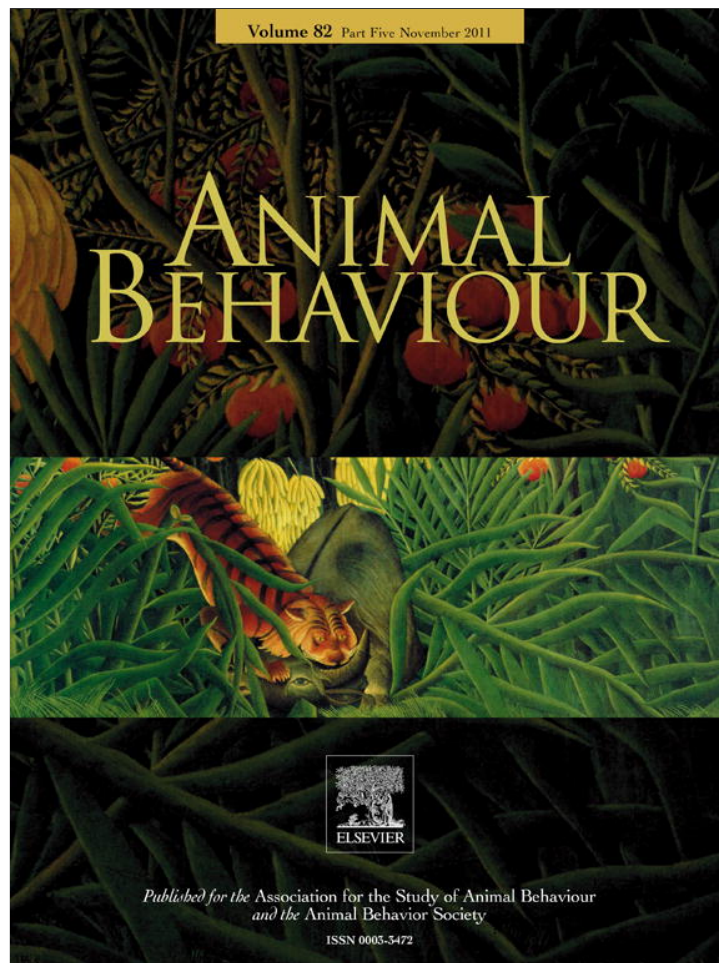


Provided for non-commercial research and education use.  
Not for reproduction, distribution or commercial use.



This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

<http://www.elsevier.com/copyright>



Contents lists available at ScienceDirect

## Animal Behaviour

journal homepage: [www.elsevier.com/locate/anbehav](http://www.elsevier.com/locate/anbehav)

## Sometimes tool use is not the key: no evidence for cognitive adaptive specializations in tool-using woodpecker finches

I. Teschke<sup>a,1</sup>, E. A. Cartmill<sup>b,2</sup>, S. Stankewitz<sup>c,3</sup>, S. Tebbich<sup>d,\*</sup>

<sup>a</sup> Department of Behavioural Ecology and Evolutionary Genetics, Max Planck Institute for Ornithology, Seewiesen

<sup>b</sup> Department of Psychology, University of Chicago

<sup>c</sup> Department of Animal Behaviour, Freie Universität Berlin

<sup>d</sup> Department of Cognitive Biology, University of Vienna

### ARTICLE INFO

#### Article history:

Received 11 March 2011

Initial acceptance 21 April 2011

Final acceptance 21 July 2011

Available online 3 September 2011

MS. number: 11-00229

#### Keywords:

adaptive specialization

*Cactospiza pallida*

Darwin's finch

physical cognition

problem solving

reversal learning

tool use

The use and manufacture of tools has been considered to be cognitively demanding and thus a possible evolutionary driving factor of intelligence. Animal tool use provides the opportunity to investigate whether the use of tools evolved in conjunction with enhanced physical cognitive abilities. However, success in physical tasks may simply reflect enhanced general learning abilities and not cognitive adaptations to tool use. To distinguish between these possibilities, we compared general learning and physical cognitive abilities between the tool-using woodpecker finch, *Cactospiza pallida*, and its close relative, the small tree finch, *Camarhynchus parvulus*. Since not all woodpecker finches use tools, we also compared tool-using and nontool-using individuals, predicting that domain-specific experience should lead tool-using woodpecker finches to outperform nontool-users in a task that is similar to their natural tool use. Contrary to our predictions, woodpecker finches did not outperform small tree finches in either of the physical tasks and excelled in only one of the general learning tasks, and tool-using woodpecker finches did not outperform nontool-using woodpecker finches in the physical task closely resembling tool use. Our results provide no evidence that tool use in woodpecker finches has evolved in conjunction with enhanced physical cognition or that domain-specific experience hones domain-specific skills. This is an important contribution to a growing body of evidence indicating that animal tool use, even that which seems complex, does not necessitate specialized cognitive adaptations.

© 2011 The Association for the Study of Animal Behaviour. Published by Elsevier Ltd. All rights reserved.

One of the fundamental issues in the study of human and animal cognition concerns the factors that drove the evolution of intelligence. The ability to use tools has played a long-standing role in this debate because it is thought to be cognitively demanding (e.g. Parker & Gibson 1977; but see also Hansell & Ruxton 2008). If tool use poses a cognitive challenge and provides a selective advantage in acquiring resources, then we expect that enhanced cognitive abilities should have evolved to increase the effectiveness of tool use or to enable the development of such an unusual technique in the first place. The notion that tool use should be linked to cognition related to learning

about the physical world, for example the spatial relationships between objects, connectivity and weight, has been particularly popular among cognitive biologists and has generated many studies of tool-related cognition in primates (Hauser et al. 1999; Povinelli 2000; Santos et al. 2006; Martin-Ordas et al. 2008; Seed et al. 2009; reviewed in Visalberghi & Tomasello 1998) and birds (e.g. Auersperg et al. 2009; reviewed in Kacelnik et al. 2006; Seed et al. 2006; von Bayern et al. 2009; Emery & Clayton 2009; Taylor et al. 2009; Wimpenny et al. 2009).

Some of the strongest evidence supporting the notion that tool use is linked to enhanced cognition is the strong correlation between tool use and brain size in birds (Lefebvre et al. 2002) and primates (Reader & Laland 2002) as well as the demonstration that the cerebellum of tool-using birds has a higher degree of folding than that of nontool-using birds (Iwaniuk et al. 2009). However, these findings provide only an indirect demonstration of the connection between tool use and enhanced cognition and must be confirmed by behavioural experiments conducted in a comparative framework.

Several comparative methods are available to demonstrate that a given character, behaviour or cognitive trait is related to an

\* Correspondence: S. Tebbich, Department of Cognitive Biology, University of Vienna, Althanstrasse 14, Vienna 1090, Austria.

E-mail address: [sabine.tebbich@univie.ac.at](mailto:sabine.tebbich@univie.ac.at) (S. Tebbich).

<sup>1</sup> I. Teschke is at the Department of Behavioural Ecology and Evolutionary Genetics, Max Planck Institute for Ornithology, Eberhard-Gwinner Str. 11, 82319 Seewiesen, Germany

<sup>2</sup> E. Cartmill is at the Department of Psychology, University of Chicago, 5848 S. University Avenue Box 6, Chicago, IL 60637, U.S.A.

<sup>3</sup> S. Stankewitz is at the Department of Animal Behaviour, Freie Universität Berlin, Takustr. 6, 14195 Berlin, Germany

ecological condition (e.g. Harvey & Pagel 1991; Shettleworth 1998). Traditionally, two approaches have been taken: the search for character divergence in closely related species under differing selective conditions and the search for convergence among unrelated species resulting from similar selective conditions. While these approaches are both necessary and informative by themselves, it is only through the synthesis of the two and by gathering repeated observations of the same pattern of convergence and divergence in several evolutionarily distinct groups that we can possibly begin to understand the true relationship between tool use and physical cognition.

The woodpecker finch, *Cactospiza pallida*, is a tool-using species that habitually uses twigs or cactus spines to poke arthropods out of tree holes (Eibl-Eibesfeldt 1961). However, not all adult woodpecker finches use tools in the wild: the tool-using ability and ecological relevance of this behaviour varies with habitat and food abundance. Differences in the ecological relevance of tool use can be attributed to differences in seasonal fluctuations in food availability: woodpecker finches obtain about half of their prey using tools in arid habitats during the dry season while those birds inhabiting humid areas hardly ever use tools (Tebbich et al. 2002). The variation in the ability to use tools in natural populations covaries with the ecological relevance of this behaviour in different habitats and seems to be a function of learning opportunities early in ontogeny. Tebbich et al. (2001) conducted a study on social learning in woodpecker finches and one of the main findings was that all juvenile woodpecker finches (but not adults) developed some form of tool use even without opportunities for socially learning the technique, indicating that this ability is based on a specific genetic predisposition and is not dependent on social learning. The juveniles from this study were from a humid habitat where only a low percentage of tool-using woodpecker finches can be observed in the wild compared to dry areas where all individuals are capable of using tools. This indicates that given the right conditions, woodpecker finches from both dry and humid habitats have a predisposition to learn to use tools. Possibly the differences in tool-using ability that arise in adulthood are due to different learning opportunities in these respective habitats.

In this study, we compared cognition in woodpecker finches, a species that habitually uses tools, and the closely related nontool-using small tree finch, *Camarhynchus parvulus*. These two Darwin's finch species are both members of the tree finch clade within the Darwin's finches and are sympatric throughout Santa Cruz Island of the Galápagos archipelago; these similarities minimize the influence of contextual variables as explanations for differences in cognitive abilities. The main objective of this study was to see whether this species pair shows a divergence in cognitive abilities apart from the genetic predisposition to acquire tool use. Such cognitive divergence could be limited to evaluation of inputs from the domain of tool use, for example, appreciation of the functional relationship between a tool and the object that it moves.

To test the hypothesis that tool use in woodpecker finches coevolved with enhanced physical cognitive abilities, we compared their performance with that of small tree finches in two tasks designed to test different aspects of physical cognition. We define physical tasks as problems that can be learned only by discriminating a cue pertaining to physical relationships (spatial relationships, contact, surface continuity). The physical tasks that we used can be assigned to two subcategories. One task was a physical problem specific to the use of tools and required the subject to pull one of two hooks to obtain a food reward. The task aimed to test the bird's ability to learn about the necessary spatial relationship between the hook tool and the reward. This was deemed to be a problem specific to the use of tools because it required the subject to manipulate a tool appropriately to obtain a reward and to be sensitive to the appropriate relationship between the tool and reward: both are challenges

that are presumably posed during the natural tool use of woodpecker finches. The other task was a more general physical problem that did not require the manipulation of a tool with the beak but nevertheless required sensitivity to a physical cue, surface continuity, to solve the entire task series. Here, birds had to avoid moving a food reward over a discontinuous surface, because this would have caused the food to drop into the surface where it was inaccessible. We consider this to be a more general physical problem because sensitivity to the physical cue involved is probably not essential for success in natural tool use. Testing physical cognition with paradigms that involve physical problems specific to tool use and problems that are more general facilitates more precise specification of the level of adaptation, allowing us to determine whether tool use evolved in conjunction with a general increase in physical cognition, or whether the cognitive adaptations are specific to problems related to the use of tools.

Those subjects that solved the initial problem posed by either of the physical tasks were subsequently tested in a series of transfer tasks to see whether they could transfer acquired knowledge to novel situations, an ability that is highly relevant in their natural feeding ecology. Typically, the goal of such transfer tasks is to differentiate dichotomously between high- and low-level cognitive solutions to a given problem. Low-level cognitive solutions generally only entail the use of simple perceptual features of a problem while high-level cognitive strategies involve mental representation of the underlying physical properties and forces involved in the problem in an abstract and conceptual manner (Seed & Byrne 2010). However, in previous studies that have adopted this approach, only a very small number of the tested animals (if any) succeeded in solving the entire task series. While this is useful in staking out the cognitive limits of a species, the low number of successful individuals also limits the opportunity to make quantitative comparisons. Thus, here we deliberately designed simple transfer task series which did not systematically preclude the exclusive use of a procedural rule.

It is also possible that tool use may have evolved in conjunction with general cognitive abilities leading to a domain-general enhancement of learning or that neither general nor specialized cognitive adaptations evolved along with tool use. Therefore, in addition to the physical tasks, we also tested birds in two general learning tasks: one tested the ability to unlearn a previously learned association while the other tested performance in a novel operant task. The classification of these problems is rooted in the rationale that ultimate success in these problems is not dependent on discrimination based on physical relationships between elements of the problem but rather on trial-and-error learning (box-opening task) or colour discrimination (reversal task). These two tasks provided us with clues as to what extent general learning abilities might differ between species and how this might fit into an explanation of the inter- and intraspecific patterns found in the specialized physical tasks. Some of the results of the species comparison in the physical and general learning tasks have been summarized in a different context, as part of a larger data set to support our argument that the fast radiation of Darwin's finches and high number of unusual feeding adaptations found in this group could be the result of a highly flexible stem species of the clade (see Tebbich et al. 2010).

In the present paper we test the hypothesis that tool use evolved with enhanced physical cognitive abilities and present more detailed analysis informing this specific question. We also present results shedding light on an entirely new question, namely on whether experience with the use of tools hones specialized cognitive abilities in woodpecker finches. It is known that domain-specific experience improves performance in tasks testing domain-specific tool-related cognitive abilities in some nonhuman primates (Hauser et al. 2002; Spaulding & Hauser 2005). Since not all woodpecker finches acquire the technique of tool use in their lifetime, we were presented with a unique opportunity to investigate how domain-specific

experience with tools influences the physical cognitive abilities within this species. We used this natural variation to tease apart species competence and individual ability honed by experience in physical task performance. This allowed us to investigate both the species comparison and the effect of ontogeny on physical cognition and formulate specific predictions about the relationship between tool use and cognitive abilities. We predicted that if there is a cognitive adaptation that evolved in conjunction with tool use in woodpecker finches, then both those woodpecker finches with and without tool-using experience should outperform small tree finches and furthermore, if experience hones physical cognition, then tool-using woodpecker finches should perform best of all. In particular, we expected experience to enhance the tool-using woodpecker finches' performance in the cane task which is most relevant to natural tool use. Additionally, if the adaptation is specialized for physical cognition, then we predicted that all test groups should perform similarly in the general learning tasks. However, if there is no cognitive adaptation that evolved with tool use in woodpecker finches, then we would expect both woodpecker finch groups to perform similarly to small tree finches. Finally, it is possible that experience alone might improve the performance of tool-using woodpecker finches in which case we would expect only tool-using woodpecker finches to outperform small tree finches in physical tasks but not nontool-using woodpecker finches.

To our knowledge, this is the first fully controlled comparison of physical- and general cognitive abilities between closely related tool-using and nontool-using species in which both species have been tested in parallel using identical experimental paradigms. This study will permit us to draw conclusions about the interactions between tool use and cognitive abilities on both ontogenetic and evolutionary timescales.

## GENERAL METHODS

### *Study Area, Subjects and Housing*

The study was carried out at the Charles Darwin Research Station on Santa Cruz Island in the Galápagos Archipelago, Ecuador from October 2007 to March 2008 and September 2008 to January 2009. A total of 18 woodpecker finches and 16 small tree finches were mist-netted for this study. Following capture, finches were first kept in a small habituation cage (0.5 × 0.5 m and 1 m high) for up to 5 days. Thereafter, the birds were maintained in outdoor aviaries (3.9 × 2 m and 3 m high or 2 × 1 m and 2 m high). Aviaries were furnished with natural branches and an experimental table on which the apparatus was presented.

Birds were kept singly and visually isolated from each other on a diet of mashed hardboiled egg, grated carrot, mixed with commercial bird food mix (Orlux). Additionally, the birds received fresh fruit and fresh moths daily following testing. Not all birds participated in all experiments. A summary clarifying the order of experiments and the participation of each bird in each experiment is given in [Appendix Table A1](#). Tool-using ability of woodpecker finches was always assessed prior to participation in experiments ([Appendix 1](#)).

### *Basic Experimental Procedure*

Experiments were conducted in the home aviaries of the birds and food was removed from their aviaries 2 h before testing. The apparatus was always baited out of sight of the subject and, for each trial, placed onto the experimental table within the home aviary. The experimenter then left the room and observed the trial via a camcorder (JVC GZ-MG130EK Hard disk camcorder). All experiments except experiment 2 were recorded with the camcorder.

### *Ethical Note*

Permission to conduct this study was granted by the Galápagos National Park and the Charles Darwin research station (Project PC-16-07, Permit No. PR.PT.P004.R02). Birds were caught during the nonbreeding season and were held individually. Individual holding was deemed not to be stressful since these birds are solitary outside of the breeding season. Eight of the woodpecker finches were held in long-term captivity ( $\geq 1$  year) for breeding purposes related to conservation. All other birds were held for the minimum amount of time required to complete the experiments, and then released at their site of capture. Subjects were kept at 100% of their free-feeding weights by monitoring weight every 3 days and adjusting each individual's diet accordingly. Water was available ad libitum for drinking and bathing. To assess the impact of extended periods in captivity on the wellbeing of the birds, we radiotracked eight of the birds upon release and one additional bird that had spent no time in captivity over a 10-day to 2-week period. This also served as a pilot test of the utility of this method for use in a planned translocation of the critically endangered and closely related mangrove finch, *Cactospiza heliobates*. Transmitters were made by Holohil Systems Ltd. (Ontario, Canada; Model LB-2N) and weighed 0.42 g which is approximately 2% of the bird's average body weight and thus well below the widely used 5% guideline (Cochran 1980). We were able to observe three birds from the wet zone, which had spent a year or more in captivity, for more than 10 days. These birds quickly resumed feeding and territorial behaviours such as singing and nest building, suggesting that they readjusted well following release. The remaining birds either removed the transmitter or could not be relocated shortly after release. For two birds, we could confirm that the transmitter had been removed since we were able to relocate the discarded transmitters.

### *Experiment 1: The Reversal Task (General Learning)*

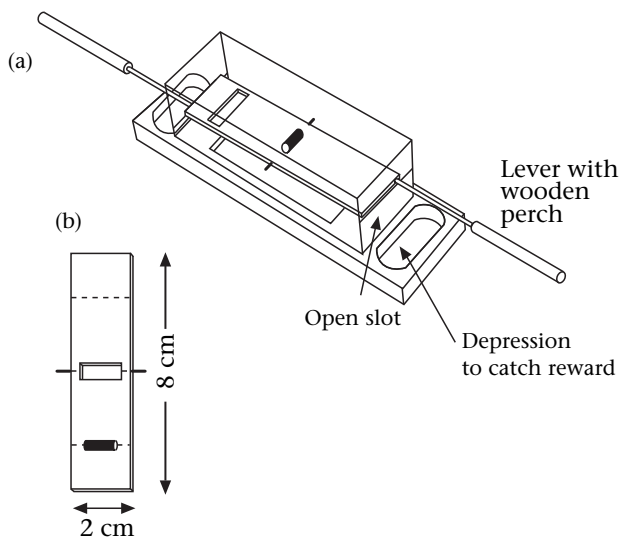
This experiment consisted of two phases: an initial 'acquisition phase' and a 'reversal phase'. In the acquisition phase, subjects were given a choice between two lids of different colours (orange and blue), one of which was the rewarded S+ stimulus. Once a subject met criterion (see General experimental procedure) in the initial colour discrimination, the colour–reward contingency was reversed in the reversal phase.

The apparatus consisted of two feeders covered with coloured lids that were mounted 10 cm apart on a wooden base. In each trial a reward was placed in one of the feeders, the lids were placed on the feeders and birds were then allowed to remove one of the two lids. A transparent Perspex divider prevented the birds from removing the lid of both feeders. Eight small tree finches, six nontool-using and 10 tool-using woodpecker finches were tested in this experiment. In the acquisition phase of the reversal task, one nontool-using woodpecker finch was mistakenly stopped after not having reached full criterion (see [Results](#): logistic regression line with negative slope in [Fig. 3a](#), middle row). Since this did not affect our main conclusions, we retained the data point in the analysis.

### *Experiment 2: The Seesaw Task (Physical Nontool Use)*

This task was designed to test sensitivity to surface continuity. However, the initial task could also be solved using the spatial relationship between elements of the task. The apparatus consisted of a spring-loaded, horizontal seesaw platform encased in a clear Perspex box ([Fig. 1](#)).

Perch-levers were attached to the left and right side of the seesaw so that the platform could be tilted by perching on either lever. A reward encased in clear plastic tubing rested on the seesaw at the



**Figure 1.** The seesaw apparatus depicted in (a) the Food-central condition and (b) the Gap-central condition.

beginning of each trial. Perching on either of the levers caused the reward to roll down the length of the platform towards the bird (Supplementary movie S1). The seesaw platform always contained a hole (gap) and an error was made when the bird tilted the platform so that the food rolled into the hole and was inaccessible. In contrast, birds were rewarded when they perched on the side that caused the food to roll down the continuous portion of the seesaw platform and out of the apparatus. Six small tree finches, six nontool-using and six tool-using woodpecker finches participated in this experiment.

#### Initial task

There were two conditions in this experiment: Food-central and Gap-central (Fig. 1a, b). The initial test condition was balanced according to the three test groups (small tree finches, tool-using and nontool-using woodpecker finches). Upon solving the initial task, a bird was subsequently presented with the unfamiliar condition as the transfer task.

#### Apparatus malfunctions

In 2.7% of trials during the initial task, the apparatus malfunctioned, meaning that the birds were sometimes not rewarded when they should have been or were rewarded when they should not have been (Appendix Table A2). In most cases extra trials were given to the birds that experienced such problems as compensation. Nevertheless, we tested whether the occurrence of apparatus malfunction was significantly higher in any one group and whether there was a correlation between proportional success and proportional occurrences of apparatus malfunction using R version 2.9.1 (R Development Core Team 2009; Table A2 contains the raw data used for this analysis). Significance in the former test would have indicated that the malfunction occurrences were unevenly distributed between groups. This would have been problematic if a high number of malfunction trials occurred mainly in the one group that did not learn the task, since then we would not know whether the inability for this group to learn was an artefact of unequally distributed malfunction trials. Significance in the latter comparison would have indicated that the number of malfunction trials was negatively related to success: the more malfunction trials a bird was exposed to, the less it was able to learn about the task.

Although we found no significant relationship between malfunction rate and success (see Results), we excluded trials in which

the apparatus malfunctioned from the GLMM analyses detailed below.

#### Experiment 3: The Cane Task (Physical Tool Use)

This experiment involved food retrieval contingent on making a choice between two canes, only one of which could retrieve the reward, and was designed to investigate subjects' sensitivity to the functional relationship between a reward and the tool used to obtain it. In each condition, one food reward was inside the hooked portion of the cane and one outside of it (with the exception of transfer task 4 in which both rewards were inside the hooked portion of both canes) and the reward could only be retrieved by pulling the cane in the correct functional relationship with the food (Fig. 2 and Supplementary movie S1).

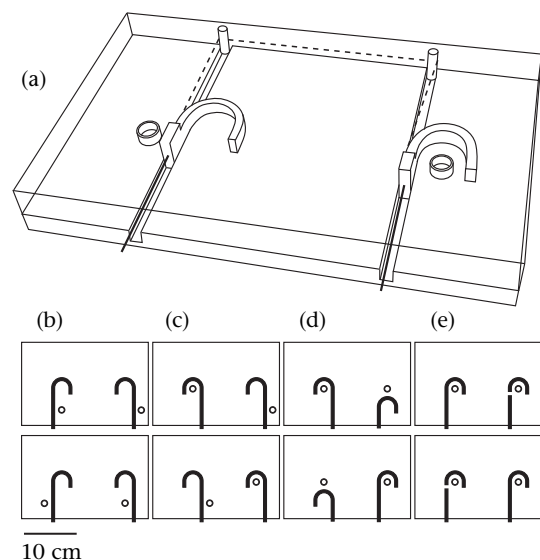
Each cane was fitted into a groove so that they could only be pulled forward in a straight line and could not be flipped. Also, the canes were attached to one another by a string that was threaded around two spindles at the back end of the apparatus (Fig. 2a). This ensured that pulling one cane caused the other to retract simultaneously into the apparatus, thus imposing a mild form of forced choice. The entire apparatus was covered with a clear Perspex lid so that the food and canes were visible from above but the food could not be reached until it had been pulled out from underneath the lid. Six small tree finches, six nontool-using and six tool-using woodpecker finches participated in this experiment.

#### Transfer tasks

Those subjects that solved the initial version of the task were then tested in four further variations of the initial task (transfer tasks: Fig. 2b–e).

#### Erroneous configurations in first transfer task

During the first session of the first transfer task, four birds (two small tree finches, one nontool-using- and one tool-using woodpecker finch) erroneously received 4–10 trials that did not conform to the intended configuration of this task. In all cases, it was likely that the erroneous configurations should have made it easier for the birds to solve the task. However, none of them reached the success criterion or even approached it in their first session and the birds



**Figure 2.** (a) The cane task apparatus with canes and rewards arranged as in the initial test condition. (b)–(e) Transfer tasks 1–4 (first row: left-hand side is correct option; second row: right-hand side correct).

were subsequently given an extra block of the first transfer task to make up for the erroneous first session.

#### *Experiment 4: The Novel Box-opening Task*

This task was meant to test the ability to solve a novel operant task where success can be achieved through perseverance and application of a wide repertoire of motor behaviours to the problem but without necessarily paying attention to relational cues. In particular, the latter feature of the task distinguishes it from the physical tasks, which could be solved by forming rules based on the relationship between physical cues of the task. The apparatus was a box made of opaque, white Perspex with a transparent lid. The lid was hinged to the back edge of the box and overlapped the front edge of the box (Appendix Fig. A1). The box could be opened by pushing the protruding lip of the lid upwards. Before testing, birds were habituated to the box by feeding from it once while it was open. Subjects were given six sessions of 25 min, receiving up to three sessions per day. A bird was successful and testing was ended when it opened the box and ate the reward. If a bird did not make contact with the box during a session, it was rehabilitated to the box as described above and the session was repeated. A bird was given up to two extra sessions upon failing to make contact with the box in any one session. Eight small tree finches, eight nontool-using and 10 tool-using woodpecker finches participated in this experiment.

We scored the following variables from video footage for each individual: 'latency to make contact with the box in the first session' (s), 'total length of testing' (s), 'success' (opening box and gaining access to the food reward), 'number of pecks to the sides of the box' (with closed beak), 'number of pecks to the top of the box' (on the Perspex lid), and 'number of bites to any part of the box' (with open beak). A new variable called 'number of contacts' was created by summing all actions on the box (including pecks to all regions, bites, tool use) for each individual.

#### *Modified box-opening task*

Five small tree finches that failed in the original task were presented with a modified version of the task. This box was the same as the original except that the box sides were transparent and the lid was opaque. Here, the birds could see the food from the side, increasing the likelihood that they would accidentally nudge the lid upwards with their head, thereby solving the problem. We presented this modified box to control for the possibility that the small tree finches' inability to solve the original problem was due to lack of physical strength.

#### *General Experimental Procedure*

Experiments 1–3 were two-choice learning experiments involving the same basic procedure. These tasks were conducted in blocks of 10 trials. In each trial, the subject was given 5 min to make a choice between two options for which the correct side was randomized and counterbalanced right and left. Where there was more than one condition in the initial phase (experiments 1 and 2), starting conditions were balanced according to test groups.

In the initial task, subjects were given at least 140 trials to meet the success criterion. In the physical tasks, subjects that solved the initial task were subsequently tested in one or more transfer tasks. Since the emphasis here was on what the birds could immediately apply to the new version of the problem, birds were only given up to 30 trials to meet criterion in the transfer tasks. In the reversal task (experiment 1), there were two test phases and the birds were given 140 trials to meet the success criterion in each phase.

To meet the success criterion, a bird had to make 15 or more correct choices within two consecutive blocks of 10 trials. Specifically,

the number of correct responses in one of the two blocks had to be at least seven consecutively correct and in the other at least eight or in one block all 10 correct. This criterion was derived using a Monte Carlo simulation (details in Tebbich et al. 2007). Further details of the experimental procedure are given in Appendix 1.

Some subjects developed a positional bias, probably as a result of intermittent reinforcement. When a subject developed a positional bias, defined as six consecutive choices of one side, we employed a side bias correction procedure ('correction trials') until that subject chose the nonpreferred side once, whereupon we reverted to the normal pseudorandomized trial schedule.

#### *Statistical Analysis for Two-choice Experiments*

For all experiments, Fisher's exact test was used to test for group differences in the proportion of individuals to meet the success criterion (see above). If this test did not show a significant difference between tool-using and nontool-using woodpecker finches we pooled these groups and compared the pooled woodpecker finches with small tree finches.

We also compared learning speed and success probability between groups ('group' = small tree finches, tool-using woodpecker finches, nontool-using woodpecker finches) in the initial phases of the seesaw and cane tasks and for both phases of the reversal task using generalized linear mixed models (GLMM, Baayen 2008). To this end, a separate model was constructed for the initial phase of each experiment (and for each phase of the reversal task). Initial (full) models always included 'group' and 'trial number' as fixed effects and 'subject' as a random effect as well as the interactions between fixed effects up to the second order. The full model for the seesaw task also incorporated 'condition', referring to test condition in the initial task (Gap-central versus Food-central), as a fixed effect. Finally, to control for possible subject differences in learning speed, each model also included random slopes in addition to random intercepts (Schielzeth & Forstmeier 2009).

For all three GLMM model analyses, we started with a full model comprising all fixed effects and all possible interactions between them up to the highest order (second order). In the first step, we used likelihood ratio tests to compare the deviance of the full model with that of the null model which comprised only the random intercept and slope (Dobson 2002). This tests whether the full model explains significant variation in the data (results are given in Appendix Tables A3 and A4). This was the case in all three analyses. In the next step, we tested for the significance of the second-order interaction between 'group' and 'trial number'. In the case of the seesaw task model, we first tested the third-order interaction between 'trial number', 'group' and 'condition'. If the interaction was not significant, we removed it from the model and tested for significance of the main effects or significance of the second-order interactions in the seesaw model. In the seesaw model, we eliminated all nonsignificant second-order interactions and tested only the main effects.

We tested and report the significance of main effects only when we did not find a significant interaction between them and we indicate their results from models with the interaction removed. When testing for the significance of the main effect 'group' (categorical variable), we also used likelihood ratio tests, whereas when we tested for the main effect of 'trial number' (a continuous variable), we used the *z* test (provided by lmer). In cases where we found a significant main effect of group, post hoc testing comparing two groups at a time was conducted with the *z* test (Table A5).

GLMMs were fitted in R version 2.9.1 (R Development Core Team 2009) using the function lmer of the R package lme4 version 0.999375-31 (Bates & Maechler 2009). Because the response variable was binary (success/failure), we specified binomial errors and the 'logit' link function. Likelihood ratio tests were used to compare

models and were derived using the R function 'anova' with the argument 'test' set to 'chisq'. To increase the reliability of these tests, we set the argument 'REML' of the function 'lmer' to 'F'. Prior to fitting the model, we  $z$  transformed trial number to a mean of 0 and a standard deviation of 1.

We were primarily interested in the interaction between trial number and group since a significant interaction between these factors would have indicated that groups learned at different speeds. A significant main effect of group indicated that groups differed in their overall probability to choose successfully, a main effect of condition for the seesaw model indicated that one condition was easier to learn than the other, while a main effect of trial number simply indicated overall learning.

Prior to analysis, side bias correction trials were removed from all data sets and trials in which the seesaw apparatus malfunctioned were excluded from the seesaw data set. Some model results are reported in the main text (but see Appendix Tables A3–A5 for full results).

Recently it has been argued that the reliability of  $P$  values associated with fixed effects is questionable in the context of mixed models (Bolker et al. 2009). We therefore always backed up our conclusions with analysis of an alternative measure of learning speed ('point of steepest slope' calculated for each individual using the coefficients derived through logistic regression, Appendix 1) and an alternative measure of overall success probability ('percentage errors' for each individual) using one-way ANOVAs and the nonparametric Kruskal–Wallis test. The results of these analyses are reported in the main text only when they did not confirm the GLMM results (descriptive statistics and full results are contained in Appendix Tables A6–A9).

#### Analysis of Perseverance in Two-choice Experiments

In a recent paper (Tebbich et al. 2010), we proposed that perseverance in the natural feeding ecology of woodpecker finches, which often requires long bouts of pecking that are only rewarded after relatively long periods, might lead to a natural insensitivity to non-reward which could hinder their learning capacities and explain the fact that they were unable to outperform small tree finches in our two-choice learning experiments. Here we attempted to test this idea by looking at the perseverance with which woodpecker finches continued a positional bias when such a strategy yielded 0% success, namely while the side bias correction procedure of the two-choice learning tasks was being implemented. To this end, we examined sequences of choices to one side during application of the side bias correction procedure, asking (1) whether any of the groups persisted in choosing the nonrewarded side in a significantly higher percentage of trials than the others and (2) whether there were differences in the mean length of side bias correction sequences. A 'sequence' was defined as one or more consecutive choices of the biased side during the correction procedure, that is, following six consecutive choices to one side in which an individual was rewarded in approximately 50% of trials for this strategy, and breaks between sessions were disregarded.

For each individual, we calculated the percentage of total trials given during the side bias correction procedure and the mean correction sequence length by dividing the total number of side bias correction trials by the total number of side bias sequences. These two response variables were analysed using one-way ANOVAs with 'group' ('group' = small tree finches, tool-using woodpecker finches, nontool-using woodpecker finches) as the explanatory variable. The analyses were conducted separately for each task since not all birds were tested in all tasks.

Before applying the ANOVA test, we tested data for homogeneity of variance using the Bartlett's test (Snedecor & Cochran 1989) and

the Fligner–Killeen test (Conover et al. 1981). If either of these tests indicated a departure from homogeneity of variance ( $P < 0.05$ ), transformations were performed on the data in question, the best transformation was selected and the one-way ANOVA was then performed on the transformed data. Otherwise, testing was always performed with the original data and a Kruskal–Wallis test was always used to back up results. A summary of the statistical tests is given in Table 1 and descriptive statistics for the perseverance data are in Table 2.

## RESULTS

### Experiment 1: The Reversal Task

All individuals met criterion in the acquisition task within 80 trials or less and all except one met criterion during the reversal task in less than 140 trials. Thus, we did not test for group differences in the proportion of successful individuals.

Groups did not differ in their speed of learning either in the acquisition phase (acquisition phase model: group\*trial number:  $\chi^2_2 = 0.18$ ,  $P = 0.916$ ; Fig. 3a) or in the reversal phase (reversal phase model: group\*trial number:  $\chi^2_2 = 0.55$ ,  $P = 0.761$ ; Fig. 3b) although individuals clearly learned in both phases of the task (acquisition phase model: trial number:  $z = 8.315$ ,  $P < 0.0001$ ; reversal phase model: trial number:  $z = 8.597$ ,  $P < 0.0001$ ). Although the overall success probability did not differ between groups in the acquisition phase (group:  $\chi^2_2 = 0.18$ ,  $P = 0.915$ ; Fig. 3a), there was a significant difference between groups in the reversal phase (group:  $\chi^2_2 = 6.18$ ,  $P = 0.046$ ; Fig. 3b). Small tree finches had a significantly higher success probability than nontool-using woodpecker finches ( $z = 2.24$ ,  $P = 0.025$ ) and tool-using woodpecker finches ( $z = -2.18$ ,  $P = 0.029$ ) but there was no significant difference between tool-using and nontool-using woodpecker finches ( $z = 0.21$ ,  $P = 0.832$ ). The main effect of group was confirmed as a trend by the ANOVA and Kruskal–Wallis tests of percentage errors by group (one-way ANOVA:  $F_{2,21} = 3.11$ ,  $P = 0.065$ ; Kruskal–Wallis test:  $\chi^2_2 = 5.17$ ,  $P = 0.076$ ).

### Experiment 2: The Seesaw Task

#### Initial task

Of the six tool-using and six nontool-using woodpecker finches tested in the initial seesaw task, only two tool-using and no nontool-

**Table 1**

Test statistics comparing perseverance by group in terms of percentage correction trials and correction sequence length

Task	One-way ANOVA*	Kruskal–Wallis test
<b>Percentage correction trials</b>		
Reversal task		
Acquisition phase	$F_{2,21}=1.04$ , $P=0.371$	$\chi^2_2=3.02$ , $P=0.221$
Reversal phase	$F_{2,21}=0.88$ , $P=0.428$	$\chi^2_2=1.24$ , $P=0.537$
Seesaw task	$F_{2,15}=0.48$ , $P=0.630$ cosine	$\chi^2_2=0.80$ , $P=0.671$
Cane task	$F_{2,15}=2.31$ , $P=0.134$	$\chi^2_2=3.52$ , $P=0.172$
<b>Correction sequence length</b>		
Reversal task		
Acquisition phase	No appropriate transformation	$\chi^2_2=3.20$ , $P=0.202$
Reversal phase	$F_{2,21}=0.69$ , $P=0.514$	$\chi^2_2=1.06$ , $P=0.590$
Seesaw task	$F_{2,15}=0.58$ , $P=0.574$ $1/\log(x+1)$	$\chi^2_2=0.88$ , $P=0.643$
Cane task	$F_{2,15}=3.26$ , $P=0.067$ square root	$\chi^2_2=5.11$ , $P=0.078$

\* If the data did not meet the criterion for homogeneity of variance (criterion given in main text), the data were transformed until a suitable transformation was found. In this case, the ANOVA was conducted with the transformed data, and the transformation used is given with the results; otherwise, the test was performed on the original data. Transformations were always conducted on the response variable. If no suitable transformation could be found, only a Kruskal–Wallis test was performed.

**Table 2**  
Descriptive statistics quantifying perseverance by group in terms of percentage correction trials and correction sequence length (mean ± SE)

Task	Groups		
	TU	NTU	STF
<b>Percentage correction trials</b>			
Reversal task			
Acquisition phase	10.50±1.85	9.47±4.71	5.14±2.69
Reversal phase	12.29±3.18	13.83±5.76	7.17±2.15
Seesaw task	20.47±5.59	30.14±10.60	15.75±3.09
Cane task	5.49±1.87	15.44±4.90	10.86±2.18
<b>Correction sequence length</b>			
Reversal task			
Acquisition phase	2.60±0.37	2.92±1.72	1.47±0.75
Reversal phase	4.04±1.03	3.97±1.63	2.38±0.87
Seesaw task	5.20±0.95	9.86±5.07	3.97±0.64
Cane task	1.61±0.38	4.92±1.64	3.29±0.57

TU = tool-using woodpecker finch, NTU = nontool-using woodpecker finch, STF = small tree finch.

using woodpecker finches solved it (Fisher's exact test:  $P = 0.454$ ). A significantly higher proportion of small tree finches (five of the six) than pooled woodpecker finches reached the success criterion in the initial seesaw task (Fisher's exact test:  $P = 0.013$ ).

Groups did not learn at significantly different speeds (group\*trial number:  $\chi^2_2 = 4.29$ ,  $P = 0.117$ ; Fig. 3c) but there was a significant difference in the overall success probability (group:  $\chi^2_2 = 8.85$ ,  $P = 0.012$ ) with small tree finches having a significantly overall higher success probability than nontool-using woodpecker finches ( $z = 3.50$ ,  $P < 0.001$ ) but not than tool-using woodpecker finches ( $z = -1.34$ ,  $P = 0.181$ ). Furthermore, there was no significant difference between tool-using and nontool-using woodpecker finches ( $z = 1.43$ ,  $P = 0.153$ ). Overall, subjects tended to improve over the sequence of trials (trial number:  $z = 1.85$ ,  $P = 0.065$ ) with the suggested increase being seemingly similar in all groups. Subjects made

fewer errors in the Gap-central condition than in the Food-central condition ( $z = 4.125$ ,  $P < 0.001$ ).

*Transfer task performance*

Five of six small tree finches and two of 12 woodpecker finches (both tool-users) advanced to the transfer task. None of these birds passed the transfer task.

*Apparatus malfunctions*

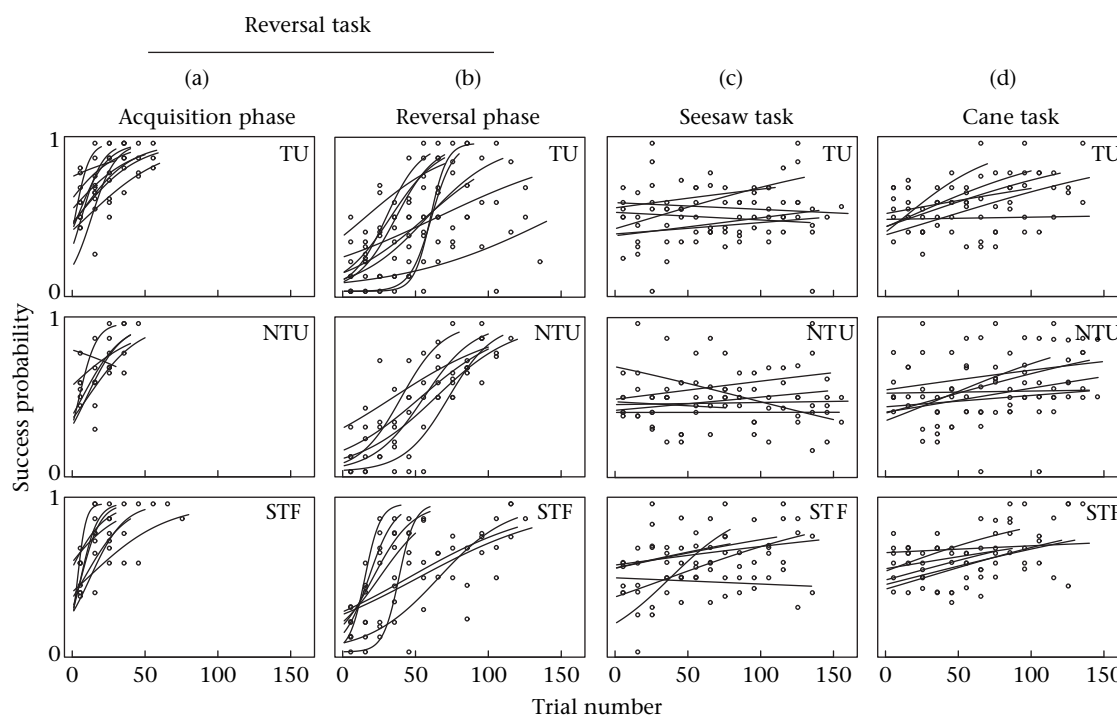
We found no significant group differences in the proportional occurrences of malfunction trials in the initial seesaw task (one-way ANOVA:  $F_{2,15} = 0.800$ ,  $P = 0.468$ ), and there was no significant correlation between proportional malfunction occurrences and proportional correct trials (Pearson product-moment correlation:  $r_{16} = -0.163$ ,  $P = 0.519$ ). This supports the interpretation that the group learning differences that we found reflect actual group differences in learning ability and are not simply an outcome of varying learning conditions.

*Experiment 3: The Cane Task*

*Initial task*

Eight of 12 woodpecker finches (three nontool-users and five tool-users) and all six small tree finches attained the success criterion in this task. Testing showed that there was no significant difference either in the proportion of tool-using and nontool-using woodpecker finches to reach criterion in the initial task (Fisher's exact test:  $P = 0.545$ ) or in the proportion of pooled woodpecker finches and small tree finches to solve the task (Fisher's exact test:  $P = 0.245$ ).

The speed of learning did not differ between groups (group\*trial number:  $\chi^2_2 = 0.76$ ,  $P = 0.683$ ; Fig. 3d) but overall, birds improved their performance over the sequence of trials (trial number:  $z = 5.31$ ,  $P < 0.0001$ ). The overall success probability differed between groups ( $\chi^2_2 = 6.91$ ,  $P = 0.032$ ), with small tree finches outperforming



**Figure 3.** Logistic regression models of success probability across trial number for two-choice learning tasks. (a) Acquisition phase and (b) reversal phase in the reversal task, (c) the seesaw task and (d) the cane task. Each row refers to one of the three test groups (TU = tool-using woodpecker finches; NTU = nontool-using woodpecker finches; STF = small tree finches). Points refer to the proportion of correct trials per block of 10 trials for each individual.



nontool-using woodpecker finches ( $z = 2.90$ ,  $P = 0.004$ ), but not tool-using woodpecker finches ( $z = -1.26$ ,  $P = 0.209$ ). There was no significant difference between the two woodpecker finch groups ( $z = 1.61$ ,  $P = 0.107$ ). This was confirmed as a trend by the ANOVA test of proportional errors by group ( $F_{2,15} = 3.54$ ,  $P = 0.055$ ) but not by the Kruskal–Wallis test ( $\chi^2_2 = 4.25$ ,  $P = 0.120$ ).

#### Transfer performance

The five tool-using and three nontool-using woodpecker finches and the six small tree finches that solved the initial task were subsequently tested in four transfer tasks. Only one subject, a tool-using woodpecker finch, met criterion in the first and second transfer tasks (Appendix Fig. A2). In the third transfer task, three small tree finches and six woodpecker finches (five tool-users and one nontool-user) met criterion, whereas no birds solved the fourth task.

Detailed analysis of the one tool-using woodpecker finch that passed three of the four transfer tasks showed that this bird might have reached success by employing a trial-and-error strategy: it switched between canes during a given trial in a higher percentage of all its transfer trials (58.9%) compared to the mean  $\pm$  SE percentage of transfer trials in which all other birds switched between canes in their transfer trials ( $7.4 \pm 1.4\%$ ), and 71.7% of the transfer trials in which this bird switched between canes in a trial were correct, although there was no significant relationship between switching and success for this bird ( $\chi^2_1 = 0.176$ ,  $P = 0.675$ ).

#### Experiment 4: The Novel Box-opening Task

Of the 18 woodpecker finches and eight small tree finches tested in this task, four nontool-using and four tool-using woodpecker finches successfully opened the box while none of the small tree finches were successful in doing so (comparison of the proportion of successful pooled woodpecker finches and small tree finches, Fisher's exact test:  $P = 0.031$ ).

Small tree finches were slower to make contact with the box in their first session (median = 133 s, range 31–1500) than woodpecker finches (median = 30 s, range 6–1500; Mann–Whitney  $U$  test:  $U = 33$ ,  $N_1 = 8$ ,  $N_2 = 18$ ,  $P = 0.030$ ) and they made contact with the box less frequently (median = 1.27 contacts/min, range 0.49–2.93) than woodpecker finches (median = 2.32 contacts/min, range 0.49–25.24; Mann–Whitney  $U$  test:  $U = 31$ ,  $N_1 = 8$ ,  $N_2 = 18$ ,  $P = 0.022$ ). Most of the contacts were pecks to the top and the sides of the box but three woodpecker finches also stabbed at the lid frequently with sticks or pieces of wire, even though this application of tools was never successful. Both species directed the majority of their pecks to the lid (small tree finches: median = 95.80 pecks, range 80.84–99.03; woodpecker finches: median = 91.87 pecks, range 71.67–100) and only a very small percentage to the sides of the box (small tree finches: median = 0.25 pecks, range 0–10.19; woodpecker finches: median = 1.02 pecks, range 0.00–8.05) and the percentage of side contacts did not differ between species (Mann–Whitney  $U$  test:  $U = 68$ ,  $N_1 = 8$ ,  $N_2 = 18$ ,  $P = 0.849$ ).

Taking the successful individuals into account, we found no correlation between contact frequency and success time (two-tailed Spearman rank correlation:  $r_s = -0.17$ ,  $N = 8$ ,  $P = 0.778$ ).

Four out of five small tree finches solved the modified version of the box-opening task within six sessions, with two of the birds even solving it in less than 40 s.

#### Perseverance in the Two-choice Experiments

We found no evidence for a difference between groups in perseverance as measured by the percentage of side bias trials in

either phase of the reversal task or in either of the physical tasks (Table 1). Furthermore, there was no significant difference in the mean length of side bias sequences in the reversal phase of the reversal task or in the initial phase of the seesaw task. A suitable transformation could not be found for the mean correction sequence length of the acquisition phase of the reversal task but the nonparametric Kruskal–Wallis test suggested that there was no significant difference between groups. There was a nonsignificant trend for a difference in groups in the initial phase of the cane task (one-way ANOVA with square root-transformed data:  $F_{2,15} = 3.26$ ,  $P = 0.067$ ). All summary and test statistics for this analysis can be found in Tables 1 and 2.

## DISCUSSION

Overall, our experiments provide no evidence that tool use in woodpecker finches evolved in conjunction with enhanced physical cognitive abilities, nor did we find any clear evidence indicating that tool-using experience in woodpecker finches hones physical cognitive abilities. Neither tool-using nor nontool-using woodpecker finches outperformed small tree finches in either of the physical tasks, nor did tool-using woodpecker finches outperform nontool-using conspecifics in any of the tasks. In fact, contrary to our expectations, small tree finches outperformed woodpecker finches in some aspects of both physical tasks: significantly more small tree finches than woodpecker finches solved the seesaw task, and in both of the physical tasks, small tree finches made the fewest errors overall, although this difference was only significant with respect to nontool-using woodpecker finches. With few exceptions most woodpecker finches had more experience with physical tasks prior to their exposure to the cane and seesaw tasks than small tree finches (Appendix Table A1). Thus it is all the more surprising that small tree finches performed as well as, or better than, woodpecker finches in the cane and seesaw tasks. The competence of small tree finches in the cane task was particularly surprising since the underlying physical problem presented in this task, namely sensitivity to the functional spatial relationship between tool and reward, is such an ecologically relevant one for tool-using woodpecker finches in the wild. Despite the adept performance of the nontool-using species, the fact that one condition of the seesaw task was easier to solve than the other illustrates the important role that attendance to simple cues (in this case, probably proximity to the food) played for both species in solving physical tasks.

The only task in which woodpecker finches outperformed small tree finches was the box-opening task: only woodpecker finches were able to solve this task because they made significantly more contact with the box. In a previous publication (Tebbich et al. 2010), we speculated that woodpecker finches' success in the box-opening task might be caused by their extractive foraging ecology, since their long bouts of energetic pecking at wood to gain access to a prey item should require them to be unusually persistent. Additionally, such perseverance might also be advantageous in the evolution and ontogenetic development of tool use in woodpecker finches because acquiring and practising tool use successfully probably requires a high level of perseverance. In particular, it is known that extracting a prey item with tools takes significantly longer than using a conventional foraging technique (Tebbich et al. 2002). On the flip side, we also proposed that such perseverance might impede learning flexibility in the two-choice learning tasks because high perseverance is likely to be associated with a low sensitivity to absence of reinforcement (Tebbich et al. 2010). To test this notion, we assessed whether woodpecker finches are less sensitive to nonreward by analysing the number and average length of side bias correction sequences by group. We

could not confirm this prediction: there was no significant difference between groups in the percentage of correction trials relative to the total number of trials or in the mean length of sequences in which individuals persisted in choosing one side.

The direct statistical comparisons between tool-using and nontool-using woodpecker finches did not yield evidence hinting at an improvement of physical cognition via experience. However, other aspects of our analysis suggest that such experience might still bring about subtle differences in physical task performance that we could not detect because of the low sample size. For one thing, we consistently found that small tree finches were significantly better in the physical tasks (make fewer errors overall) than nontool-using woodpecker finches but not tool-using ones. Furthermore, in the cane task, the only bird that solved all three transfer tasks was a tool-user, which suggests that this tool-using individual might have learned a more general rule. A close analysis of this bird's behaviour indicated that it applied a strategy involving observation of the moment-to-moment effect of its cane-pulling actions on the movement of the reward and thereby did not assess the solution in advance. It is possible that such a heuristic strategy is improved by tool-using experience, but on the basis of one bird, we cannot say for sure whether this is the case. To examine more fully the relationship between experience and cognition, a larger sample size of tool-using and nontool-using woodpecker finches is needed.

Although we did not find an enhancement of either sensitivity to the functional spatial relationship between tool and reward (cane task) or an appreciation of surface continuity (seesaw task) in tool-users, it is possible that future work will demonstrate a connection between tool use and other cognitive abilities not covered in the current study. One possibility is that woodpecker finches have enhanced cognitive abilities that are specific to active tool use, but for obvious reasons we were unable to assess this in a comparison with a nontool-using species.

The fact that small tree finches outperformed woodpecker finches in one of the general learning and even the physical tasks alters our concept of tool use in woodpecker finches and highlights the importance of investigating character divergence in closely related species. While a comparison between distantly related tool-using species indicated that the capacity to solve physical problems evolved in conjunction with tool use in woodpecker finches (Tebbich & Bshary 2004), it now seems much more plausible that capacities such as high flexibility shared by the clade were coopted for the specific woodpecker finch niche (Tebbich et al. 2010).

This conclusion is timely, coinciding with findings from a recent paper in which it was postulated that physical intelligence in corvids preceded the evolution of tool use in New Caledonian crows, *Corvus moneduloides*, and evolved in another context, possibly complex sociality (Bird & Emery 2009). In Darwin's finches, we consider it likely that certain cognitive characteristics such as high explorative tendency or flexibility in the stem species coupled with an untapped resource were the prerequisites for the development of tool use. In contrast to Bird & Emery (2009), we are not implying that anything more cognitively sophisticated than enhanced trial-and-error learning and/or exploration preceded tool use in Darwin's finches. Indeed, the failure of Darwin's finches to transfer rule learning to different versions of a physical task, even when very simple perceptual cues were available, indicates that the basic cognitive abilities of Darwin's finches are at a lower level compared to corvids (reviewed in Emery & Clayton 2009). This study has contributed further evidence showing that tool use, even that involving some degree of task-specific modification and selectivity (Tebbich & Bshary 2004), is not necessarily associated with sophisticated cognition; rather, there are simpler cognitive routes

that are capable of producing successful and seemingly complex tool use.

## Acknowledgments

S.T. was supported by the Austrian Science Fund (FWF, Project No. V95-B17) and I.T. by the German Research Foundation (DFG, Project No. TE628/1-1). We are grateful to Birgit Fessl for support in all facets of this study and to Roger Mundry for his detailed statistical advice. Caroline Raby, Viviana Morales, Mari Cruz Jaramillo, Tania Quisingo Chiza, Paola Buitron Lopez and Eduardo Sandoval provided valuable field assistance and help with experiments. Patrick Meidl provided vital support in organizing vast amounts of data. Thanks to Eric Koch and Andy Burnley for constructing the experimental apparatus, Theo Weber for graphics depicting the cane and seesaw apparatus and finally to Henrik Brumm for valuable comments on the manuscript.

## Supplementary Material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.anbehav.2011.07.032.

## References

- Auersperg, A. M. I., Gajdon, G. K. & Huber, L. 2009. Kea (*Nestor notabilis*) consider spatial relationships between objects in the support problem. *Biology Letters*, **5**, 455–458.
- Baayen, R. 2008. *Analyzing Linguistic Data*. Cambridge: Cambridge University Press.
- Bates, D. & Maechler, M. 2009. *lme4: Linear Mixed-effects Models Using Eigen and Splus*. Vienna: R Foundation for Statistical Computing.
- von Bayern, A. M. P., Heathcote, R. J. P., Rutz, C. & Kacelnik, A. 2009. The role of experience in problem solving and innovative tool use in crows. *Current Biology*, **19**, 1965–1968.
- Bird, C. D. & Emery, N. J. 2009. Insightful problem solving and creative tool modification by captive nontool-using rooks. *Proceedings of the National Academy of Sciences, U.S.A.*, **106**, 10370–10375.
- Bolker, B. M., Brooks, M. E., Clark, C. J., Geange, S. W., Poulsen, J. R., Stevens, M. H. H. & White, J.-S. S. 2009. Generalized linear mixed models: a practical guide for ecology and evolution. *Trends in Ecology & Evolution*, **24**, 127–135.
- Cochran, W. 1980. Wildlife telemetry. In: *Wildlife Management Techniques Manual* (Ed. by S. D. Schemnitz), pp. 507–520. Washington, DC: The Wildlife Society.
- Conover, W. J., Johnson, M. E. & Johnson, M. M. 1981. A comparative study of tests for homogeneity of variances, with applications to the outer continental shelf bidding data. *Technometrics*, **23**, 351–361.
- Dobson, A. J. 2002. *An Introduction to Generalized Linear Models*. Boca Raton: Chapman & Hall/CRC.
- Eibl-Eibesfeldt, I. 1961. Über den Werkzeuggebrauch des Spechtfinken *Camarhynchus pallidus* (Scalder und Salvini). *Zeitschrift für Tierpsychologie*, **18**, 343–346.
- Emery, N. J. & Clayton, N. S. 2009. Tool use and physical cognition in birds and mammals. *Current Opinion in Neurobiology*, **19**, 27–33.
- Hansell, M. & Ruxton, G. D. 2008. Setting tool use within the context of animal construction behaviour. *Trends in Ecology & Evolution*, **23**, 73–78.
- Harvey, P. H. & Pagel, M. D. 1991. *The Comparative Method in Evolutionary Biology*. Oxford: Oxford University Press.
- Hauser, M. D., Kralik, J. & Botto-Mahan, C. 1999. Problem solving and functional design features: experiments on cotton-top tamarins, *Saguinus oedipus oedipus*. *Animal Behaviour*, **57**, 565–582.
- Hauser, M. D., Santos, L. R., Spaepen, G. M. & Pearson, H. E. 2002. Problem solving, inhibition and domain-specific experience: experiments on cottontop tamarins, *Saguinus oedipus*. *Animal Behaviour*, **64**, 387–396.
- Iwaniuk, A. N., Wylie, D. R. & Lefebvre, L. 2009. The comparative approach and brain-behaviour relationships: a tool for understanding tool use. *Canadian Journal of Experimental Psychology*, **63**, 150–159.
- Kacelnik, A., Chappell, J., Weir, A. A. S. & Kenward, B. 2006. Cognitive adaptations for tool-related behaviour in New Caledonian crows. In: *Comparative Cognition: Experimental Explorations of Animal Intelligence* (Ed. by E. A. Wasserman & T. R. Zentall), pp. 515–528. Oxford: Oxford University Press.
- Lefebvre, L., Nicolakakis, N. & Boire, D. 2002. Tools and brains in birds. *Behaviour*, **139**, 939–973.
- Martin-Ordas, G., Call, J. & Colmenares, F. 2008. Tubes, tables and traps: great apes solve two functionally equivalent trap tasks but show no evidence of transfer across tasks. *Animal Cognition*, **11**, 423–430.
- Parker, S. T. & Gibson, K. R. 1977. Object manipulation, tool use and sensorimotor intelligence as feeding adaptations in *Cebus* monkeys and great apes. *Journal of Human Evolution*, **6**, 623–641.

- Povinelli, D. J. 2000. *Folk Physics for Apes*. Oxford: Oxford University Press.
- R Development Core Team 2009. *R: a Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. <http://www.R-project.org>.
- Reader, S. M. & Laland, K. N. 2002. Social intelligence, innovation, and enhanced brain size in primates. *Proceedings of the National Academy of Sciences, U.S.A.*, **99**, 4436–4441.
- Santos, L. R., Pearson, H. M., Spaepen, G. M., Tsao, F. & Hauser, M. D. 2006. Probing the limits of tool competence: experiments with two non-tool-using species (*Cercopithecus aethiops* and *Saguinus oedipus*). *Animal Cognition*, **9**, 94–109.
- Schielzeth, H. & Forstmeier, W. 2009. Conclusions beyond support: overconfident estimates in mixed models. *Behavioral Ecology*, **20**, 416–420.
- Seed, A. M. & Byrne, R. 2010. Animal tool-use. *Current Biology*, **20**, R1032–R1039.
- Seed, A. M., Tebbich, S., Emery, N. J. & Clayton, N. S. 2006. Investigating physical cognition in rooks, *Corvus frugilegus*. *Current Biology*, **16**, 697–701.
- Seed, A. M., Call, J., Emery, N. J. & Clayton, N. S. 2009. Chimpanzees solve the trap problem when the confound of tool-use is removed. *Journal of Experimental Psychology: Animal Behavior Processes*, **35**, 23–34.
- Shettleworth, S. J. 1998. *Cognition, Evolution, and Behavior*. London: Oxford University Press.
- Snedecor, G. W. & Cochran, W. G. 1989. *Statistical Methods*. 8th edn. Ames: Iowa State University Press.
- Spaulding, B. & Hauser, M. D. 2005. What experience is required for acquiring tool competence? Experiments with two callitrichids. *Animal Behaviour*, **70**, 517–526.
- Taylor, A. H., Hunt, G. R., Medina, F. S. & Gray, R. D. 2009. Do New Caledonian crows solve physical problems through causal reasoning? *Proceedings of the Royal Society B*, **276**, 247–254.
- Tebbich, S. & Bshary, R. 2004. Cognitive abilities related to tool use in the woodpecker finch, *Cactospiza pallida*. *Animal Behaviour*, **67**, 689–697.
- Tebbich, S., Taborsky, M., Fessl, B. & Blomqvist, D. 2001. Do woodpecker finches acquire tool-use by social learning? *Proceedings of the Royal Society B*, **268**, 2189–2193.
- Tebbich, S., Taborsky, M., Fessl, B. & Dvorak, M. 2002. The ecology of tool use in the woodpecker finch (*Cactospiza pallida*). *Ecology Letters*, **5**, 656–664.
- Tebbich, S., Seed, A. M., Emery, N. J. & Clayton, N. S. 2007. Non-tool-using rooks, *Corvus frugilegus*, solve the trap-tube problem. *Animal Cognition*, **10**, 225–231.
- Tebbich, S., Sterelny, K. & Teschke, I. 2010. The tale of the finch: adaptive radiation and behavioural flexibility. *Philosophical Transactions of the Royal Society B*, **365**, 1099–1109.
- Teschke, I. & Tebbich, S. 2011. Physical cognition and tool-use: performance of Darwin's finches in the two-trap tube task. *Animal Cognition*, **14**, 555–563.
- Visalberghi, E. & Tomasello, M. 1998. Primate causal understanding in the physical and psychological domains. *Behavioural Processes*, **42**, 189–203.
- Wimpenny, J. H., Weir, A. A. S., Clayton, L., Rutz, C. & Kacelnik, A. 2009. Cognitive processes associated with sequential tool use in New Caledonian crows. *PLoS ONE*, **4**, e6471.

## Appendix

### Sex of Experimental Subjects

Because male and female woodpecker finches are monomorphic it was not possible to determine the sex of all birds. Thus, we were not able to include 'sex' as a variable in our analysis.

### Determining Tool Use Abilities of Woodpecker Finches

We assessed the tool-using abilities of all woodpecker finches prior to participation in experiments as described in Tebbich et al. (2001). The procedure entailed placing food into drilled holes within a natural log and presenting the baited log to subjects over a series of sessions. Tools and tool material were abundant in each individual's aviary and tools were inserted into the holes at the beginning of each session. Observation sessions ranged from 20 to 30 min and the normal food was removed 2 h prior to the start of a session. An individual was categorized as a nontool-user if it did not show successful tool use within 530 min of observation. This time period seemed appropriate, since out of seven birds tested for tool use, all displayed tool use within 90 min of observation, and five displayed tool use within the first 30 min of observation.

## Two-choice Experiments

### General experimental procedure

The number of trials in which the reward was presented consecutively on one side never exceeded three, except in the case of a side bias correction procedure (see below). Choices were 'forced', meaning that choosing one option precluded subsequent choice of the other option, thus increasing the cost of making a wrong decision. In the cane task, however, only a mild form of forced choice was imposed since a bird could reverse its decision as long as one cane was not pulled all the way out of the apparatus. If a correct decision was made, the experimenter waited until the bird had finished eating the reward before entering the room and removing the apparatus. Rewards were either mealworm pieces, moth pieces, small pieces of boiled egg or bird food and varied from subject to subject. In the cane task, food rewards were placed in small white cups so that the reward was salient and to ensure that the food slid out of the apparatus easily. In the seesaw task, food rewards were inserted into clear plastic tubing so that they would easily roll along the surface of the seesaw platform. In both cases, the birds were given ample opportunity to learn that the reward was contained in these receptacles prior to the start of testing.

### Rehabilitation procedure

If no approach to the apparatus was made within 5 min during a given trial, the bird was given a rehabilitation trial in which a reward was placed somewhere on the apparatus where the bird could easily take it without operating the apparatus. If the bird took the reward within 5 min, the trial was repeated. If not, the block was ended since this indicated low motivation. For this same reason, a bird was also never given more than two rehabilitation trials per block: if a bird failed to approach the apparatus within 5 min for a third time within a block, that block was ended.

### Training procedures

**Reversal task training.** Prior to testing in the acquisition task, all subjects were habituated to the apparatus and trained by a shaping procedure to remove a white lid from a feeder with the same dimensions as feeders used in the reversal task. Once they learned this reliably (criterion was to retrieve the food reward from the box within 2 min in six consecutive trials), the bird advanced to the acquisition task phase.

**Seesaw task training.** Following a habituation phase, a seesaw platform without a gap was used. We shaped the birds to jump on the levers and thus obtain the reward. Using the platform without a gap meant that the birds were always rewarded regardless of which lever they jumped onto.

**Cane task training.** In the initial training phase, the birds learned to pull a stick that looked similar to the canes but was straight and had a white cup at the end resembling the food reward cup of later testing. In the next training phase, the subjects were familiarized with the choice between two options and the forced-choice nature of the task. The apparatus was the same as the test apparatus but instead of hooks, the same straight sticks as had been used in the previous training phase were used. The sticks were attached at their apices so that pulling one caused the other to retract into the apparatus. In contrast to the test phase, only one of the sticks was baited with a reward. Once a bird attained criterion (criterion was six consecutive successes in a session) in this task, it advanced to the test phase.

Location of maximum slope for each individual

The statistical model for logistic regression is:

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x$$

where  $p$  is a binomial proportion and  $x$  is the explanatory variable (trial number). We derived the coefficients of the respective logistic regression for each individual separately. With these derived coefficients, we then calculated the location of the maximum slope (i.e. a measure of how quickly they learn), for each individual to characterize the learning process as follows: if the linear predictor of the logistic regression is  $\beta_0 + \beta_1 x$ , then the location of the maximum slope is  $-\beta_0/\beta_1$ .



Figure A1. Woodpecker finch in box-opening task. From above, the bird can see the food reward inside.

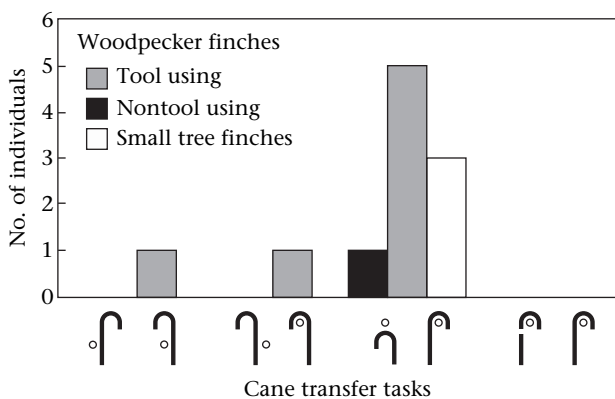


Figure A2. Cane task transfer performance. Performance is given as the number of successful subjects of each group in each transfer task. The cane and reward configurations for transfer tasks 1–4 are shown beneath the graph from left to right. Each of these diagrams shows the configuration for a trial in which pulling the right-hand cane would be correct.

Table A1 Subject's experiment participation and history

Subject	Species*	Tool-user†	Zone of origin	Experience with other experiments‡
blackL	STF	NA	Wet	1,2
blackred	STF	NA	Wet	3,4
blackwhite	STF	NA	Wet	4,3,6
blueL	STF	NA	Wet	1,2
greenL	STF	NA	Wet	1
lightblueL	STF	NA	Wet	1,2,5
orangeL	STF	NA	Wet	1,2,5
pinkL	STF	NA	Wet	1,2,5
purpleL	STF	NA	Wet	1,2,8,5,3,6
redorange	STF	NA	Wet	5,6
rosaL	STF	NA	Wet	1,2
white	STF	NA	Wet	4,3,2
yellow	STF	NA	Wet	4,3,6
yellowblue	STF	NA	Wet	4,3,6
yellowred	STF	NA	Wet	6
yellowwhite	STF	NA	Wet	4
blackpink	WPF	NTU	Wet	1,2,5,3,4,6
blueblue	WPF	NTU	Wet	1,5,6,3,4
lightgreen	WPF	NTU	Wet	1,2,8,5,6
orangegreen	WPF	NTU	Wet	1,2,5
purplegreen	WPF	NTU	Wet	1,5,6,3,4
redgreen	WPF	NTU	Wet	1,2,8,5,3,4
rosablue	WPF	NTU	Wet	1,2,8,5,3,4,6
rosapink	WPF	NTU	Wet	1,2,8,5,3,4,6
blackblue	WPF	TU	Dry	1,2,5,6
blackgreen	WPF	TU	Dry	1,2,3,4,7
bluered	WPF	TU	Dry	1,2,5,6
greengreen	WPF	TU	Dry	1,2,4,3,7
metal	WPF	TU	Wet	1,2,5,6,3,4
orangeblue	WPF	TU	Dry	1,2,5,6
purpleblack	WPF	TU	Dry	1,2,4,3,7
purplepink	WPF	TU	Dry	1,2,3,4,7
redblack	WPF	TU	Dry	1,2,5,6
redL	WPF	TU	Wet	1,2,8,5,6,3,4

\* STF = small tree finch, WPF = woodpecker finch.

† NTU = nontool-user, TU = tool-user, NA = not applicable (small tree finch).

‡ 1 = box-opening task; 2 = reversal task; 3 = seesaw task; 4 = cane task; 5 = two-trap tube (Teschke & Tebbich, 2011: physical cognitive task involving a choice of pulling one end of a stick which was inserted into a tube); 6 = modified two-trap tube (Teschke & Tebbich, 2011: same as 5, except that a smaller apparatus was used and the tube base was painted to make traps more salient); 7 = active mini trap tube (Teschke & Tebbich, 2011: same as 5 and 6 except instead of pulling the stick, birds (only tool-using woodpecker finches) had to use their own tool to move the reward); 8 = contact task (unpublished data: task that again involved pulling a stick, but required the birds to attend to the necessity of contact between the tool (stick) and the reward). The order of numbers representing experiment participation are in the order of actual presentation.

Table A2

Raw data used in the assessment of the effect of apparatus malfunction occurrences on success in the initial seesaw task for each bird

Subject	Group*	Malfunction trials	Correct trials	Total trials	Proportion of malfunctions	Proportion correct
yellow	STF	0	63	130	0.000	0.485
metal	TU	0	68	110	0.000	0.618
purpleblack	TU	2	79	130	0.015	0.608
greengreen	TU	3	60	140	0.021	0.429
purplegreen	NTU	4	75	150	0.027	0.500
redgreen	NTU	3	23	140	0.021	0.164
yellowblue	STF	1	44	80	0.013	0.550
purplepink	TU	2	59	140	0.014	0.421
white	STF	4	58	140	0.029	0.414
blackgreen	TU	4	52	140	0.029	0.371
blackred	STF	6	95	140	0.043	0.679
blackpink	NTU	6	69	150	0.040	0.460
purpleL	STF	1	46	80	0.013	0.575
rosablue	NTU	5	85	150	0.033	0.567
blueblue	NTU	5	69	160	0.031	0.431
blackwhite	STF	6	74	111	0.054	0.667
redL	TU	9	62	160	0.056	0.388

\* STF = small tree finches, TU = tool-using woodpecker finches, NTU = nontool-using woodpecker finches.

**Table A3**

Generalized linear mixed model (binomial error) showing the effect of group and trial number on success probability for the initial phases of the seesaw and cane tasks

Model terms	Cane task			Seesaw task		
	df	$\chi^2/z$	P	df	$\chi^2/z$	P
Full model	5	28.17	<b>&lt;0.0001</b>	11	27.86	<b>0.003</b>
Group*trial*condition	–	–	–	2	0.38	0.828
Group*trial	2	0.76	0.683	2	4.37	0.113
Group*condition	–	–	–	2	0.97	0.616
Trial*condition	–	–	–	1	0.05	0.824
Group	2	6.91	<b>0.032†</b>	2	8.85	<b>0.012†</b>
Trial‡	–	5.31	<b>&lt;0.0001</b>	–	1.85	<b>0.065</b>
Condition	–	–	–	1	13.77	<b>&lt;0.0003</b>

Model terms: groups = factor with three levels (small tree finch, tool-using woodpecker finch, nontool-using woodpecker finch), trial = trial number, condition = factor with two levels (Gap-central and Food-central, only relevant to seesaw task). Significant terms retained in the model are shown in bold (no significance test can be performed for factors involved in an interaction).

† For post hoc tests, see Table A5.

‡ The main effect of trial number was assessed with the z test; therefore the z statistic and no degrees of freedom are reported in this row instead of  $\chi^2$  and df.

**Table A4**

Generalized linear mixed model (binomial error) results showing the effect of group and trial number on success probability for both phases of the reversal task

Model terms	Acquisition phase model			Reversal phase model		
	df	$\chi^2/z$	P	df	$\chi^2/z$	P
Full model	5	37.43	<b>&lt;0.0001</b>	5	40.83	<b>&lt;0.0001</b>
Group*trial	2	0.18	0.916	2	0.55	0.761
Group	2	0.18	0.915	2	6.18	<b>0.046†</b>
Trial‡	–	8.32	<b>&lt;0.0001</b>	–	8.60	<b>&lt;0.0001</b>

Model terms: groups = factor with three levels (small tree finch, tool-using woodpecker finch, nontool-using woodpecker finch), trial = trial number, condition = factor with two levels (Gap-central and Food-central, only relevant to seesaw task). Significant terms are shown in bold. No tests of main effect were performed for factors involved in a significant interaction.

† For post hoc tests, see Table A5.

‡ The main effect of trial number was assessed with the z test; therefore the z statistic and no degrees of freedom are reported in this row instead of  $\chi^2$  and df.

**Table A5**

Post hoc tests for cane task model and model of reversal phase data of reversal task

Group comparisons†	Cane task	Seesaw task	Reversal task*
TU versus NTU	$z=1.61, P=0.107$	$z=1.43, P=0.153$	$z=0.21, P=0.832$
NTU versus STF	$z=2.90, P=0.004$	$z=3.50, P<0.001$	$z=2.24, P=0.025$
TU versus STF	$z=-1.26, P=0.209$	$z=-1.34, P=0.181$	$z=-2.18, P=0.029$

Significant results are shown in bold.

\* Data set consists only of data from the reversal phase.

† TU = tool-using woodpecker finch, NTU = nontool-using woodpecker finch and STF = small tree finch.

**Table A6**

Mean  $\pm$  SEM point of steepest slope according to group for the initial phases of the cane and seesaw tasks and both phases of the reversal task

Group	Cane task	Seesaw task	Reversal task	
			Acquisition phase	Reversal phase
STF	$-53.87 \pm 55.26$	$-6.70 \pm 16.02$	$4.45 \pm 2.46$	$35.17 \pm 6.46$
TU	$28.39 \pm 15.29$	$84.54 \pm 37.39$	$-1.92 \pm 5.41$	$56.23 \pm 11.26$
NTU	$3.33 \pm 38.63$	$2574.8 \pm 2481.2$	$15.92 \pm 11.54$	$53.45 \pm 5.84$

TU = tool-using woodpecker finch, NTU = nontool-using woodpecker finch and STF = small tree finch.

**Table A7**

Mean  $\pm$  SEM percentage errors according to group for the cane and seesaw tasks and both phases of the reversal task by group

Group	Cane task	Seesaw task	Reversal task	
			Acquisition phase	Reversal phase
STF	$39.30 \pm 2.24$	$43.85 \pm 4.19$	$27.14 \pm 2.22$	$47.88 \pm 3.17$
TU	$39.84 \pm 2.52$	$52.75 \pm 4.53$	$26.98 \pm 2.26$	$58.67 \pm 3.56$
NTU	$48.49 \pm 3.34$	$60.67 \pm 6.43$	$31.86 \pm 3.31$	$56.48 \pm 2.19$

TU = tool-using woodpecker finch, NTU = nontool-using woodpecker finch and STF = small tree finch.

**Table A8**

Test results of the comparison of point of steepest slope for the cane and seesaw task and both phases of the reversal task by group

	ANOVA*		Kruskal–Wallis		
	$F_{df}$	P	df	$\chi^2$	P
<b>Cane task</b>					
Group	1.39 <sub>2,15</sub>	0.386	2	3.84	0.280
<b>Seesaw task</b>					
Group	1.05 <sub>2,15</sub>	0.376	2	4.26	0.119
<b>Reversal task</b>					
Acquisition phase					
Group	1.90 <sub>2,21</sub>	0.175	2	3.14	0.209
Reversal phase					
Group	1.40 <sub>2,21</sub>	0.268	2	3.18	0.204

\* One-way ANOVA: explanatory variable = group, response variable = point of steepest slope.

**Table A9**

Test results of the comparison of percentage errors for the cane and seesaw task and both phases of the reversal task by group

	ANOVA*		Kruskal–Wallis		
	$F_{df}$	P	df	$\chi^2$	P
<b>Cane task</b>					
Group	3.54 <sub>2,15</sub>	0.055	2	4.25	0.120
<b>Seesaw task</b>					
Group	2.68 <sub>2,15</sub>	0.101	2	3.31	0.191
<b>Reversal task</b>					
Acquisition phase					
Group	1.03 <sub>2,21</sub>	0.375	2	1.93	0.381
Reversal phase					
Group	3.11 <sub>2,21</sub>	0.065	2	5.16	0.076

\* One-way ANOVA: explanatory variable = group, response variable = point of steepest slope.