism (*H8*) (Fig. 2I; chimpanzees, Mann-Whitney $U = 166.5$, $P = 0.42$; children, $U = 163$, $P = 0.54$), with individuals continuing to manipulate the dials and buttons of the puzzlebox after they had found the solution to stage 1. In the open condition, where they received rewards at all stages, both chimpanzees and children manipulated the puzzlebox slightly more, rather than less, than individuals in the scaffolded condition, despite the latter being unrewarded at the previous stage(s). Although we did not find a significant difference between the proportions of rewards scrounged at each stage in chimpanzees, they expressed clear and strong preferences for the three foodstuffs in pilot work, and olfactory holes in the doors allowed these foods to be detected in the apparatus prior to their extraction. Moreover, many of the chimpanzees performed failed attempts to access the foods by "termiting" (inserting stalks through the olfactory holes), and all 29 cases involved an attempt to reach the highest-stage food that was available. In the children and capuchins, more low-stage than high-stage rewards were scrounged, which reflects a greater motivation to retain high-grade rewards.

Thus, we found no support for the hypotheses that cumulative culture is absent in chimpanzees or capuchins because in these species the social transmission of superior solutions is hindered by scrounging, because dominant individuals monopolize key resources, because of a lack of attention to low-status innovators, because these animals satisfice, or because these animals were unable to discriminate higher-quality from lower-quality rewards. Nor can the results be easily dismissed as an artifact of captivity testing, as wild chimpanzees and capuchins have been subjected to long-term studies that reveal no unambiguous evidence for cumulative culture (*24*, *25*). Likewise, our animals cannot be described as "dysfunctional" because they have performed effectively in previous studies demonstrating social learning and tradition of noncumulative tasks (*31*, *32*).

Closer inspection of the children's behavior supports the conclusion that a package of social cognitive capabilities, encompassing teaching (largely through verbal instruction) as well as matching (e.g., imitation) and prosociality (altruism), was critical for performance at the highest level. Table 1 reveals that all children who reached level 3 received at least one form of social support and 86% received at least two types. Conversely, children who did not benefit from social support generally performed poorly in the task. These data not only provide clear and strong evidence for a cumulative cultural capability in the children but strongly link their elevated performance to their social cognition.

The puzzlebox experiment reveals clear and characteristic differences in cumulative cultural learning and patterns of social interaction among children, chimpanzees, and capuchin monkeys, highlighting sociocognitive processes that may be important for cultural transmission to "ratchet." The children responded to the apparatus as a social exercise, manipulating the box together, matching the actions of others, facilitating learning in others through verbal instruction and gesture, and engaging in repeated prosocial acts of spontaneous gifts of the rewards they themselves retrieved. In contrast, the chimpanzees and capuchins appeared to interact with the apparatus solely as a means to procure resources for themselves, in an entirely self-serving manner, largely independent of the performance of others, and exhibiting restricted learning that appeared primarily asocial in character.

Our findings, based on confirmation of predicted patterns of covariation both between and within species, constitute strong support for the view (*2*, *11*, *12*) that cumulative culture requires a package of key psychological processes—specifically, teaching through verbal instruction, imitation, and prosocial tendencies—that are present in humans but are absent or impoverished in chimpanzees and capuchins. The claim that these sociocognitive processes, rather than other effects, were directly responsible for the pattern of cumulative cultural learning observed in the children is supported by the positive relationships found between the stage reached and the amount of teaching, verbal instruction, and prosocial acts received, as well as between the stage reached and the amount of observational learning that took place (Table 1).

We reject as improbable the alternative causal hypotheses that performing well in the task caused elevated levels of the predictor variables or that some unspecified factor elevated both performance and the predictors. It is not clear why success in solving the task should cause children to imitate, be taught by, or receive rewards from others, nor how an unspecified third variable might account for our within-species data. For instance, although it is possible that the relationship between imitation and performance reflects the child's cognitive ability, this explanation cannot account for the relationships of both teaching and prosociality with performance, because in both cases the donor (of knowledge or reward) is a different individual from the learner. The most likely explanation is that aspects of human social cognition are directly responsible for the cumulative culture capability.

Human cultural traditions accumulate refinements over time, thereby producing both technology and other cultural achievements of astonishing complexity and diversity unprecedented in the rest of nature. Although numerous hypotheses have been proposed for this phenomenon, the explanation has for many years remained elusive. Our experiment provides a clear answer to this conundrum, providing strong support for the position advanced by Tomasello and colleagues (*2*, *11*, *12*, *33*) that "human social learners focus to a much greater degree than other nonhuman primates on the actual actions performed by others …[and]… that uniquely human forms of cooperation ... teaching and norms of conformity contribute to the cultural ratchet" [(*11*), p. 2413]. These findings pave the way for an exciting avenue of research into when and why this particular "package" of other-regarding sociocognitive capacities evolved.

References and Notes

1. R. Boyd, P. Richerson, *Culture and the Evolutionary Process* (Univ. of Chicago Press, Chicago, 1985).
2. M. Tomasello, in *Chimpanzee Cultures*, R. W. Wrangham, W. C. McGrew, F. B. M. de Waal, P. G. Heltne, Eds. (Harvard Univ. Press, Cambridge, MA, 1994), pp. 301–377.
3. M. Enquist, S. Ghirlanda, *J. Theor. Biol.* **246**, 129 (2007).
4. R. Boyd, P. Richerson, *Proc. Br. Acad.* **88**, 77 (1996).
5. B. G. Galef, *Hum. Nat.* **3**, 157 (1992).
6. C. M. Heyes, *Anim. Behav.* **46**, 999 (1993).
7. C. Boesch, M. Tomasello, *Curr. Anthropol.* **39**, 591 (1998).
8. C. Boesch, *Evol. Anthropol.* **12**, 82 (2003).
9. G. R. Hunt, R. D. Gray, *Proc. Biol. Sci.* **270**, 867 (2003).
10. C. Sanz, D. Morgan, *Int. J. Primatol.* **30**, 411 (2009).
11. C. Tennie, J. Call, M. Tomasello, *Philos. Trans. R. Soc. London Ser. B* **364**, 2405 (2009).
12. M. Tomasello, *The Cultural Origins of Human Cognition* (Harvard Univ. Press, Cambridge, MA, 1999).
13. K. N. Laland, *Learn. Behav.* **32**, 4 (2004).
14. S. Marshall-Pescini, A. Whiten, *Anim. Cogn.* **11**, 449 (2008).
15. L. Giraldeau, L. Lefebvre, *Anim. Behav.* **35**, 387 (1987).
16. S. Coussi-Korbel, D. Fragaszy, *Anim. Behav.* **50**, 1441 (1995).
17. S. Reader, K. Laland, *Int. J. Primatol.* **22**, 787 (2001).
18. D. Biro *et al.*, *Anim. Cogn.* **6**, 213 (2003).
19. C. Hrubesch, S. Preuschoft, C. van Schaik, *Anim. Cogn.* **12**, 209 (2009).
20. J. Henrich, *Am. Antiq.* **69**, 197 (2004).
21. K. R. Hill *et al.*, *Science* **331**, 1286 (2011).
22. G. Basalla, *The Evolution of Technology*, G. Basalla, W. Coleman, Eds. (Cambridge Univ. Press, Cambridge, 1988).
23. C. Caldwell, A. Millen, *Evol. Hum. Behav.* **29**, 165 (2008).
24. A. Whiten *et al.*, *Nature* **399**, 682 (1999).
25. S. Perry *et al.*, *Curr. Anthropol.* **44**, 241 (2003).
26. E. Herrmann, J. Call, M. V. Hernàndez-Lloreda, B. Hare, M. Tomasello, *Science* **317**, 1360 (2007).
27. We conducted two classes of analyses, cross-species comparisons and single-species analyses exploring the relationship between a predictor variable and performance. For the latter analyses, to render the

distribution continuous, we computed an "achievement rank" by combining the number of the stage reached and the number of successful manipulations (see supporting online material).
28. G. Csibra, G. Gergely, in *Processes of Change in Brain and Cognitive Development*, Y. Munakata, M. Johnson, Eds. (Oxford Univ. Press, Oxford, 2006), pp. 249–274.
29. D. Wood, J. S. Bruner, G. Ross, *J. Child Psychol. Psychiatry* **17**, 89 (1976).
30. R. L. Kendal, J. R. Kendal, W. Hoppitt, K. N. Laland, *PLoS ONE* **4**, e6541 (2009).
31. A. Whiten *et al.*, *Curr. Biol.* **17**, 1038 (2007).
32. M. Dindo, B. Thierry, A. Whiten, *Proc. R. Soc. B* **275**, 187 (2008).
33. M. Tomasello, M. Carpenter, J. Call, T. Behne, H. Moll, *Behav. Brain Sci.* **28**, 675 (2005).

Acknowledgments: Supported by the CULTAPTATION project (European Commission contract FP6-2004-NESTPATH-043434), a European Research Council Advanced Grant (EVOCULTURE, 232823) (K.N.L.), and a Royal Society Dorothy Hodgkin Fellowship (R.L.K.). Chimpanzees at M. D. Anderson Cancer Center were supported by NIH Cooperative Agreement RR-15090. We thank S. Pavonetti, A. Whiten, C. Caldwell, C. Tennie, M. Tomasello, C. van Schaik, the CCBC/EARG journal club (Durham), and members of the Laland lab for advice and/or useful comments on earlier drafts of our manuscript, and J. Zampol for drawing Fig. 1. All research was approved by the ethics committee at the University of St. Andrews and the relevant committees at the research facilities and follows the relevant legislation in the UK, USA, and France. Data are available from the authors upon request.

Supporting Online Material
www.sciencemag.org/cgi/content/full/335/6072/1114/DC1
Materials and Methods
Figs. S1 and S2
Tables S1 to S6

14 September 2011; accepted 23 January 2012
10.1126/science.1213969

The Effects of Experience and Attrition for Novice High-School Science and Mathematics Teachers

Gary T. Henry,[1]* C. Kevin Fortner,[2] Kevin C. Bastian[3]

Because of the current high proportion of novice high-school teachers, many students' mastery of science and mathematics depends on the effectiveness of early-career teachers. In this study, which used value-added models to analyze high-school teachers' effectiveness in raising test scores on 1.05 million end-of-course exams, we found that the effectiveness of high-school science and mathematics teachers increased substantially with experience but exhibited diminishing rates of return by their fourth year; that teachers of algebra 1, algebra 2, biology, and physical science who continued to teach for at least 5 years were more effective as novice teachers than those who left the profession earlier; and that novice teachers of physics, chemistry, physical science, geometry, and biology exhibited steeper growth in effectiveness than did novice non–science, technology, engineering, and mathematics teachers.

In the past two decades, the teacher labor market has dramatically changed in response to more employment opportunities for women, increased demand for teachers, and policies opening new pathways into the profession (*1*). For instance, the modal value of experience for U.S. teachers dropped from 15 years in 1987–1988 to 1 year in 2007–2008 (*2*). Additionally, turnover for beginning teachers is high: After just 5 years, nearly 50% of all novice teachers have exited the profession (*3*). This churn of beginning teachers in and out of public schools results in more students, particularly poor and/or ethnic minority students, being taught by novice teachers (*4*, *5*), and that in turn leads to reduced student achievement (*6*–*11*). Investigating the consequences of these teacher labor market conditions for high-school students' science and mathematics achievement—specifically, the effects of experience and attrition among novice teachers—will shed light on challenges facing education and career preparation in science and mathematics.

We quantified the growth in effectiveness of high-school science and mathematics teachers and the effects of those teachers who exit public school classrooms. We analyzed effectiveness using scores on standardized tests given to high-school students in three mathematics courses and four science courses. We define teachers' effectiveness in terms of the increases in their students' test scores, adjusted for the prior achievement of the individual students and for other student, classroom, and school covariates. Prior research shows that the average effectiveness of novice teachers increases during their first 3 years and flattens thereafter (*12*–*14*); and that after differences in effectiveness that are attributable to experience are removed, less effective teachers are more likely to exit the profession (*10*, *15*, *16*). We extended this research to investigate teacher effectiveness in specific high-school courses and addressed three questions: (i) To what extent do novice high-school science and mathematics teachers become more effective with additional experience? (ii) Are novice high-school science and mathematics teachers who exit public schools more or less effective than those who stay? [There is little published research on where teachers who exit public schools are subsequently employed, but it suggests that leaving for higher-paying jobs in the private sector is relatively rare (*17*).] (iii) Do the rates of change in effectiveness for high-school science and mathematics teachers vary by course?

Student outcomes are related to variations in many school-related factors, including leadership; an orderly environment; high student expectations; a focus on student outcomes; a positive school culture; parental involvement; and, most closely, to teachers' effectiveness (*18*, *19*). Teachers are the most important school-related variable explaining variation in student achievement (*12*, *18*, *20*), and teacher experience positively affects student performance (*6*–*11*). Most gains in effectiveness occur in the first 3 years of teachers' careers, with minimal increases thereafter (*13*).

It is likely that through teaching experience, trial and error, professional development, mentoring, and/or collaboration with fellow educators, teachers learn rapidly during their first few years on the job. However, some of the average increases in effectiveness that have been attributed to experience may be a statistical artifact caused by the exit of less effective early-career teachers, thereby overstating the year-to-year differences in the statistics related to experience. Recent research supports this second explanation, finding that exiting teachers are less effective than comparable teachers who remain in the profession (*10*, *15*, *16*). Here we disentangle the effects of teacher development from differential attrition among high-school teachers of science and mathematics courses.

To assess the effectiveness and attrition of novice science and mathematics teachers, we developed a data set from North Carolina containing end-of-course test scores for seven science and mathematics courses—algebra 1, algebra 2, geometry, biology, chemistry, physical science, and physics—and three other courses—English 1, U.S. history, and civics/economics—which we grouped together for the purposes of our analysis and label non-STEM (science, technology, engineering, and mathematics) courses. Our study sample included all teachers in tested subjects with less than 5 years of experience employed in any regular North Carolina public high school from 2005–2006 through 2009–2010. The most crucial feature of the data set is that students and teachers were linked on the basis of actual classroom rosters, which allowed us to match approx-

---

[1] Department of Public Policy and Education Policy Initiative at Carolina, University of North Carolina at Chapel Hill, Abernethy Hall, Campus Box 3435, Chapel Hill, NC 27599–3435, USA. [2] Department of Educational Policy Studies, College of Education, Georgia State University, 30 Pryor Street, Atlanta, GA 30303–3083, USA. [3] Department of Public Policy and Education Policy Initiative at Carolina, University of North Carolina at Chapel Hill, Abernethy Hall, Campus Box 3279, Chapel Hill, NC 27599–3435, USA.

*To whom correspondence should be adddressed. E-mail: gthenry@unc.edu

www.sciencemag.org/cgi/content/full/335/6072/1114/DC1

Supporting Online Material for

Identification of the Social and Cognitive Processes Underlying Human Cumulative Culture

L. G. Dean, R. L. Kendal,* S. J. Schapiro, B. Thierry, K. N. Laland*

*To whom correspondence should be addressed. E-mail: rachel.kendal@durham.ac.uk (R. L.K.); knl1@st-andrews.ac.uk (K.N.L.)

Published 2 March 2012, *Science* **335**, 1114 (2012)
DOI: 10.1126/science.1213969

This PDF file includes:

Materials and Methods
Figs. S1 and S2
Tables S1 to S6

METHODS

We exposed social groups of 3-4 year-old children, adult and juvenile chimpanzees and capuchin monkeys to an experimental puzzle box (Figure 1, Figure S1) that could be solved at three sequential levels to retrieve rewards of increasing desirability. The study was designed to evaluate eight separate hypotheses concerning the factors necessary for cumulative cultural learning (Table S5, below). Two experiments were conducted. The first involved presenting groups of naive subjects with the puzzlebox, across two conditions (an ‘open’ condition where groups could gain access to all stages, and a ‘scaffolded’ condition, where guards prevented access to the manipulandi associated with higher stages until performance at the lower stage reached criterion), and recording which individuals interacted with it, when and how as well as who observed these interactions. This experiment was carried out with all three species, although the capuchins experienced only the scaffolded condition. The second experiment was carried out with chimpanzees only and involved training demonstrator animals, of high and low status, to solve the box and retrieve food effectively. These trained individuals were then reintroduced into their native groups with the puzzlebox, and allowed to demonstrate successful solutions. We again monitored which individuals interacted with the puzzlebox as well as when, how and who observed these interactions. This second experiment was designed to determine whether the failure of the chimpanzees to achieve high-level solutions in the first experiment could be attributed to an absence of quality demonstration, as well as to evaluate whether the status of the demonstrator affected the likelihood of individuals adopting a behaviour pattern.

Subjects

(i) Chimpanzees. Subjects were housed at the Michale E. Keeling Center, MD Anderson Cancer Center, Bastrop, TX, USA. They were tested in the outdoor portion of their enclosures, which are octagonal corrals 24.3 metres in diameter. Chimpanzees were not food deprived before the experiment, but were not tested within an hour of a large feed.

Table S1. Chimpanzee groups participating in the experiment.

Experiment	Group Number	Condition (Exp 1)/ Demonstrator rank (exp 2)	Number of males	Number of females	Number of adults	Number of sub-adults/ juveniles	Mean age of group (yrs) (± standard error)
1	C1	Open	4	4	7	1	25 (±2.60)
1	C5	Open	4	6	8	2	19.3 (±2.03)
1	C6	Scaffolded	3	5	8	0	32.4 (±3.59)
1	C8	Scaffolded	2	5	6	1	31.6 (±6.17)
2	C2	High	7	6	12	1	26.5 (±3.39)
2	C3	Low	4	5	9	0	22.7 (±1.87)
2	C4	Low	2	9	10	1	23.5 (±3.40)
2	C7	High	2	6	8	0	31.6 (±3.39)

The 74 subjects were aged between 6 and 48 years old and were housed in 8 multi-male, multi-female groups, ranging in size from 7 to 13 individuals (Table S1).

(ii) Capuchins. Subjects were housed at the Centre de Primatologie, Strasbourg, France. The single population was tested in the outdoor portion of their

enclosure, consisting of two interconnected runs measuring 45m^2 in total. The puzzlebox was placed at the end of the larger run with access allowed to both outdoor runs during the experiment.

Table S2. Capuchins participating in the experiment. *Individuals that were removed from the group in March 2008

Name	Sex	Month/Year of birth	Age category 2007/2008	Rank 2007/2008	Rank category 2007/2008
Accroc*	Male	08/1996	Adult	1/NA	High/NA
Alila	Female	08/1999	Adult	15/3	Mid/High
Arnaud	Male	07/1998	Adult	2/1	High/High
Asson*	Female	05/1989	Adult	6/NA	High/NA
Boy	Female	01/1973	Adult	17/8	Low/Mid
Kinika	Female	06/1992	Adult	7/13	High/Low
Kiwi	Female	~1980	Adult	3/10	High/Mid
Kolette	Female	08/1999	Adult	11/9	Mid/Mid
Olive*	Female	09/2000	Adult	16/NA	Low/NA
Paola	Female	06/2001	Adult	18/11	Low/Mid
Petula	Female	04/2001	Adult	13/12	Mid/Low
Pistou	Male	04/2001	Adult	4/4	High/High
Popeye	Male	05/2001	Adult	10/5	Mid/High
Raven	Male	08/2002	Adult	8/2	Mid/High
Rosy	Female	05/2002	Adult	5/7	High/Mid
Samir	Male	05/2003	Adult	9/6	Mid/Mid
Shaka*	Female	07/2003	Adult	14/NA	Mid/NA
Velvet	Male	10/2006	Juvenile/ Subadult	21/14	Low/Low
Vicky	Female	03/2006	Juvenile/ Subadult	20/16	Low/Low
Vlad*	Male	05/2006	Juvenile/ Subadult	12/NA	Mid/NA
Wallis	Male	05/2007	Infant/ Juvenile	19/15	Low/Low
Willow	Female	08/2007	Infant/ Juvenile	22/17	Low/Low

The capuchin group was a multi-male, multi-female group with ages ranging from 0.5 years to over 30 years (Table S2). Testing was carried out in two sessions, in November - December 2007 and June 2008. During the intervening six months, five members of the group were removed to start a new colony at a separate facility. For the 2007 cohort *N*= 22, and for the 2008 cohort *N*=17.

(iii) Children. Participants were tested at three nursery schools, namely St. Andrews Nursery School, Lawhead Primary School and Westfield Nursery School, in east Fife, UK. They were tested in an area of their schools that was

separate from the main class, but was familiar to them. Where required by the school, a teacher was present in the room also, although they were requested not to speak or interact with the children during the trial sessions.

Eight groups of children were tested with group sizes of 4 and 5. The age range of the groups was 40 to 59 months. There was always a mix of sexes within the groups, although exact sex ratio varied (Table S3). The parents of all children involved in the study had signed consent forms agreeing that their child could participate.

Table S3. Child groups participating in the experiment.

Group identity	Condition	Number of males	Number of females	Mean age of groups (yrs/months) (± standard error [months])
1	Scaffolded	3	1	3.6 (± 1.5)
2	Scaffolded	3	1	4.1 (± 1.9)
3	Scaffolded	4	1	4.7 (± 1.3)
4	Scaffolded	2	2	3.9 (± 3.1)
5	Open	4	1	3.9 (± 2.0)
6	Open	1	3	4.2 (± 2.0)
7	Open	3	2	4.3 (± 2.1)
8	Open	2	2	3.8 (± 1.9)

Ethics Approval

All research was approved by the ethics committee of the University of St. Andrews, in addition chimpanzee and capuchin work was approved by the ethics committees of MD Anderson Cancer Center and the Centre de Primatologie respectively. All research complied with both the legislation of the UK and the countries in which the research was conducted.

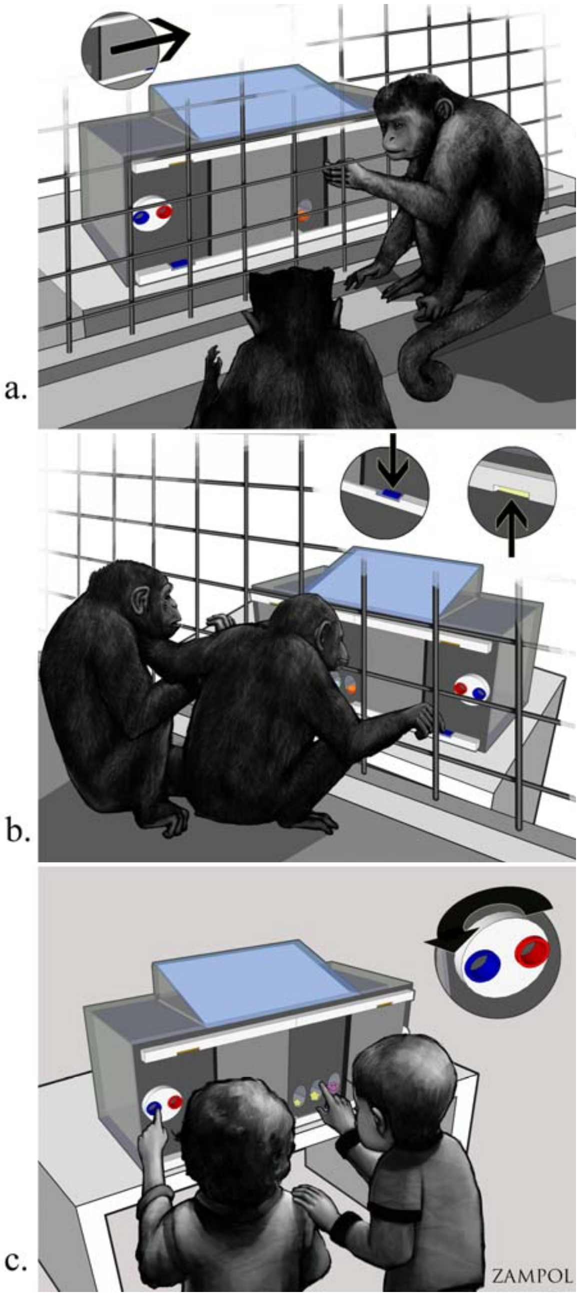


Figure S1: The puzzlebox used in the experiments, showing the three different species interacting with the puzzlebox. A- capuchins opening stage 1, B- chimpanzees pushing the down button to solve stage 2, C- children using the blue fingerhole to solve stage 3.

Apparatus

The puzzlebox used in this experiment could be solved sequentially, at three separate levels, or ‘stages’. The three stages offer successively more desirable rewards, but require more complex manipulations to solve, with each stage building upon the previous one (see Fig. S1). The box was designed symmetrically, allowing two parallel options (alternative doors could be slid left or right at stage 1, alternative buttons at the top or bottom could be depressed at stage 2, and alternative coloured finger-holes enabled the dial to rotate clockwise or counter-clockwise at stage 3) with which to complete each stage. This two-action, two-option design allowed us to distinguish between alternative social learning mechanisms.

The first stage could be opened by sliding one of the two doors outwards in a horizontal plane, the left-side door moving to the left and the right-side door to the right. This action revealed a feeding chute through which a low-level reward could be delivered, with each door revealing a separate symmetrically placed tube. The second stage could be opened by pushing one of two buttons; either the button in the top runner, upwards, or the button in the bottom runner, downwards. Depression of either of these buttons allowed the door to be slid open wider to reveal a second food tube on that side, from which a mid-level reward was delivered. Once again, there were symmetrically placed upper and lower buttons on each side of the box, and symmetrically placed mid-level feeding tubes on right and left sides. The final stage was opened by turning a dial, using either a red or blue bordered finger hole, which allows the door to be slid open even further, to reveal a third feeding tube on that side from which a high-level reward could be retrieved. Again, there were symmetrically placed dials on each side of the box, and symmetrically placed feeding tubes delivering high-level food on the right and left side. Olfactory holes were drilled into each puzzlebox door, to help ensure that the subjects were aware of the presence of the rewards behind them.

The puzzleboxes given to children, chimpanzees and capuchin monkeys differed only in size, being scaled appropriately to the mean size of the subject.

(i) Chimpanzees. The puzzlebox used with chimpanzees was 700mm (l) x 300 mm (h) x 300(w)mm, with the main frame constructed of Perspex. The doors were 220 (h) x 160 (w)mm and were made of acrylic veneered with steel for added strength. The acrylic buttons at stage two were positioned 130mm from each end of the puzzlebox and measure 40 (l) x 10 (w) mm. The dials (diameter 100mm) were positioned 50mm from each end of the puzzlebox and were also made from acrylic. The entire puzzlebox was bolted to a cart to ensure the safety of animals and experimenters and to assist in transport.

Chimpanzees were tested at an observation 'window' in the outdoor corrals. This was 1(h) x 1.93 (w)m and covered with bars 51mm apart. Subjects were able to reach through the bars and operate the puzzlebox, which was located outside the enclosure. When first presented to them, the puzzlebox was novel to all animals in the group, although the required actions were similar to those displayed by the chimpanzees when presented with other puzzleboxes (e.g. Whiten et al., 2007). The actions required to solve the puzzlebox were, therefore, likely to be in the repertoire, or similar to actions in the repertoire, of the chimpanzees. In Experiment 2, individual demonstrator training took place in the indoor enclosures of the chimpanzee facility as described below.

(ii) Children and capuchins. The puzzlebox used with capuchins and children was constructed in the same way as the chimpanzee puzzlebox, except that the doors did not require a veneer of steel. This puzzlebox measured 540 (l) x 180 (h) x 190(w) mm. The doors measured 120(w) x 115(h) mm each and, when closed, were 140mm from the end of the puzzlebox. The buttons measured 30(l) x 5(w)mm and were positioned 75mm from each end of the puzzlebox. The dials were 50mm in diameter and were positioned 90mm from the bottom of the puzzlebox and 45mm from each end.

For the capuchins, the puzzlebox was placed outside of the outdoor enclosure with capuchins being able to reach through the $50mm^2$ mesh to reach and manipulate it.

For the children, the puzzlebox was positioned on a table and children were instructed before the start of the first trial where in the room they were allowed to walk. If necessary a barrier of chairs prevented the children walking directly behind the puzzlebox, in order to ensure they did not gain visual access to the mechanisms under the control of the experimenter.

When in use, the experimenter sat behind the puzzlebox to reset and re-bait the box with the rewards. The experiments were filmed with a Sony Handicam DCR-HC27E, which was positioned behind the experimenter for the chimpanzee and capuchin trial and to one side of the box in the children trials.

Procedure

Reward preference testing

Prior to the experiment, food preference testing was carried out with the chimpanzees and capuchins in order to establish suitable low-, mid- and high-level rewards. In the case of the chimpanzees, initial trials utilised food identified in previous food preference trials carried out by Brosnan et al. (Brosnan SF, Talbot C, Ahlgren M, Lambeth SP & Schapiro SJ 2010 Animal Behaviour 79, 1229-1237; Brosnan, pers. comm.). Each chimpanzee group was tested with a separate food preference test. Testing occurred when chimpanzee groups were allowed back into their indoor enclosures following husbandry procedures. Half a kilo of three foods, *(i)* grapes, and grape-sized pieces of *(ii)* carrots, and *(iii)* apples - were each separately placed in four piles, totalling 12 piles of food spaced evenly, in a randomised order, across the floor of the enclosure. The food first consumed by each subject in the group was recorded, as well as the order in which the four piles of food were completely consumed. This was repeated three times with every experimental group prior to the commencement of the experimental trials.

The capuchin food preference testing exploited the fact that the capuchins were previously trained to exchange items and have been involved in experiments in which they choose between two options offered to them by an experimenter. Whilst the group were freely associating in their outdoor enclosure, individuals were presented with two foods (from carrot, apple and grape) and were allowed to choose one food, which they were able to consume. The order of food presentation and the hands in which foods were presented was randomised over time. Due to dominance in the group, some individuals received more tests than others as they displaced the focal individual, however, eighteen (81%) of the population each received at least five food preference tests.

In both chimpanzees and capuchins we observed an unambiguous pattern of preference, with grape being deemed most desirable, then apple, then carrot.

Children were given stickers as rewards, as is common in developmental psychology studies (e.g. Herrmann et al., 2007). Prior to the experimental sessions the children were told that during the game they might get stickers, although they were not told that these rewards would come from the puzzlebox. A pilot study with five children, none of whom took part in the main experiment, was conducted; in this study children were asked to stick a range of stickers on a piece of paper in order of desirability. Stickers were chosen for the main experiment that appeared in the hierarchies in the same order in the pilot, regardless of the exact rank each child gave the sticker. Experimental groups were told the order of desirability of stickers with small stars being bettered by large stars, which were, in turn, bettered by stickers displaying a smiling face and a glittery background.

Experiment 1

(i) Chimpanzees

All trials were one hour in duration and were conducted in the morning between 9am and 12pm. Trials were conducted at least 30 minutes after the usual morning feed of vegetables and fruit and before the provision of the chow feed.

The exact timing of the trials was randomised to control for feeding motivation of animals throughout the morning. In four instances early termination of testing was required, due to malfunctioning of the puzzlebox or a security breach. All groups were exposed to the puzzlebox for a total of 30 hours.

The trials were conducted from August to October 2007, and from August 2008 to January 2009. One trial, per group, was conducted per day as frequently as practicable.

The procedure differed across two conditions, designed to examine the importance of satisficing and conservatism (hypothesis 8, Table S5), by controlling the protocol in which rewards are given.

"Open" Condition

Two groups (*N*=8 & *N*=10) were presented with the puzzlebox with food provided at all stages. Individuals were able to manipulate the puzzlebox to any stage and receive the food reward at that level. If an individual successfully opened the puzzlebox to stage three then all manipulandi were immediately reset and the food tubes restocked. However, if animals performed unsuccessful manipulations or successfully manipulated the box and opened stage one or two then two minutes after the initial manipulation all manipulandi were reset and the food tubes restocked.

"Scaffolded" Condition

Two groups (N=8 & N=7) were presented with regulated access to parts of the puzzlebox. Here, the dial and buttons of the task were shielded, using guards, such that the subjects could only gain access to the stage one doors. When 75% of the group had successfully manipulated the doors of the puzzlebox at least five times in a trial (a criteria judged as indicating 'learning' of the technique), the guards covering the button manipulandi for stage 2 were removed. At this point the reward was removed from stage 1, thus animals must successfully manipulate stage 2 (having manipulated stage one with no reward) to receive a reward. This procedure was to be repeated in transition of individuals from stage two to stage three of the puzzlebox. As with the open condition, the manipulandi were returned to their original positions two minutes after they were first manipulated unless individuals solved the puzzlebox to the maximum level possible at the time and had received a food reward, in which case the puzzlebox was immediately reset.

(ii) Capuchins

Two, one hour, trials were conducted daily with a total of 53 trials over two time periods (November to December 2007 and June 2008). The first trial was conducted in late morning (starting 10.30-11am) and the second trial in the afternoon (starting 1.30-2.30pm) with no less than 90 minutes between trials. The capuchin group was tested using the scaffolded condition only.

(iii) Children

Each group received five trials of 30 minutes each, with one trial per day, with a space of one to three days between trials. In accordance with the testing context for the non-human primates, the children were allowed to leave the room and return to their classroom at any time. The stickers that individuals collected were placed in an opaque cup that they were allowed to carry with them. This allowed the stickers to be stored in one discrete place, but also allowed limited visual access by other members of the group, making it less likely they could assess the skill of another individual from results alone, in the same manner as non-humans would eat the food rewards they received. Four groups were tested in the open condition and four in the scaffolded condition, with conditions balanced across the three nurseries.

Experiment 2

Four groups of chimpanzees took part in the second experiment. From each of these groups a female was isolated and trained to use the puzzlebox to access stage three reliably, rapidly and consistently. In two groups (*N*=13 and 8) a high-ranking female demonstrator was trained whilst in two groups (*N*=11 and 9) a low-ranking female demonstrator was trained. Females were chosen as demonstrators as they can be isolated more easily and reintroduced to the group with less aggression, and they tend to concentrate for longer during training sessions (Whiten A, Horner V & DeWaal F, 2005. Conformity to cultural norms of tool use in chimpanzees. Nature 437, 737-740). Demonstrators of different rank were used to assess whether there was a difference in the spread of a cumulative innovation depending upon the rank of the 'innovator'.

During demonstrator training, tutee demonstrators observed demonstrations by the experimenter and the trainer at the facility. Rewards were handed to the chimpanzee once the trainer had demonstrated how to get to the stage. In addition further rewards, where necessary, including fruit, yoghurt and peanut butter, were placed on the button and dial of the puzzlebox to scaffold learning. Training sessions never took more than 20 minutes and the animals were then reintroduced carefully back into their groups to avoid any violence towards them. Animals were judged to have learned to use the puzzlebox when they could reach stage three on six successive attempts, for three trials all of which were conducted on different days.

The trials in the second experiment were three hours in duration, each group receiving eight trials, which were randomised between morning (8.30-11.30am) and afternoon (2-5pm) sessions. One trial was conducted per day over two weeks with a space between trials of one to three days. During trials a maximum of one small feed of vegetables and fruit was given by the care staff. This was insufficient to satiate the subjects or distract them for more than approximately five minutes.

In two groups, one with a low-ranking demonstrator and one with a high-ranking demonstrator, rewards were available at all levels for the first four trials and in the subsequent four trials there was food only available at the final stage. In the other two groups rewards were only available at the final stage for the first four trials and were available at all stages for the next four. This reward regime replicated the manipulation of ‘open’ and ‘scaffolded’ conditions in Experiment 1, but within rather than between subjects.

Demonstrator performance

All trained chimpanzee demonstrators solved the task consistently during the open diffusion trials, giving a mean of 150.9 (standard error ± 20.4) demonstrations reaching stage three per trial.

Data Collection

All data were coded from the video taken during the experimental trials. A second observer coded 2% of the data coded in each species. Inter-observer reliabilities were >94% for all recorded behaviour. All occurrences sampling was used to record each time an individual contacted the puzzlebox, and each unsuccessful and successful manipulation of the functionally relevant parts (stage 1-3) of the puzzlebox. Unsuccessful and successful manipulations were defined as those in which an individual did not and did retrieve a food reward, respectively. In each case the identity of the individual interacting with the puzzlebox was recorded as was the identity of the individuals in proximity to the puzzlebox (defined as an area of 1.5m around the puzzlebox) when the events occurred. In addition, the latency at which all individuals arrived and left the area defined as proximity was recorded. Any aggression (defined as any interaction in which one individual struck another, displayed or exhibited an aggression face) or scrounging (defined as one individual removing food from the hand of another individual or from the puzzlebox before the individual who opened the door retrieved it) that took place within the area in proximity was recorded.

Table S4: The definitions of codes and additional clarifications that were coded from the video. Inter-observer reliability was calculated from both the code and additional comments combined.

Code	**Additional comments noted**	**Definition**
Contact	The area of the puzzlebox (e.g. 'left door' or 'top').	An individual touches the puzzlebox, but does not operate any of the moving parts of the puzzlebox.
Unsuccessful manipulation	Right/ left door	An individual opens the right/left door in the two minutes before the food reward has been replaced and therefore receives no food reward.
	Down on right/left The method of pushing the button (i.e. pushing with hands or biting)	An individual pushes on the down button on the right/left after another individual has pressed it, but before it has been reset.
	Up on right/left The method of pushing the button (i.e. pushing with hands or biting)	An individual pushes on the up button on the right/left after another individual has pressed it, but before it has been reset.
	Dial on right/left The method of turning the dial (i.e. red or blue hole)	An individual turns the dial after another individual.
Successful manipulation	Right/ left door. Stage to which door is pushed. Note whether the individual takes the food or not	An individual pushes the door open to reveal a reward.
	Up on right/left. The method of pushing the button (i.e. pushing with hands or biting)	An individual either pushes the up button or bites the button, unlocking the second stage of the puzzlebox.
	Down on right/left. The method of pushing the button (i.e. pushing with hands or biting)	An individual either pushes the down button or bites the button, unlocking the second stage of the puzzlebox.
	Dial on right/left. The method of turning the dial (i.e. red or blue hole)	An individual turns the dial to unlock the third stage of the puzzlebox.
Altruism	Identity of individual that donates reward and individual that receives it.	An individual gives a reward it has obtained from the puzzlebox to another individual.
Aggression	Identity of individual perpetrating aggression and those being attacked.	Any interaction in which one individual strikes another, displays or exhibits an aggression face.
Scrounging	Identity of the scrounger and the victim	An individual removes food from the hand of another individual or from the puzzlebox before the individual who opened the door retrieves it.
Teaching	Method of teaching (i.e. verbal, gestural or a mixture)	An individual produces a gesture or vocalisation (or both) that functions to facilitate learning in another individual by imparting knowledge about the solutions to the puzzlebox.
Vocalisation		*Non-human primates:* an individual produces a food call (as defined in capuchins by Fragaszy et al., 2004 and chimpanzees by Slocombe & Zuberbühler, 2005).
	Children: The words spoken by the individual or a description of the	*Children:* an individual produces a vocalisation, either a verbal or non-verbal.

	vocalisation if non-verbal.	

Analyses

All analyses were carried out using the *R* statistics package (R-Development-Core-Team). The data were tested for normality using a Shapiro's test and non-parametric tests were used only where the assumptions of parametric tests were violated. Below we provide further detail, where necessary, detailing how the eight hypotheses outlined, in the main text, were evaluated.

To allow greater resolution in the assessment of the performance of individuals, rather than analysing data on a 0-3 scale based upon the puzzlebox stage the individual achieved, a species-specific 'achievement rank' was calculated for each individual. The 'achievement rank' ranks individuals first upon the stage that they achieved and differentiates further between individuals by the number of times they successfully manipulated the puzzlebox at that stage. In the case of a tie at this point, the number of successful manipulations performed at previous puzzlebox stages is used to differentiate between the individuals. This has the advantage that it renders the distribution continuous, which is better suited to analyses and affords greater statistical power.

Table S5: Eight alternative hypotheses specifying why humans, but not other animals possess cumulative culture and the extent to which each is supported by comparing the performance of capuchins, chimpanzees and children.

Hypotheses	**Capuchins**	**Chimpanzees**	**Children**	**Hypothesis supported?**
Social Cognition				
1. A lack of teaching in non-human primates hinders ratcheting (*2,6,13*)	No direct teaching events. Mother—juvenile-offspring dyads are significantly more likely to have reached a different stage than	No direct teaching events. Individuals significantly more likely to scrounge from their juvenile offspring than from their mother. No significant difference	Substantive teaching, with a significant correlation between the number of teaching events received and achievement rank.	Supported

	mother —adult-offspring dyads.	between the stage reached by mother—juvenile-offspring dyads and mother—adult-offspring dyads		
2. Communication in non-human primates is not sufficient to support ratcheting (*13*).	Few food calls emitted. No increase in recruitment following calls.	Few food calls emitted. No increase in recruitment following calls.	All instances of teaching involve vocalization. Significant correlation between amount of verbal instruction and achievement rank.	Supported
3. Lack of imitation in non-humans hinders ratcheting (*1,2,6*).	Do not match recently observed actions.	Do not match recently observed actions	Match recently observed actions. Significant correlation between proportion of matching manipulations and achievement rank.	Supported
4. Lack of prosociality in non-humans hinders ratcheting (*2,13*).	No voluntary donation of rewards.	No voluntary donation of rewards.	Frequent voluntary donation of rewards. Significant relationship between gifts received and achievement rank.	Supported
Social Structure				
5. Scrounging, or being scrounged from, hinders learning (*20*).	No correlation between scrounging and achievement rank. Positive correlation between number of times scrounged from and achievement rank.	Positive correlation between scrounging, and number of times scrounged from, and achievement rank.	Positive correlation between scrounging, and number of times scrounged from, and achievement rank.	Not supported
6. Dominants monopolise resources preventing low rankers from gaining access to the task (*17*).	Dominant individuals use the puzzlebox significantly more than low rankers in 2007, but not in 2008.	Low and mid rankers use the puzzlebox significantly more than high rankers.	No significant difference between the number of manipulations performed by low and high rankers	Not supported
7. Lack of attention to low rankers and/or juveniles hinders diffusion (*18,19*).	No significant difference between the amount of attention paid to individuals of different rank or age.	No significant difference between the amount of attention paid to individuals of different rank or age.	No significant difference between the amount of attention paid to individuals of different rank.	Not supported
Non-Social Cognition				
8. Non-human animals are conservative and satisfice (*8,16*).	Individuals perform a significant number of non-conservative manipulations.	Receiving rewards at all stages does not hinder performance relative to scaffolded condition. Individuals perform a significant number of non-conservative manipulations.	Receiving rewards at all stages does not hinder performance relative to scaffolded condition. Individuals perform a significant number of non-conservative manipulations.	Not supported

Additional methods for hypothesis testing

Hypothesis 1

A lack of teaching in non-human primates hinders the spread of cumulative innovations throughout the population

We defined teaching by direct instruction as ‘any instance in which an individual engaged in an act that clearly functioned to facilitate learning in another individual', in this instance by imparting knowledge about the solutions to the puzzlebox task. In the capuchins and chimpanzees we went on to consider more subtle forms of 'teaching', such as *scaffolding*, defined as facilitating learning in others through acting in a manner that functions to draw attention to the task or rewards, or create learning opportunities for others. We specifically considered scaffolding afforded by tolerated theft by comparing the frequency of food transfer from mothers to juveniles to that from juveniles to mothers.

Hypothesis 2

Communication insufficient to support ratcheting

With regard to the analysis of the recruitment potential of food-calls, we computed the rate (arrivals/min) of animals entering proximity to the task in the two minutes following an individual in proximity emitting a food call, and compared this to the baseline rate of individuals entering proximity throughout the trial. In children, we compared the success of individuals who had received verbal instruction with those that had not.

Hypothesis 3

Lack of imitation or other complex forms of social learning in non-humans hinders the spread of cumulative innovations throughout the population

To test whether observational learning played any role in the acquisition of solutions to the puzzlebox, we examined whether individuals manipulated the box in a matching manner, either because they copied the actions of others at the puzzlebox (i.e. imitation) or because they made the same parts of the box move in the same way (object-movement re-enactment, emulation). As physical access to the puzzlebox was often blocked when other individuals were interacting with it, the analysis determined whether an individual matched the manipulations of another individual who had been manipulating the puzzlebox immediately prior to their manipulation. As there was little progression beyond stage 1 in experiment 1 with the chimpanzees, this analysis was carried out using data from the second experiment, utilising those occasions when a skilled demonstrator left the proximity of the task and another individual manipulated the puzzlebox, provided both the demonstrator and observer had been in proximity to the puzzlebox for at least a minute. For the capuchins, analysis focussed on occasions where individuals skilled at stage two left the puzzlebox, after having been observed by another individual in proximity for at least one minute, and who went on to contact the puzzlebox in the subsequent minute. As children left the puzzlebox less frequently than other species, all instances of skilled children leaving the puzzlebox were considered until a time at which all individuals in the group had learned to open stage three. Once again, we focused on occasions where the first child had been observed by another child in proximity for at least one minute, and where the second child went on to contact the puzzlebox in the subsequent minute. In all cases, all classes of manipulations by the 'demonstrator' (e.g. slide left door to left, push left upper button etc) were recorded in the minute preceding it leaving the puzzlebox, and all manipulations by the observer in the subsequent minute were recorded. Those manipulations that matched those performed by the demonstrator were classified as 'matching', while those that had not been performed by the demonstrator were classified as 'non-matching' (Table S6).

Table S6: Actions performed by a demonstrator and the actions that were classed as matching if performed by an observer after observing that demonstrators action. All other actions were classified as non-matching.

Demonstrator's action	**Matching actions**
Contact puzzlebox (+ location on box touched)	Contact puzzlebox (+ same location on puzzlebox)
Unsuccessful/Successful right door	Contact right door: *touches but does not move door*
	Unsuccessful right door: *opens right door before it has been reset*
	Successful right door: *opens right door*
Unsuccessful/Successful left door	Contact left door: *touches left door but does not move door*
	Unsuccessful left door: *opens left door before it has been reset*
	Successful left door: *opens left door*
Unsuccessful/Successful down button on right (+method of pushing the button- i.e. pushing button with hands or biting)	Contact down on right (+ same method of pushing the button): *touches but does not move down button on right, using the same method*
	Unsuccessful down on right (+ same method of pushing the button): *pushes down on right, but before it has been reset, using same method*
	Successful down on right (+ same method of pushing the button): *pushes down on right, using the same method*
Unsuccessful/Successful down button on left (+method of pushing the button- i.e. pushing button with hands or biting)	Contact down on left (+ same method of pushing the button): *touches but does not move down button on left, using the same method*
	Unsuccessful down on left (+ same method of pushing the button): *pushes down on left, but before it has been reset, using same method*
	Successful down on left (+ same method of pushing the button): *pushes down on left, using the same method*
Unsuccessful/Successful up button on right (+method of pushing the button- i.e. pushing button with hands or biting)	Contact up on right (+ same method of pushing the button): *touches but does not move up button on right, using the same method*
	Unsuccessful up on right (+ same method of pushing the button): *pushes up on right, but before it has been reset, using same method*
	Successful up on right (+ same method of pushing the button): *pushes up on right, using the same method*
Unsuccessful/Successful up button on left (+method of pushing the button- i.e. pushing button with hands or biting)	Contact up on left (+ same method of pushing the button): *touches but does not move up button on left, using the same method*
	Unsuccessful up on left (+ same method of pushing the button): *pushes up on left, but before it has been reset, using same method*
	Successful up on left (+ same method of pushing the button): *pushes up on left, using the same method*
Unsuccessful/Successful dial on right (+method of turning the dial- i.e. red or blue hole)	Contact dial on right (+same method of turning the dial): *touches but does not move the dial on right*
	Unsuccessful dial on right (+same method of turning the dial): *turns dial on right after*

	another individual, using the same method
	Successful dial on right (+same method of turning the dial): *turns dial on right after another individual, using the same method*
Unsuccessful/Successful dial on left (+method of turning the dial- i.e. red or blue hole)	Contact dial on left (+same method of turning the dial): *touches but does not move the dial on left*
	Unsuccessful dial on left (+same method of turning the dial): *turns dial on left after another individual, using the same method*
	Successful dial on left (+same method of turning the dial): *turns dial on left after another individual, using the same method*

To measure whether social learning was occurring at the first stage, we used option-bias analysis (Kendal, R.L., Kendal, J.R., Hoppitt, W. & Laland, K.N. 2009. Identifying Social Learning in Animal Populations: A New 'Option-Bias' Method. *PLoSOne* 4(8): e6541) at the level of opening left door or opening right door, testing whether individuals in a group were more likely to use one option (opening one door), more than the other, which is likely to occur if social learning is occurring. This method is more powerful than conventional inferential statistics (Kendal et al., 2009).

Hypothesis 4

Lack of prosociality in non-humans hinders the spread of cumulative cultural traits

We recorded the number of altruistic events performed by each individual, defining an altruistic event as any instance in which an individual voluntarily gives a reward of any stage, accessed by themselves, to another individual.

Hypothesis 6 and 7

Dominant individuals monopolise resources hindering lower ranking individuals from gaining access, thereby limiting the number of individuals with the chance to *solve the task. Lack of attention to low-ranking and/or juvenile individuals hinders learning from potentially skilled sections of the population*

Individuals were divided into rank categories, high, medium and low for chimpanzees and capuchins and high and low for children. For chimpanzees ranks were based upon data that had been previously gathered on aggression during reintroductions and on feeding priority. Capuchin data were gathered on displacement rates at a single monopolisable food source. Child data were gathered by asking teachers to rank pupils on a scale of most socially dominant-least socially dominant and bold-shy.

Hypothesis 8

Non-human animals are conservative and satisfice, such that once they have a solution that rewards them they do not change it

We compared the performance of individuals in the open and scaffolded conditions in experiment 1, in both the children and chimpanzees. (As we only had access to one capuchin group, we were unable to make this comparison in the capuchins). We reasoned that, if individuals do satisfice then individuals in the scaffolded condition should manipulate the puzzlebox more at higher stages than individuals in the open condition, since the latter would still be receiving rewards at the lowest stage, and be inhibited from further learning. Expectations for the rate of manipulating each part of the apparatus were derived from performance in early trials. For the chimpanzees, the manipulations in the first three trials after the scaffolded groups had stopped receiving a reward at a lower stage were compared to the same time period in the open condition. For the children, the first 10 minutes of the trial in which individuals in the scaffolded groups did not receive rewards, at the lower stages, any more were compared to the same time period in the open condition.

Additional Results

General performance

In experiment 1, following 30 hours of presentation of the cumulative puzzlebox, in 1-hr sessions, to each of the four groups of chimpanzees, only a single individual in a single group reached stage 3. In the same group and one other group, a single individual reached stage 2, in a third group two individuals reached stage 2, whilst the remaining group witnessed multiple solvers at stage 1, but not at higher levels. Likewise, in groups with trained demonstrators (experiment 2), although multiple individuals solved stage 1, the solutions to stages 2&3 did not spread. Thus the experiments provide no evidence for cumulative cultural learning in any chimpanzee group, including in experiment 2, where trained demonstrators performed stages 1-3 proficiently. The chimpanzees were clearly capable of solving the apparatus at higher stages (stage 2-3), as witnessed by the performance of innovative individuals in three groups, as well as the trained demonstrators, but in no group is there any evidence that these solutions spread to a second individual. A virtually identical pattern is observed in the capuchins, where after 53 hours (year 1: 28, year 2: 25), no individual reached stage 3, whilst only two individuals reached stage 2, and the majority of individuals solved only stage 1. These findings stand in stark contrast to those of the children, where despite a far shorter exposure to the apparatus (2.5 hours), five of the eight groups had at least two individuals (out of maximum 5) who reached stage 3, with multiple solvers at stage 2 in all these groups, providing clear and strong evidence for a cumulative cultural capability. Of the groups not reaching stages 2 & 3, two expressed little interest in the box, whilst in the third the children initially exhibited interest, leading to widespread stage 1 solutions, before interest waned (see below for discussion).

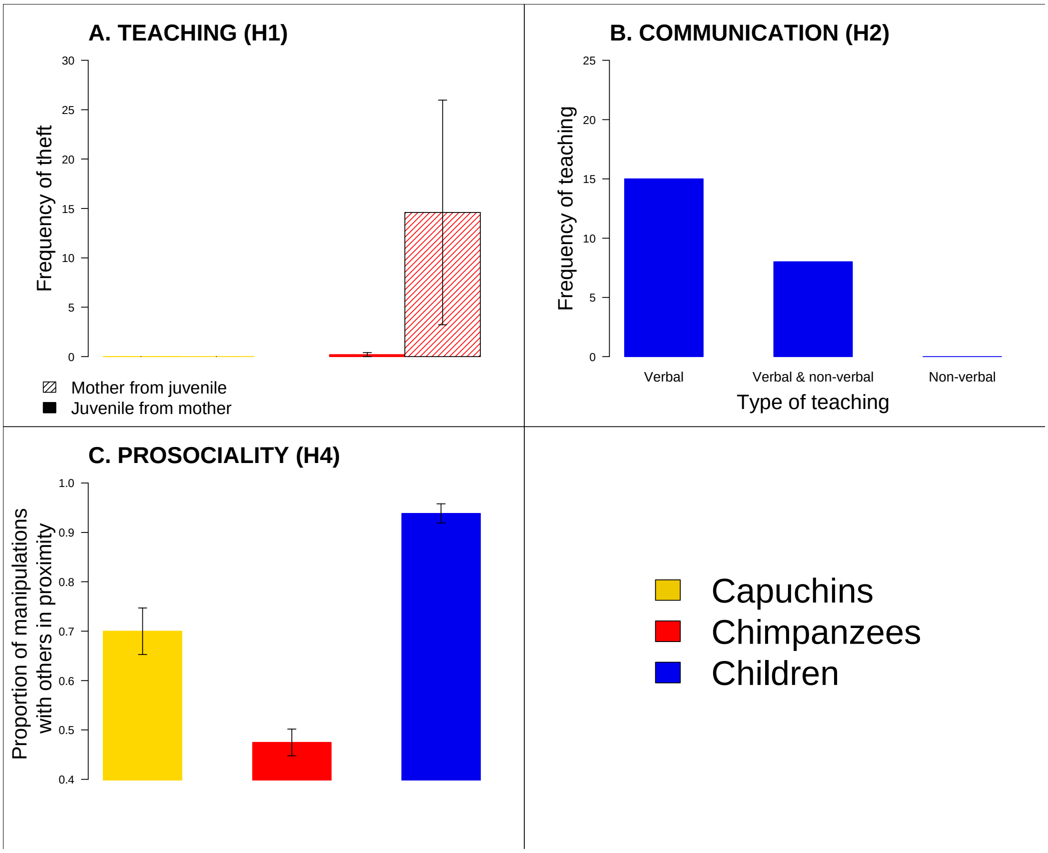
Hypothesis 1

Figure S2A shows that we observed substantially greater rates of tolerated theft of extracted food by mothers from offspring than vice-versa in chimpanzees (*Wilcoxon* W=16, P=0.026) and no tolerated theft in mother-infant pairs of capuchins.

Hypothesis 2

All teaching events by children involved verbal instruction and approximately one third involved gesture (Figure S2B).

Figure S2. Additional analyses.



Hypothesis 3

Using the option-bias method, there was no evidence of social learning of door choice (left vs right) in capuchins (*Option bias* χ^2= 546.5, *P*=1). However, there was evidence of social learning by chimpanzees in experiment one at the level of door choice, that is, stage 1 (*Option bias* χ^2= 941.6, *P*= 0.021). Combined with the lack of evidence for cumulative cultural learning in chimpanzees, and the low levels of matching at higher stages, these findings support the view that chimpanzees are capable of social transmission but not cumulative culture.

Hypothesis 4

A greater proportion of the manipulations by children were at the same time as another individual than either chimpanzees or capuchins (*Kruskal-Wallis* χ^2=39.56, *df*=2, *P*<.001; Figure S2C).

Hypothesis 5

There was no evidence that scrounging negatively affected the performance of either those individuals scrounging or those that were victims of scrounging. In capuchins there was no significant correlation between the number of times an individual scrounged from another and their achievement rank (*Spearman's Rank Correlation*: *rho*=0.34, *S*=1170.5, *P*=0.12). Achievement rank in chimpanzees was positively correlated with the number of scrounging events an individual perpetrated (*Spearman's Rank Correlation*: *rho*= 0.41, *S*=35466.2, *P*=0.0005). The achievement rank of children was significantly positively correlated with the number of times an individual scrounged from others (*Spearman's Rank Correlation*: *rho*= 0.84, *S*=1165.90, *P*<0.001).

Hypothesis 8:

We found no evidence that non-human animals are conservative and satisfice, such that once they have a solution that rewards them they do not change it. There were two capuchins that got to stage 2, thus suggesting that not all individuals act conservatively (i.e. remained at level 1). Across the entire population the number of non-conservative manipulations (that is, manipulations different from the first solution) performed by individuals (mean= 39.94 standard error= 22.41) was significantly different to zero (*Mann-Whitney test: U*=78, *P*=0.002). Likewise, analysing whether chimpanzees act conservatively after they have learned to get to the first stage reveals that individuals do not always act conservatively. Across the populations the number of non-conservative manipulations performed by individuals (mean= 76.71 non-conservative actions, standard error= 42.37) was significantly different to zero (*Mann-Whitney test:* U=253, P<0.001).

Children did performed a significantly larger proportion of non-conservative actions (mean=0.34) than chimpanzees (mean=0.06) or capuchins (mean=0.18) (*Kruskal-Wallis:* χ^2=6.60, *df*=2, P=0.037). Whilst this might be interpreted as a difference in the conservative tendencies of the three species, other interpretations are possible. For instance, the elevated number of non-conservative actions performed by the children likely represents their elevated performance in general, which requires a degree of non-conservative behaviour, and hence may be attributable to the socio-cognitive processes discussed in the main text.

Results indicating that capuchins and chimpanzees recognised that the higher quality resources were superior to the lower quality resources.

In the pre-experiment food preference trial, capuchins showed a clear preference for grapes over apples and over carrots. During the trials a higher proportion of stage one rewards (carrot) were able to be scrounged than stage two (apple) rewards (*Wilcoxon* W=103, P=0.003).

Pre-trial testing revealed that chimpanzees preferred grapes to apples and apples to carrots. This supplemented other sources which also concluded this order of food preference (Brosnan, pers. comm.). During the trials there was no significant difference between the proportion of food that individuals allowed to be scrounged at each stage (*Kruskal Wallis:* χ^2=1.05, *df*=2, *P*=0.59). There were 29 instances of 'termiting' behaviour in which individuals probed the olfactory holes in the puzzlebox doors with small sticks or grass. There was at least one instance of this behaviour in seven of the eight groups (mean=3.63 instances per group, standard error= 1.16), with all instances occurring at the highest stage that was stocked with food.

Pre-trial testing revealed that children consistently preferred smiley face stickers to large stars to small stars. During the trials the proportions of rewards found that were stolen at stage one (mean=0.20, standard error=0.031) and stage two (mean=0.17, standard error=0.027) were significantly greater to the proportion of the rewards found that were stolen at stage three (mean=0.09, standard error=0.03) (*Kruskal-Wallis:* χ^2= 6.88, *df*= 2, *P*=0.032).

Failure of 2 groups of children to interact with the cumulative task

There was a notable finding with the children that in one group of children no participants solved the puzzlebox and another group, in the scaffolded condition, did not qualify as having solved the first stage to progress to the second stage. These results contrast markedly with other species in which all but one of the chimpanzees and 15 out of 22 capuchins learned to solve stage one. Shyness in children of an unfamiliar experimenter and neophobia of the puzzlebox may partially account for the lack of manipulations in some individuals, but are unlikely to account for a group-level effect. In contrast to the chimpanzees and capuchins studied, who live in colonies that regularly take part in a range of extractive foraging experiments the children had not taken part in similar experiments. Whilst shyness or neophobia are individual traits, a group conformity effect may operate, whereby if one child does not step forward and operate the puzzlebox, others will also refrain from doing so, and/or anxiety may spread socially. This lack of solving in these two groups of children may, therefore, also be due to the same socio-cognitive processes responsible for the increased ability to solve the puzzlebox, with children operating as a group and observing the performance of other individuals around them.

This conclusion is supported by the observation that, in one of the groups, children engaged in a game, which was invented after one child dropped the cup he had been given to store stickers during the first trial. This became known as the ‘cup game’ among the group and consisted of throwing the cup, following it and recovering it. Following its invention the game spread to all group members, distracting them from the puzzlebox and providing a new social activity during the trials.