Bending and Unbending of Tools by New Caledonian Crows (*Corvus Moneduloides*) in a Problem-Solving Context

Magdalena van Buuren[1,*], Jessica Alfredsson[2], Berenika Mioduszewska[3,4], Sabine Tebbich[1,¶], and Auguste M.P. von Bayern[3, 5,*,¶]

[1] Department of Behavioural Biology, University of Vienna
[2] Department of Psychology, Umeå University
[3] Max Planck Institute for Ornithology
[4] Messerli Research Institute, Medical University of Vienna, University of Vienna
[5] Department of Zoology, University of Oxford

* Corresponding authors (Email: magdalena.vanbuuren@gmail.com and avbayern@orn.mpg.de)
¶ Sabine Tebbich and Auguste M.P. von Bayern contributed comparably to this work and therefore share the last authorship.

Abstract – New Caledonian crows (NCCs) are proficient tool manufacturers and flexible problem solvers. A study reporting innovative hook bending by a captive individual in 2002 (Weir et al., 2002), has since been cited as a remarkable case of causal reasoning in animals. Recently, however, tool bending has been found to form part of the species' natural behavioral repertoire. Yet, further studies examining the flexibility of this behavior in controlled novel problem-solving contexts are lacking. We tested 17 captive NCCs, in one tool bending and three unbending problem-solving experiments. In the first study, two groups of five subjects were trained to select inflexible straight or hooked tools respectively for retrieving food from an apparatus. In the subsequent test, the crows were supplied with the wrong, but now pliable, type of tool. Three subjects made hooks while one subject straightened tools consistently. In the second study, 13 subjects, including five from Study 1, were tested. Ten subjects received an unbending task with a Z-shaped tool without prior training. Both angles needed to be straightened for the tool to become sufficiently long for reward retrieval. Four subjects solved the task at least once, and two did so consistently. Eight subjects, including the previous solvers, were challenged with yet another apparatus, requiring novel unbending behavior. One crow succeeded in both unbending tasks consistently. Our results show that NCCs can flexibly bend or unbend pliant material to create tools for solving novel tasks. The underlying cognitive mechanisms seemingly varied across individuals although trial-and-error learning prevailed.

Keywords – New Caledonian crows, Tool manufacture, Innovation, Novel problem solving, Bending, Unbending, Trial-and-error learning

New Caledonian crows (NCCs) are noted for their proficient and flexible tool use and tool manufacture (Chappell & Kacelnik, 2004; Hunt, 1996, 2014; Hunt & Gray, 2002; Knaebe et al., 2017; St. Clair & Rutz, 2013) but also for their flexible problem-solving skills and cognitive performance (Taylor et al., 2007, 2009, 2012; Teschke et al., 2013; von Bayern et al., 2009, von Bayern et al., 2018). The first report of innovative tool manufacture in birds, namely the spontaneous bending of a hook by the captive NCC called Betty in 2002, was a significant finding in the field of animal cognition (Weir et al., 2002). This first recorded tool bending occurred in an (unintended) novel problem-solving context, during a tool choice experiment, in which the hook tool required to lift a basket from a transparent, vertical tube had been

removed by the subject's interfering mate. Deprived of an appropriate tool, Betty spontaneously bent the available straight wire into a hook and then used it to retrieve the reward. Remarkably, after Betty initially used the straight wire unsuccessfully to probe the apparatus, she immediately proceeded to wedge it in the tape securing the base of the tube and to bend it without apparent haptic exploration.

To examine whether this was mere serendipity, the experimenters reproduced the circumstances with a single straight wire several times and, indeed, Betty continued to successfully modify and use the wire in the same manner, i.e., by wedging one end under some tape securing the tube or under her foot and then pulling the rest of the wire up to create a hook at the distal end. Throughout this series of trials, the specific steps of the behavioral sequence she used were not rewarded, except for the extraction of the rewarded bucket. Therefore, her performance could not be merely explained by a series of reinforcements (Weir & Kacelnik, 2006). However, it was possible that the first manipulation she attempted happened to be successful and that she immediately formed an association between that manipulation and extracting the reward.

In a follow-up study, Weir and Kacelnik (2006) tested the same subject in another bending task, but now with strips made from tougher aluminum that could not be modified with the previously used technique. With this new material, Betty first attempted to use the same technique, i.e., wedging one end of the tool under the base of the apparatus, then pulling it up to create a hook. After this approach failed, Betty adapted her technique to bend the proximal end with the beak, while securing the strip by wedging it in tape mounting the tube. If her behavior had been the result of simple associations from trial-and-error learning, one would expect her behavior to increase in variability, the "try-something-different" effect, during the extinction period for that association (Neuringer et al., 2001). Instead, the change in her behavior was centered on a single step of the entire sequence, but not the others. It therefore appeared that her behavior was goal-oriented, i.e., meant to produce a hook, instead of following a particular sequence associated with previous success (Bluff et al., 2007).

To further investigate the flexibility and functional understanding underpinning Betty's bending behavior and to establish that she would not apply bending when counterproductive, she was subsequently tested in two tasks where unbending the tool would be required to retrieve the food. The first task consisted of a horizontal transparent tube that contained food and a U-shaped aluminum strip that was too wide to fit into the mouth of the tube. In the second task, the bent tool was too short for reaching the food positioned inside a transparent tube. The results of the first unbending task were inconclusive, but in trials 3 and 4 of the second unbending task, Betty straightened the tool by pushing it against the lip of the tube and then extracted the reward. Sadly, Betty died after these first few trials of the second task, so no further insights could be gained. However, despite the hook bending behavior being reinforced prior to those tasks, Betty did not attempt to make a hooked tool in these tasks when this was counterproductive. It is therefore unlikely that she formed a simple association between pliant material and out-of-reach food that would trigger a learned specific behavioral response.

Betty's behavior was considered at the time an impressive example of "folk physics" (Emery & Clayton, 2005) and causal understanding in animals (Emery & Clayton, 2004; Weir & Kacelnik, 2006), which inspired a series of experiments with children (Beck et al., 2011, 2014; Cutting et al., 2011, 2014; Neldner et al., 2017; Whalley et al., 2017) and other large-brained birds (e.g., Bird & Emery, 2009; Laumer et al., 2017). However, the study was also subject to criticism suggesting that simpler mechanisms, such as shaping, could not be excluded (Shettleworth, 2009). Further, the conclusions were predicated on bending behavior not being part of the natural behavioral repertoire of NCCs, and thus a novel tool modification for solving a specific task.

Recent field research on wild NCCs has however revealed that they routinely employ bending techniques similar to the ones Betty had used (Klump et al., 2015; Rutz et al., 2016). The studies examined NCCs' remarkable hook crafting behavior, which involves creating hooks by trimming the base of detached side branches and found that 10 out of their 18 subjects (kept temporarily captive) bent the shaft of their newly crafted hook tools by applying force to the stem. Their findings suggest that modifying shaft curvature has a functional role and that wild NCCs adjust it deliberately, possibly to move the hooked working end in center of their vision. Most of the observed birds achieved this by securing the tool with

their feet and then bending the shaft with their beak. Yet, another technique consisted of wedging the tool upright and then pulling the shaft sideways with the beak, which Betty had done as well. Klump et al. (2015) also observed another crow flexing the shaft by pushing the tool end against a hard surface. These findings have significant implications for the conclusions previously drawn from Betty's hook bending. Her behavior in those experiments could have stemmed from some predisposition to bending flexible stick-like objects, and/or that, as a wild caught individual, she had learned those techniques before her capture.

Our aim was to assess whether captive NCCs, both wild-caught and reared in captivity, can solve a novel problem by creating a tool with the correct functional properties through modification of pliant materials by bending or unbending. We exposed our subjects to one bending and three unbending tasks in controlled experiments to assess whether both bending and unbending is present in more than a single NCC individual. Our tasks employed a similar paradigm as in Weir et al. (2002) and Weir and Kacelnik (2006); i.e., an extractive task, the solution to which required the instrumental modification of a pliant tool. However, it differed in several methodological aspects, which are discussed below, to get better insights into the underlying cognitive mechanisms. Although our data was collected before the discovery that bending forms part of the natural behavioral repertoire of NCCs, we believe it might add valuable information to the discussion.

Experiment 1 consisted of a bending and an unbending task. The subjects were provided with pre-experience in a more controlled fashion than Betty's experience before her first hook making. Further notable differences were the more naturalistic appearance of the apparatuses (tree trunks instead of transparent tubes), the removal of the basket from the bending task to decrease the cognitive distance between subject and reward, as well as use of different tool materials, which were either rigid or flexible just at both ends.

Experiment 2 followed on from Experiment 1's unbending task and aimed at improving it by making the apparatus simpler, but also making accidental success less likely by requiring two unbending modifications of the tool. In Experiment 3, we adjusted the apparatus design of Experiment 2 by making it considerably wider to prevent the unbending technique that had emerged from trial-and-error learning.

We closely monitored the subjects' path to first success and reliable replication, as well as their mistakes and cases of unnecessary modification, as they could also provide insights into the underlying cognitive mechanisms. We predicted that if the main mechanism was instrumental learning through trial-and-error, then the formation of the successful behavior would include mistakes (e.g., probing with non-functional shapes, failure to modify the tool) and a gradual establishment of reliable solving. If the solving was spontaneous, continuous and with few mistakes, as with Betty, we predicted that the main mechanisms would be similar to those described in the introduction, such as causal inference and/or pre-learned routines. While through instrumental learning, the final behavior might be goal-directed (de Wit & Dickinson, 2009) and compatible with causal understanding, here we specifically distinguish between causal understanding that precedes and underlies the final behavioral pattern and the causal understanding that might follow it.

Experiment 1: Bending and Unbending Tasks with Tree Trunk Apparatuses

Method

Subjects

The subjects were ten NCCs, five females and five males of various ages, which could not be determined precisely (see Table 1). They were wild-caught in family groups (both adults and juveniles) in 2003 and in the beginning of 2010, except for Uék, who hatched in captivity and was hand-raised in 2004. All subjects were tested in 2012. The subjects were housed in family units with up to two juveniles, or in pairs. Each aviary comprised an outdoor area between 15 and 32 m², and was connected to a lit and heated indoor chamber of about 7 m² in which the experiments were conducted. The aviary was enriched with gravel, woodchips, wooden toys, potted plants, a bird bath and branches. The birds had free access to both indoor and outdoor parts during the day, but at night were kept in the indoor chamber, which was on a LD

12:12h schedule with naturalistic sunlight lamps (Arcadia ®). They were fed every morning in the chamber with a mixture of minced beef, curd, rice, oatmeal, insect powder, egg powder, seed oil, vitamins and minerals (Korvimin®), as well as Nutribird® pearls or moistened cat kibble, scrambled egg and a piece of fruit (apple, orange or banana). Water was available ad libitum in the indoor compartment year-round, and additionally outdoors during summer.

Table 1

Sex, Ages and Repartition of the Subjects into the Bending-Trunk and Unbending-Trunk Groups in Experiment 1, of the Unbending-Tube in Experiment 2, and Unbending-Box Tasks in Experiment 3.

Age	Bending-trunk		Unbending-trunk		Unbending-tube		Unbending-box	
	Female	**Male**	**Female**	**Male**	**Female**	**Male**	**Female**	**Male**
Juvenile	Liane°	–	–	Jungle°	Calypso▲	Crusoe▲	*Calypso*▲	*Crusoe*▲
Adult	Annie-Claude*	Barney*	Uék▲	Papaye°	*Coralie°*	*Captain-Hook °*	Liane°	*Jungle°*
		Mango°	Tortue°		*Mermaid°*	*Papaye°*	Tabou°	*Papaye°*
		Aigaios°	Coquille°		*Tabou°*	*Mangroove°*	Tortue°	*Aigaios°*
					Tortue°			
					Tumulte°			

Note. Both Aigaios and Coquille started as naïve subjects in their respective groups. Birds marked with an asterisk were wild-caught in 2003, those wild-caught in 2010 are marked with a circle (°), and those born in captivity and hand-raised with a triangle (▲). In italics are subjects with experience with pliant tools prior to testing.

In accordance with the German Animal Welfare Act (Section V, Article 7; Borchert, 1998), no specific research permission was required for this non-invasive behavioral study.

We divided the subjects randomly into two groups, namely Bending and Unbending (hereafter Bending-trunk and Unbending-trunk) (see Table 1), which were tested with a different apparatus each (i.e., the Unbending-trunk or Bending-trunk apparatus; see Figure 1A, 1B)[1]. The subjects had no prior experience with hook tools or pliable materials since they had been captured.
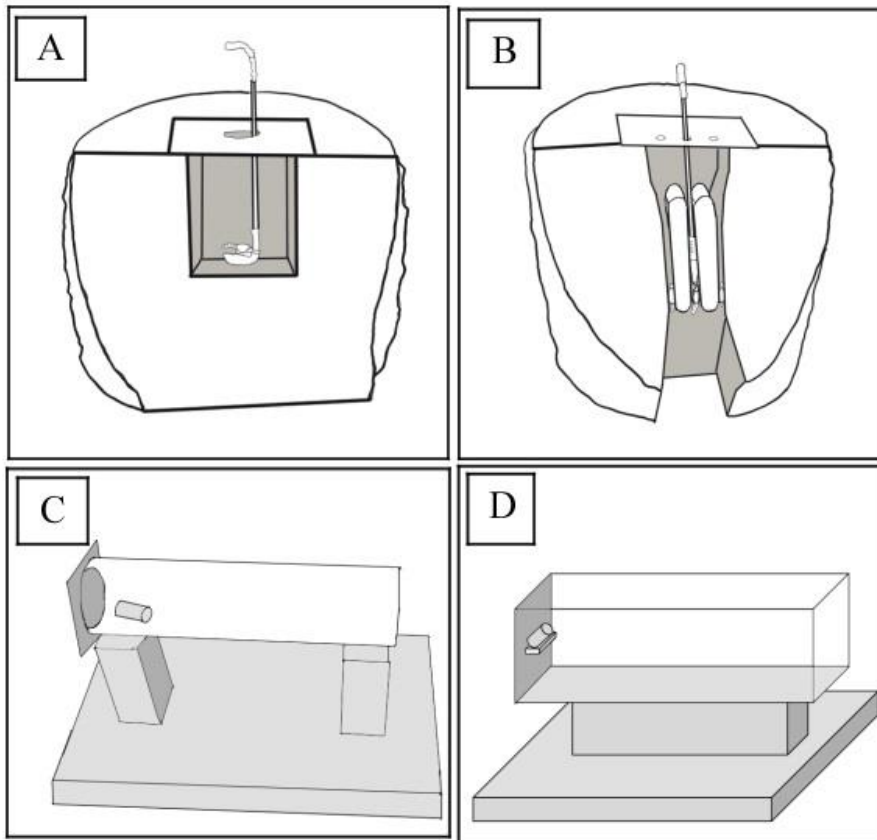
Materials

Both apparatuses were carved from wooden tree trunks to provide a naturalistic, possibly more intuitive setting for the subjects than transparent tubes. During training and testing, the apparatuses (35 cm wide x 25 cm high) were placed on the floor of the indoor chamber. The Bending-trunk apparatus (see Figure 1A) had a cavity (10 x 12 cm), covered by a transparent, but hatched acrylic plate so the three rewards were visible. The three rewards inside that cavity could be accessed through a hole on top of the apparatus when using an appropriate tool. The opening was designed to prevent retrieval of the reward without a hook as much as possible, although some subjects still managed to retrieve the food with a straight tool. When using a hook, the reward would be nestled in the hooked part and dragged up against the slanted back of the cavity until it reached the opening (see Video 1).

---

[1] Note that the study had been planned in a counterbalanced design testing both groups in both tasks with reversed order, but due to time constraints, the counterbalance was never conducted, and hence the study became a between-subjects design.

Figure 1

Experimental Apparatuses Used in the Three Experiments

Note. (A) Apparatus for the Bending-trunk task, represented with an inserted hooked tool and a reward to be pulled through the top. (B) Apparatus for the Unbending-trunk task, represented with a tool inserted in the middle groove pushing the reward down. (C) Apparatus for the Unbending-tube task, with the plastic capsule enclosing the reward at the end of the tube. (D) Apparatus for the Unbending-box task, with the capsule at the end of the box on a small platform.

The Unbending-trunk apparatus (see Figure 1B) had three openings at the top, leading each to a narrow tunnel (about 20 cm long) in which a reward was placed. Those tunnels were also covered by hatched acrylic plates to render the three rewards visible. The design would allow only straight tools to pass through the tunnels to dislodge the rewards and push them out of the tunnel ends, thus making them available at the bottom of the apparatus.

The tools for both apparatuses were made of thin bamboo (20 cm long; approx. 0.7 cm diameter) and multiple interwoven pipe cleaners attached to both ends with tape (see Figure 2A). These pipe cleaner ends were rigid (fortified by additional wire) during the initial tool selection training stage, but pliable during the test stage. The tools were placed inside a wooden frame (25 cm x 25 cm x 2 cm) onto the floor next to the apparatuses. The food rewards were dead beetle larvae (*Zophobas morio*) together with a piece of beef enclosed in a pistachio shell with transparent tape.
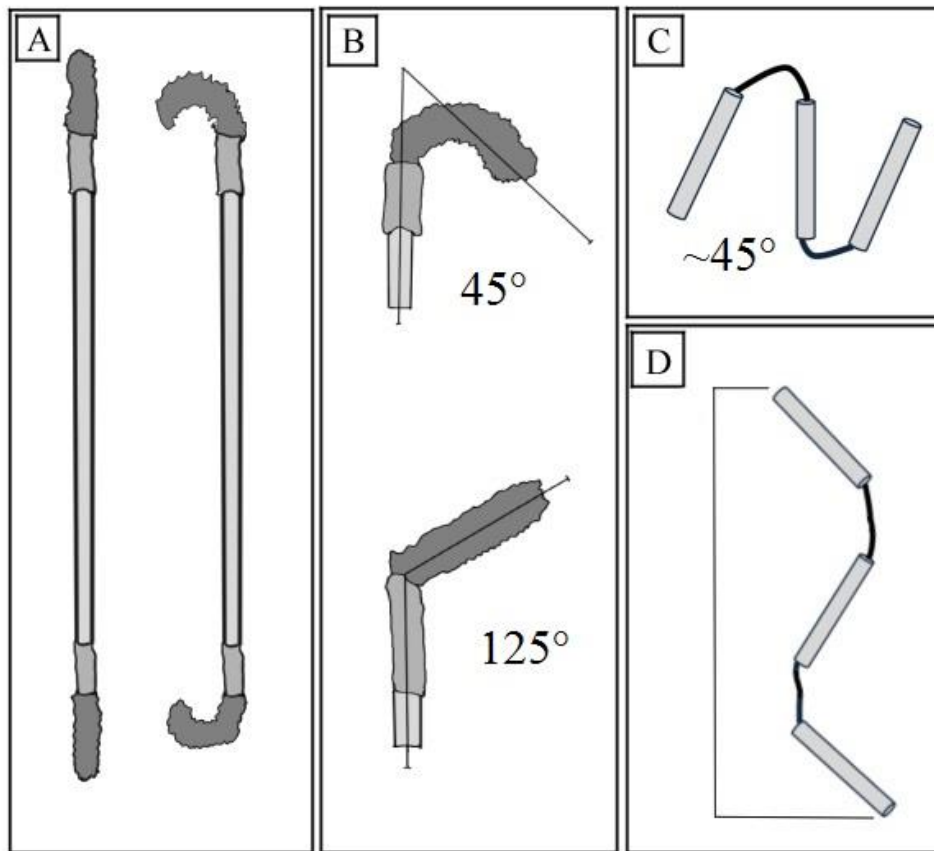
Experimental Phases and Procedures

Habituation. The subjects were habituated to the setup individually in six 10 min trials, a maximum of twice a day, in the morning and the afternoon. The trials were preceded by 45 min of food deprivation. During habituation, the apparatuses were not covered by the hatched acrylic plates, allowing the subjects access to the rewards inside. Additional rewards were placed on top of the Bending-trunk and the bottom

of the Unbending-trunk apparatus to encourage the subjects to approach. The subjects moved on to the next phase if they had taken all rewards throughout at least the last three trials.

Figure 2

Test Tools and Methods of Measurement

Note. (A) Tools, straight and hooked, used in Experiment 1. The bamboo was 20 cm in length, the pipe cleaner 3cm on each side and 0.8 cm in diameter. (B) Method to measure the bend of the tools in Experiment 1: the angle between a straight line going through the middle of the bamboo and a straight line going through the middle of the pipe-cleaner tip. (C) The Z-shaped tool as presented at the beginning of each test in Experiment 2 and 3. (D) The tool in the typical functional state produced by the subjects. The total length is measured from tip to tip.

Tool Selection Training. Prior to testing, all but one subject randomly selected in each group were trained to retrieve rewards from an apparatus using unmodifiable functional tools (straight for the Unbending-trunk group, and hooks in the Bending-trunk group). The two naïve pilot individuals were exposed to the apparatus and received food from it, but were not provided with tools in order to remain naïve to the shape of the functional tool. They were included to pilot whether NCC would be able to create a tool for solving a task without having experienced the required shape for solving that task before, with the objective to replicate a similar study that had been conducted on children (Beck et al., 2011). Following this training, all subjects were tested with their respective experimental setups (see Figure 1A, 1BB). After the testing phase, the naïve individuals (who both failed to solve) subsequently underwent the same tool selection training the other subjects had experienced before, and were then tested again.

During training trials, the subjects were presented with a choice of six straight and six hooked tools made from solid, unmodifiable material, which were placed in a frame on the ground. The functional tool was straight for the Unbending-trunk group and hooked for the Bending-trunk group. Subjects received six

tools of both types to ensure that they were available in sufficient numbers in case some were carried off and misplaced. During each trial, the subject could extract three rewards from the apparatus.

The subjects received one or two sessions per day, consisting of two 8 min trials, following 45 min of food deprivation. At the end of the first out of two trials of a session, the apparatuses were re-baited and all tools returned to the frame. To proceed to the test phase, all subjects had to participate in at least 15 training sessions and reach the following success criteria: they had to retrieve a minimum of 30 rewards (i.e., half the rewards available) within the last ten sessions (20 trials), and not have failed to retrieve any available reward within the last six sessions. Additionally, mistakes were taken into account: the subject must not have carried a non-functional tool outside the frame, or made insertion attempts with a wrong tool within the last six trials. Subjects that did not retrieve any food in 15 successive sessions were excluded.

Tests. During this phase, only the wrong, non-functional type of tool for the respective setup was provided in the wooden frame, but now the tool ends were pliable. In the Unbending-trunk group, all 12 tool ends were hooks and, in the Bending-trunk group, all 12 were straight. The procedure remained identical to the training phase, but the trials now lasted a minute longer (i.e., 9 min). Every three sessions (i.e., six trials), a training session was interspersed to keep the subjects motivated. In total, they received 12 test sessions (24 trials) and four interspersed training sessions including one after the last test trial. The naïve subjects received such training sessions as well, but without tools and with the rewards placed on top of the apparatus. To solve a trial, the subject had to modify the tool to make it functional and retrieve at least one reward (out of three) with it. A subject reached criterion for success by solving three trials in a row. The success criterion was used to determine whether a subject solved the task consistently.

Scoring and analysis

The experiments were video-recorded and all tools that were modified and/or used for probing were collected. A single coder (not blind) scored all the videos for the following information:
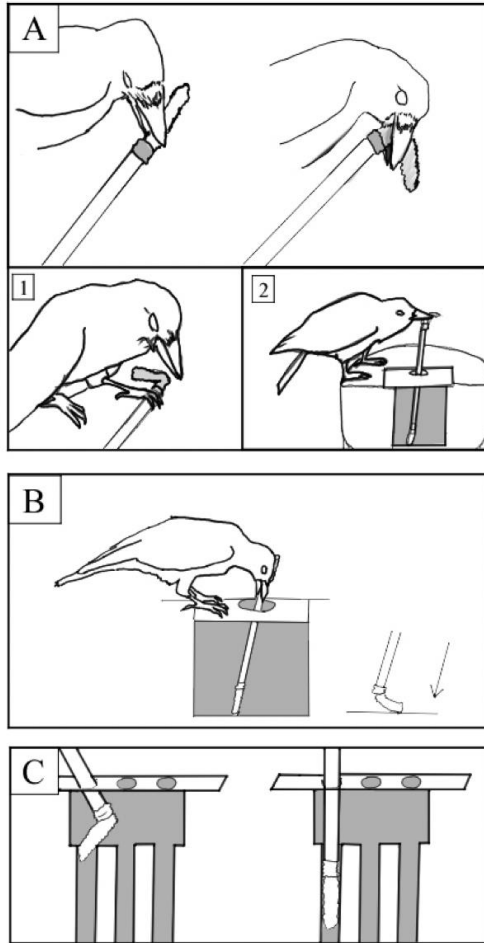(1) Every modification event with the duration and modification technique used. Three different techniques used were assessed and typified from the videos: a) the "beak" technique, where the subject used the beak to bend or straighten one flexible end, while holding the tool with a foot, or having it inserted in the apparatus (Video 1), b) the "pushing" technique, where the subject held the tool with the beak and pushed the distal and flexible end on a hard surface (Video 2) and c) the "insertion" technique, where the subject forced a bent tool into one of the grooves of the apparatus, thereby straightening an end (Video 3 and Figure 3).
(2) The degree of bend of each modified tool (see Figure 2B.).
(3) How often and how long each modified or unmodified tool was used to probe the apparatuses.
All tests were performed in R statistical software (version 3.5.1 R Core Team 2018).

To assess inter-observer reliability a second coder coded 6% of videos, which were randomly selected. The intra-class class correlations (ICC) were calculated using the "icc" command from the "irr" package. The ICCs were good to excellent: duration of unmodified tool use (ICC = .82, $p < .001$), modified tool use (ICC > .99, $p < .001$) and duration of tool modification (ICC = .74, $p < .001$).

To investigate whether successful subjects of the Bending-trunk group learned that unmodified tools were not as efficient (although there were instances of subjects succeeding to retrieve food with an unmodified tool) as modified tools to get the reward, we used a linear mixed model (LMM) with a normal distribution (identity link) fitted with the "glmmTMB" package. We used the duration of unmodified tool use per trial for each subject in seconds as response. As predictor variable, we used number of solved trials prior to the current trial. For example, if a subject solved trials 1, 2, but not 3, the assigned solved trials values are 0 for trial 1 (no preceding solved trial), 1 for trial 2, 2 for trial 3, and still 2 for trial 4. As control variables we used the trial number and the session number. Subject ID was added as random intercept, but the predictors were not included as random slope within subjects as the resulting model failed to converge. The assumptions of normality and homogeneity of the residuals were assessed by plotting the residuals. Because it appeared the assumptions were violated, the response was log-transformed, which solved the issue.

Figure 3
Modification Techniques used by Successful Subjects in Experiment 1

Note. (A) The grab (top left) and twist (top right) of the beak technique used both for bending and unbending: (1) holding the proximal end underfoot and (2) propping the distal end in some crevice or narrow hole. (B) The pushing technique in the Bending-trunk task. (C) The insertion technique in the Unbending-trunk task.

A second LMM was used to determine whether subjects in the Bending-trunk group modified their tools faster with increasing modification experience. The response was the average duration of tool modification per trial. The predictor variable was the number of completed tool modifications per trial prior to the current trial (on the same principle described above for the increasing number of solved trials), and both trial and session number were used as control predictors. Subject ID was added as random intercept, without random slopes within, as that model failed to converge. The visual examination of residuals showed departure from the assumptions of homogeneity and normality so the response was log-transformed, which again solved the issue.

As there was only one successful subject in the Unbending-trunk group, we tested the correlation between tool modification duration and the increasing number of successful tool modifications with a Spearman's rank test. For the Spearman's rank tests, the *p*-values are reported with a Bonferroni correction.

For the Bending-trunk group, we used a Generalized Additive Mixed Model (GAMM) ("mgcv" package, "gamm" function) to determine whether the bend of hooks had an impact on retrieval durations. This model was chosen because the relation between extraction duration and tool angle was likely not linear. As dependent variable, we included the probing duration per tool end divided by the number of rewards extracted with that tool. As a fixed effect, we included the angle of the tool. Subject ID was added

as a random intercept and tool angle as random slope within subject. The visual inspection of residuals showed departure from the assumptions of normality and homogeneity, so the dependent variable was log-transformed which resolved the issue.

Results

Bending-trunk Task

During the training phase, three out of four individuals reached criterion after 25, 31 and 37 sessions and were subsequently tested. The fourth individual, Liane, was excluded after 31 sessions due to health issues. The naïve individual (Aigaios) who initially did not go through tool selection training, failed the test and was therefore trained subsequently (25 sessions). Thus, a total of four subjects were tested in the Bending-trunk group. Aigaios and Mango solved the task for the first time on their first trial whereas Barney's first solved trial was his sixth trial (see Table 2). Tool use durations are summarized in Table 3. Three subjects reached the success criterion (solving three trials in a row; see Table 2). During the interspersed training sessions, when choosing a tool, all subjects except for Aigaios always selected a hook if they made a choice. In the first trials of Session 2 and 3, Aigaios first attempted to bend a straight tool (rigid) before selecting a hook and extracting the rewards. In the second trial of Session 3 as well as trials one and two of Session 4 (after test trial 24), he did not attempt to bend the rigid tools and selected the hooks directly.

In the test trials, instances of subtle modifications not followed by reward extraction were recorded. Because of the pliant nature of the tool ends, probing or pushing the pipe-cleaner against their cheeks, while holding the bamboo would sometimes slightly bend the tool (by less than 20°) seemingly accidentally (see Table 4; for further information, see supplementary material). For the successful modifications, the change of tool shape was more pronounced (see Table 4). After tool selection training, Aigaios used the beak technique for the first time from his first trial and in the following trials, exclusively. Both Barney and Mango used the pushing technique exclusively to make their tools (for details about the tools, see Table 4). For the three successful subjects, the first tool they made using their respective technique was also the first tool used to obtain a reward. Aigaios modified 12 tool ends, which he did not use (in his second test phase after tool selection training). This occurred from trial 4 onward when he probed for a reward with an already functional tool and then modified the proximal end (the handle) with his beak, seemingly in order to get the pipe cleaner out of the way and hold onto the bamboo better.

The duration of tool modification did not decrease significantly with the number of successful modifications for the successful subjects (LMM: E = -0.03 (95% CI [-0.06, 0.01]), p = .11; see Figure 4), indicating that the subjects did not modify their tools significantly faster with increased experience. For all successful subjects of the Bending-trunk group, the duration required to extract a reward with a functional tool varied from 1 s to 140 s. The angle of the tool did not have a significant effect on this duration (GAMM: R^2 = 0.02, p = .27), which indicates that the bend of the tool did not affect its usefulness significantly. All successful subjects used unmodified tools before they modified a tool for the first time, but also after (see Figure 5), and this unmodified tool use decreased significantly with the number of solved trials (LMM: E = - 0.20 (95% CI [-0.30, -0.10]), p < .001).

Table 2

Solved and Failed Trials in Subjects with at Least One Solved Trial

		Legend
🟩		Solved trials
⬜ (grey)		Failed trial
🟧		Failed but produced a functional tool
♦		Extracted reward with non-functional tool
<u>1</u>		Trials with >5s use of non-functional tools
*		Invalid due to experimenter error

Condition	Subject	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
Bending-trunk	nAi	<u>1</u>	2	<u>3</u>	4	5	6	7	8	9	10	11	12	13	14	15	16	<u>17</u>	18	<u>19</u>	20	<u>21</u>	22	23	24
	Ai	1	2♦	3	4	5	6	7	8	9	10♦	11	12*	13	14	15	16	<u>17</u>	18	19	20	21	<u>22</u>	23	24
	Ba	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	<u>16</u>	17	18	19	20	21	22	23	24
	Mg	<u>1</u>	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	<u>17</u>	18	<u>19</u>	20	21	22	23	24
Unbending-trunk	Uék	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
Unbending-tube	Cal	<u>1</u>	2	<u>3</u>	4	5	6	<u>7</u>	8	<u>9</u>	10	11	12	13	14	15									
	Cru	<u>1</u>	2	<u>3</u>	4	5	6	<u>7</u>	8	<u>9</u>	10														
	Pa	1	2	3	4	5	6	7	8	9	10														
	Tu	<u>1</u>	2	<u>3</u>	4	5	6	<u>7</u>	8♦	9	10	11	12	13	14	15									
Unbending-box	Cal	<u>1</u>	2	<u>3</u>	4	5	6	7	8	9	10	11													
	Cru	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19					
	Pa	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15									
	Ju	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20♦	21	22	23	24
		25	26	27	28	29	30																		

Note. nAi refers to the tests for naïve Aigaios.

Table 3

Summary of Tool Use of Both Functional, Non-Functional Modified and Unmodified Tools

Task	Subject	Solved trials	Unmodified tool use average (s)	NFM tool use average (s)	Functional tool use average (s)
Bending-trunk	Aigaios	21/23	22.8 ± 49.8	-	70.3 ± 38.9
	Barney	15/24	63.2 ± 126.8	-	101.3 ± 62.8
	Mango	8/24	101.5 ± 105.3	-	91.5 ± 43.8
Unbending-trunk	Uék	11/24	32.2 ± 46.5	-	51.6 ± 44.3
Unbending-tube	Calypso	3/15	13.1 ± 12.3	48.9 ± 73.4	75.5 ± 46.3
	Crusoe	8/10	16.0	221.8 ± 246.7	83.3 ± 72.0
	Papaye	6/10	109.7 ± 97.5	48.4 ± 39.5	8.3 ± 8.1
	Tumulte	1/10	17.6 ± 9.3	55.0 ± 61.0	51.5
Unbending-box	Crusoe	10/19	15.2 ± 21.1	39.4 ± 41.3	115.1 ± 127.0
	Jungle	1/30	31.2 ± 45.6	37.2 ± 25.9	12.0 ± 16.3

Note. NFM (Non-Functional Modification): Use of non-functional modified tools in Experiment 2 and 3 (Z-shaped tool with one straight angle).

Table 4

Number of Modified Tools Used Successfully, Used Unsuccessfully or Used Not at all in Experiment 1

Task	Subject	Modified used successfully	Modified used unsuccessfully	Modified not used	Modification duration average (s) (used tools)
Bending-trunk	Annie-Claude	0	**4 (167°)**	**4 (169°)**	**-**
	Naive Aigaios	0	**2 (170°)**	**3 (167°)**	**-**
	Aigaios	26 (118°)	9 (123°)	12 (128°)	3.0 ± 2.3
	Barney	16 (100°)	1 (100°)	**10 (161°)**	8.3 ± 5.6
	Mango	9 (121°)	1 (100°)	**10 (162°)**	41.4 ± 13.8
Unbending-trunk	Uék	11 (173°)	4 (100°, 145°, 2 folded)	2 (150°)	Beak: 10 ± 5.2 Insert: 6.1 ± 4.9

Note. The average bend is given in parentheses. Slight modifications (<20°) obtained while probing or pushing the pipe-cleaner against the cheek are indicated in bold font.

Figure 4

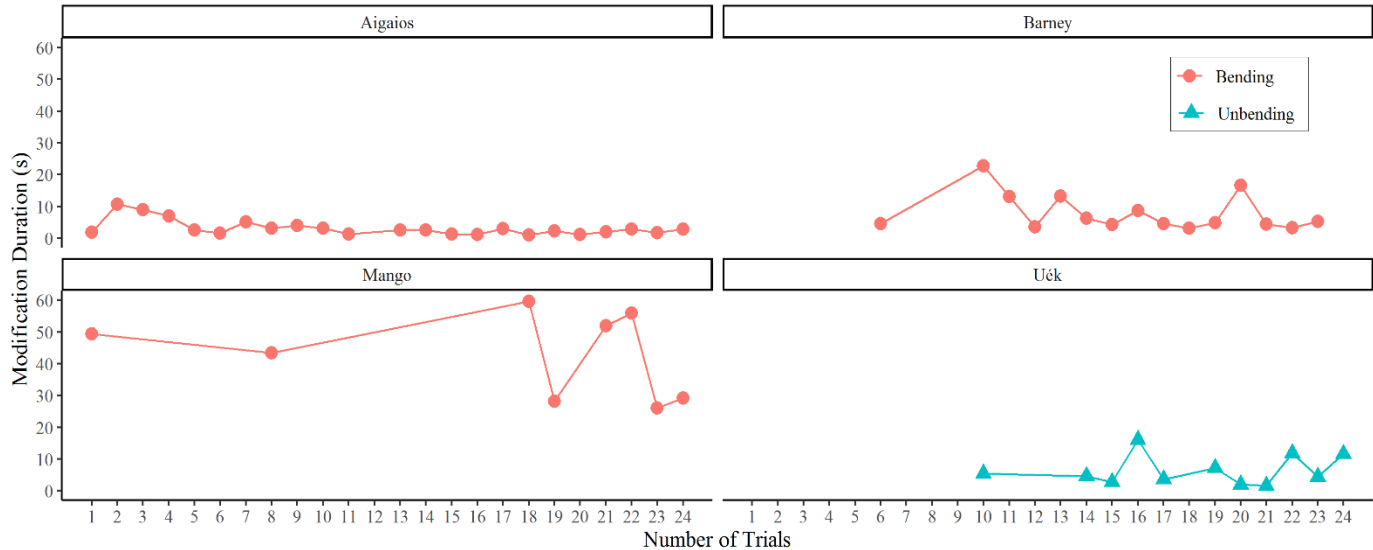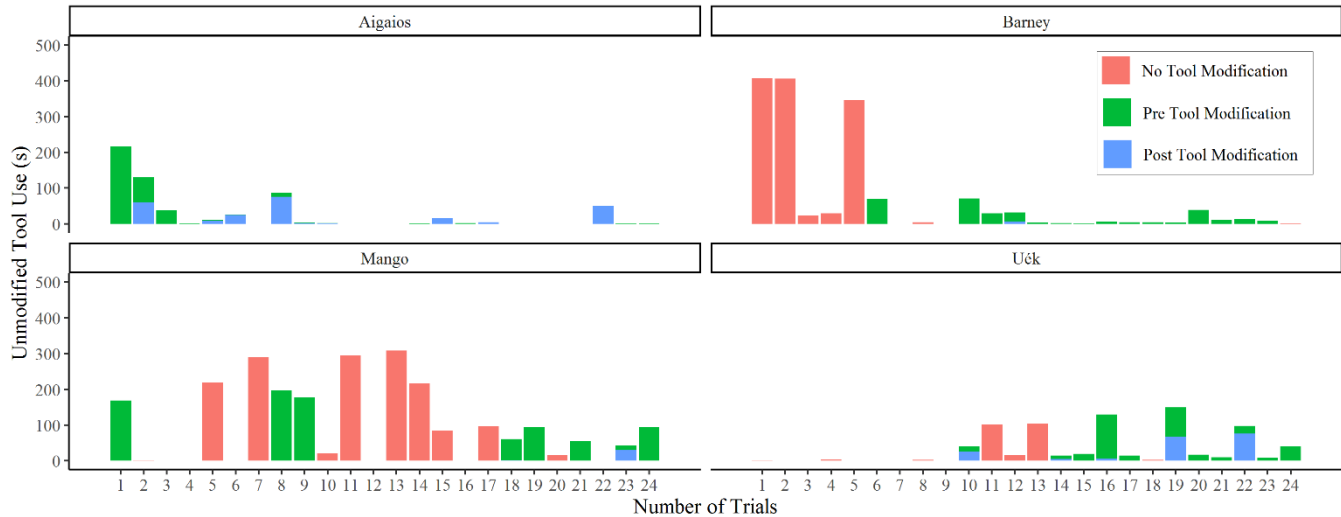*Tool Modification Duration Averages (s) Per Subject and Trial in Experiment 1 for Successful Subjects of Both Groups*

Figure 5

Unmodified Tool Use in Experiment 1 for Successful Subjects of Both Groups

Note. Aigaios, Barney, and Mango from the Bending-trunk group and Uék from the Unbending-trunk group. The unmodified tool use is represented in trials without tool modification, and with tool modification (before and after the tool is modified).

Unbending-trunk Task

During the training phase, three out of four individuals reached criterion after 30, 26, and 21 sessions and proceeded to testing. The naïve subject (Coquille) failed the test and was trained, then failed again and did not proceed with additional tests. Of these three subjects, only Uék solved the task. She started solving the task from trial 10 onwards and succeeded 11 times in total. She reached the success criterion in trial 16 (see Table 2).

Uék was the only subject in her group who modified tools. She used both the beak and the insertion techniques. Her insertion technique involved forcing her bent tool into one of the grooves of the apparatus, resulting in a straightened tool end (see Figure 4). Out of all 11 solved trials, 9 trials were solved using this insertion technique and two trials using the beak technique (including the first solved trial). All subjects selected the straight tools during the interspersed training sessions. In total, Uék modified and used 15 tools (see Table 4). The duration of tool modification was not significantly correlated to the number of successful modifications (Spearman: rho = 0.05, $p > .99$; see Figure 4).

Uék sometimes applied those modification techniques counterproductively (for details, see supplementary material). Uék attempted to insert unmodified tools in the apparatus (both at the top and from the bottom where the rewards fall) before her first successful modification. The time Uék spent trying to insert the tool in the apparatus increased significantly with the number of successful trials (Spearman: rho = 0.61, $p = .005$; see Figure 5).

Discussion

Three (out of four) successful subjects of the Bending-trunk group were capable of solving the task consistently, i.e., three times in a row. The way the first successful modifications were achieved varied across individuals and, in some cases, implied the involvement of chance. The gradual improvement of their performance, i.e., their decreasing use of unmodified tools, supports that individual learning formed part of the underlying mechanism. Other measures of learning not showing significant differences could be due to the low data point numbers. Our successful subjects' performance appeared to be underlined by different mechanisms which required individual analyses of their first modifications, their paths to proficiency, and their mistakes.

In the Bending-trunk group, the first modification of Barney and Mango, described as the "pushing" technique, could have emerged accidently, given that their behavioral patterns resembled normal probing behavior. Aigaios's first modification technique, described as the "beak" technique, seemed to be the least affected by chance, considering that it was applied directly to the pliant material and that it did not occur concurrently with other common behaviors such as probing.

Aigaios's first modification with the "beak" technique emerged after tool selection training, whereas he had never shown such behavior during the preceding test when he was naïve about the required tool shape. Still, a simple associative account could explain his first hook bending event: Aigaios could have pushed the pipe-cleaner away from his face in order to acquire a better grip on the tool or a better view of the tool tip on the other end. He then might have associated the resulting shape (hooked) with reward extraction from the previous training and therefore swiftly used it to extract the reward. However, this explanation is not totally satisfying as Aigaios never adjusted the grip of his tool during his naïve test. It is also unlikely that that he learned this type of hook making in the wild as suggested in Rutz et al. (2016), since he failed to do it before experiencing hooked tools.

The finding that Aigaios was successful only after he had gone through the tool selection training suggests that it provided an essential experience, such as the formation of a mental template of the required tool shape. A recent publication shows that NCCs are capable of mental template matching (Jelbert et al., 2018). NCCs were trained in dropping paper in a device that would then dispense a reward. Afterwards, some subjects learned that only large pieces of paper were rewarded while other subjects learned that small pieces were rewarded. NCCs were capable of consistently tearing off paper pieces of appropriate size. The authors argue that a drive for the transmission of tool design in wild NCCs could be such mental template

matching and that it might be the mechanism underlying Betty's first hook bending. It is worth noting that if a similar mechanism was involved here with Aigaios, it does not necessitate comprehension of the physical properties of hooks, only a mental template of the shape associated with the context of the task.

The path to consistent success varied in the three successful subjects of the Bending-trunk group. Mango's numerous failures and slow learning curve (i.e., the slow decrease in modification duration and in the use of unmodified tools across trials after first success) point to instrumental learning through trial-and-error. Barney and Aigaios, on the other hand, were both able to immediately re-apply their modification techniques following their first successful bending of a hook, when probing with straight tools in subsequent trials. In addition, they achieved maximum efficiency (i.e., modification speed) during their first five solved trials. Contrary to Aigaios, Barney had not been tested naïvely, so whether he had learned this technique in the wild cannot be ruled out. While Aigaios and Barney were consistent rather quickly (i.e., after one solved trial), both they and Mango continued to make apparent mistakes.

All successful subjects from the Bending-trunk group had used unmodified tools before solving the problem for the first time, which was to be expected given the complexity of the task. However, much like Betty, they also continued to use unmodified tools following their first successful modification. For Aigaios, who successfully used unmodified tools twice to extract a reward, it might have remained a viable method albeit less efficient. A possibility is that the successful subjects had not yet learned the final behavioral pattern necessary to solving the task (i.e., that probing with unmodified tools was not necessary), but another is that NCCs might employ a two-step heuristic strategy when selecting tools (Hunt et al., 2006). In the wild, when selecting a tool to probe a hole, they usually choose twigs or stems within close proximity of the target. This first tool's length is not related to the depth of the hole. The second tool, if the first is too short, is typically longer, and therefore more appropriate. The authors argue that this heuristic strategy might explain why Betty also probed the apparatuses with a straight tool before shaping a hook. It is possible, that her and our subjects tried the already available tool, tested it via probing and then proceeded with modification if necessary. This heuristic strategy might also explain why Aigaios attempted to modify a straight tool before selecting a hook in the interspersed training trials, after this modification technique was established. It is possible it had become his primary step in this problem-solving strategy.

One out of three subjects succeeded in the Unbending-trunk group. Uék first solved the task using the beak technique, which she then only used another time successfully in trial 22. During most of her solved trials, she used the insertion technique, which was similar to the pushing technique in that it was likely just the direct result of probing the apparatus. Uék's pattern of failures also showed the progressive establishment of the final successful behavioral pattern, which suggests instrumental learning via trial-and-error. Uék's modification technique was based on inserting the tool into the apparatus, which is likely the reason why the time she spent attempting to insert this tool increased significantly. However, the actual modification once the tool was inserted did not change, perhaps because she failed to identify which tactile feedback were useful as she could not see the tip of the tool. It is also possible the number of data points was too low to detect such correlation.

Uék also made mistakes, such as attempting reward extraction with unmodified tools while a modified tool was available, going as far as removing a modified tool from the apparatus to do so. In trial 22, she bent a straight tool end with the beak technique instead of using it to extract rewards. Her overall behavior showed that, contrary to Aigaios, she could not apply the beak technique consistently to solve the task, but that her main strategy was to simply force the tool inside the apparatus.

Two issues linked with the affordances of the experimental setup of both groups might have impacted the results. First, using identical but rigid tools in the training might have slowed the NCCs' discovery of the pliant properties because it involved a change in previously known properties, which had to be noticed through reassessment. An additional problem with the Unbending-trunk apparatus was that the inner grooves were obscured by the tool while probing. It therefore was impossible to see what the problem with the tool was (i.e., not sliding in the groove smoothly because of its shape). Both those issues were addressed in the design of Experiments 2 and 3, which were conducted without tool selection training, and in the structure of the apparatuses, which could be overviewed easily.

Experiment 2: Unbending Task with the Horizontal Tube Apparatus

Method

Subjects

We tested ten individuals with the tube setup in October 2014 (Table 1). The subjects were adults wild-caught in the beginning of 2010. The juvenile individuals (Crusoe and Calypso) were born in 2013, hand-raised and housed together. They were selected as they had been unsuccessful or not tested in Experiment 1. All subjects had experience with manipulating pliant materials, except for the hand-raised individuals.

The housing conditions were the same as in Experiment 1, except for the addition of enrichment devices with various hollow materials (wood, plastic, etc.) that the NCCs could probe. In accordance with the German Animal Welfare Act (Section V, Article 7; Borchert 1998), no specific research permission was required for this non-invasive behavioral study.

Materials

The tube apparatus was a horizontal acrylic tube (20 cm in length and 4 cm in diameter). Two wooden battens, screwed to a board (20 x 30 cm), propped up the tube to a height of 10 cm (see Figure 1C). The reward was placed 2 cm from the back of the tube. The reward was one dead beetle larva (*Zophobas morio*) encased in a plastic capsule for easier retrieval.

The tool used was constructed from three pieces of bamboo (5 cm long, 0.6 cm in diameter) connected by wire (1.5 cm long). At the beginning of each trial, the tool was provided in the Z-shaped form, folded at each junction at a 45–50° angle (see Figure 2C). The tool was long enough to retrieve the reward only if both angles were straightened (angles > 95–100°, depending on the size of the bird). The functional lengths measured were 13.5–18 cm. The rationale behind requiring two steps was to decrease the likelihood of accidental modifications leading to a functional tool.

Experimental Procedures

Habituation. During the habituation, the reward was placed near the opening of the respective apparatus, so the birds could easily retrieve it with the beak. Each trial lasted 10 min. To reach habituation criterion, the individuals had to retrieve the reward in three consecutive habituation trials.

Tests. The subjects were always tested individually in a lit indoor chamber they were well habituated to. The subjects received a minimum of 10 trials, and a maximum of 15 trials. The apparatus was placed on the floor or a table where available. The entire room was filmed by two cameras. In addition, the experimenter monitored the trial in real time using a baby phone camera. Each trial lasted a maximum of 12 min or stopped as soon as the subject retrieved the reward. If the reward or tool became inaccessible, or if the subject uncovered a cached stick (a behavior hard to prevent in NCCs who cache tools in captivity), the trial was paused for rectification. The subjects were tested once or twice a day (typically once in the morning and once in the afternoon). A trial was considered solved when the subject retrieved the reward with the straightened tool. A subject reached criterion by solving three successive trials, so that purely accidental solving would not be considered as success and the path to reaching this success criterion was examined.

When an individual solved a trial for the first time, they received another trial immediately, regardless of the number of trials that day. In addition, they would be tested at least five more times (during the subsequent three days) regardless of the number of previous trials (i.e., even exceeding maximum trial number). These measures aimed at determining whether the subject could replicate the behaviors leading to solving the trial instantly or whether their solving was more likely of accidental nature and thus could not be replicated straight away.

Three subjects (one had passed away) that solved at least one Unbending-tube trial were tested again a year after the initial tests to check their memory before Experiment 3 (Unbending-box). If during this second series of tests, they failed the tube task three times in a row, they were given a single easier trial, with a non-functional tool that required only one modification step (i.e., it had only a single bend that had to be unbent).

Training. From reviewing of the video-taped trials, it turned out that some subjects either did not manipulate the tool or never in a way resulting in modification and therefore likely failed to learn about its physical properties, i.e., its pliability. Thus, subjects that had failed to modify the tool throughout the first five trials (see Results section) were given experience with the flexible properties of the tool preceding each of the following five trials, to exclude failures simply due to the subjects' ignorance of the pliability of the material. The experience involved a four-part tool very similar in appearance to the tool provided during tests, but constructed with slightly more flexible wire in between the rigid parts, which was attached to the wall of their aviary in its folded form. A beetle larva hung from the tip of the tool to attract the subjects, slightly out of reach. Retrieving the worm unbent the tool. The aim was to familiarize the subjects with the pliability of this type of bamboo sticks combined with wires, but outside the experimental context and without it being used as a tool. The training trials lasted a maximum of 12 min, and were conducted on the same days as the five remaining trials (i.e., one training trial in the morning, and one test trial in the afternoon). To pass a training trial successfully, a subject had to retrieve the larva. If it failed, the training was repeated immediately until the subject was successful.

Scoring and Analysis

From the video recordings and the retrieved tools, one experimenter coded:
(1) The modification events (technique employed and duration of the modification, as well as time from first contact to the final, functional shape in case of interruptions).
(2) The degree of bend of each modified tool and its resulting length (as a straight line from one tip to the other; see Figure 2D).
(3) Frequency and duration of insertion in the apparatus of each functional tool, non-functional modified tool (only one angle straightened), and unmodified tool.

To assess inter-observer reliability, a second coder coded 10% of videos, which were randomly selected. The intra-class class correlations (ICC) were calculated using the "icc" command from the "irr" package. The ICCs were good to excellent: duration of unmodified tool use (ICC = .83, $p < .001$), non-functional modified tool use (ICC = .99, $p < .001$), modified tool use (ICC = .79, $p < .001$) and duration of tool modification (ICC = .97, $p < .001$).

We also determined if the subjects had learned that both angles needed to be straightened. To do so, we investigated whether their use of non-functional modified tools decreased with the number of successful modifications, i.e., modifications resulting in a functional tool. This number was for each trial, the number of successful modifications prior to the current trial. For example, if a subject made functional tools on trials 1, 2, but not 3, the assigned number of modification values are 0 for trial 1 (no preceding modifications), 1 for trial 2, 2 for trial 3, and still 2 for trial 4. We used an LMM with the "glmmTMB" package, with the duration of probing the apparatus with a non-functional modified tool as response. As a fixed effect, we added the number of successful modifications, and the trial number as a control variable. Subject ID was added as random intercept, and the number of tool modifications and trial numbers as random slope within "Subject." This full model did not converge, so the random slopes were removed. The residuals departed from the assumptions of normality and homogeneity so the response was log-transformed.

We also used a Spearman's rank test to determine the correlation between duration of tool modification and increasing number of successful modifications. It was applied to subjects with sufficient data points. For all Spearman's rank tests, the p-values are reported with a Bonferroni correction for repeated tests.
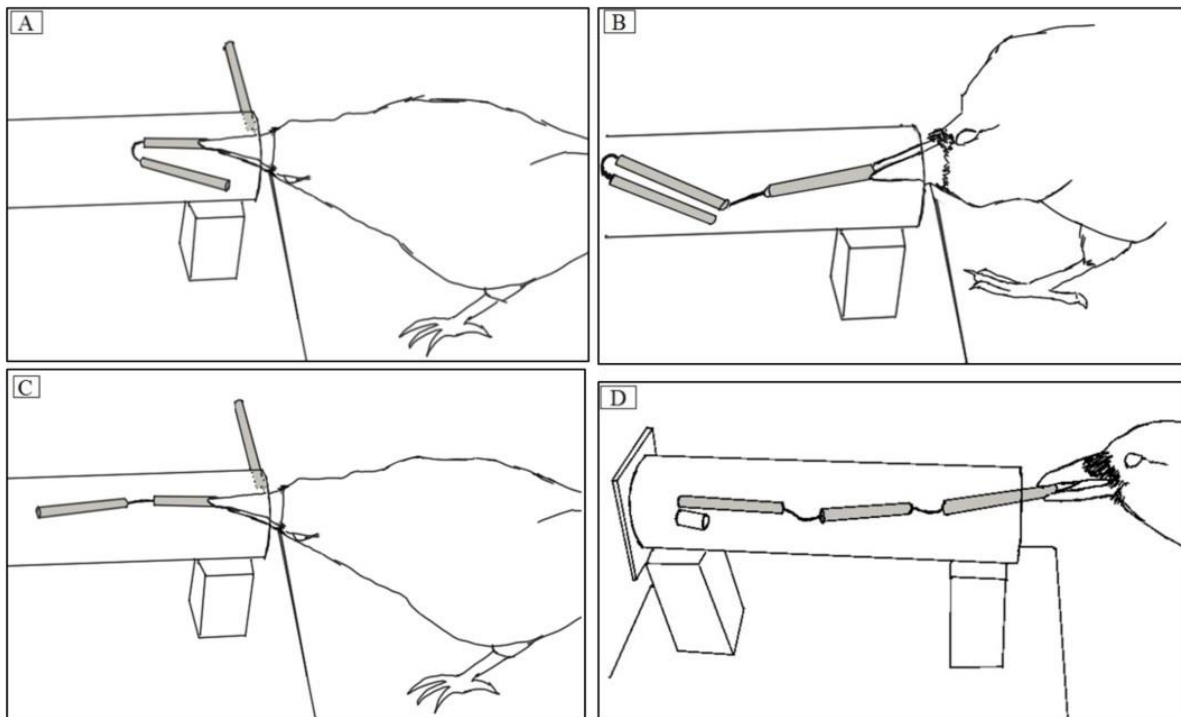
Results

All ten subjects reached criterion during habituation and were subsequently tested. During their first five trials, three subjects solved at least one trial (Crusoe, Papaye, Tumulte) (see Table 2) and were therefore never trained. The other seven did not modify any tools during their first five trials, and were therefore given experience with the tool material. Only one of them (Calypso) subsequently solved the task.

To modify the Z-shaped tool, the subjects forced one angle inside the tube while the second one was still outside. As they pushed the tool further inside, the wall of the tube levered the second angle straight. Usually after probing with the non-functional modified tool (only one angle straightened), they flipped the tool around and straightened the second angle by again using the wall of the tube as a lever ("shove-and-flip" technique; see Video 4 and Figure 6). An exception was Papaye's third trial and first complete modification, where he straightened one angle, then accidentally wedged the tool under the platform of the apparatus and then, while freeing it, straightened the second angle. Crusoe's modifications took on average 245.8 ± 248.9 s, Papaye took 170.5 ± 141.5 s, Calypso took 148.5 ± 243.4 s and Tumulte took 37 s. The modification duration was neither significantly correlated with the number of complete modifications for Crusoe (Spearman: rho = 0.52, $p = .88$) nor for Papaye (Spearman, rho = 0.37, $p > .99$).

Figure 6

Shove-And-Flip Technique Shown with the Tube Apparatus in Experiment 2

Note. (A) Subject pushes one angle inside the apparatus, using tube wall as a lever, which straightening the second angle. (B) Subject pushes the non-functional modified tool (one angle straight) toward the reward. (C) After flipping the tool around, the subject uses the tube to straighten the second angle. (D) The straightened tool is long enough to extract the reward.
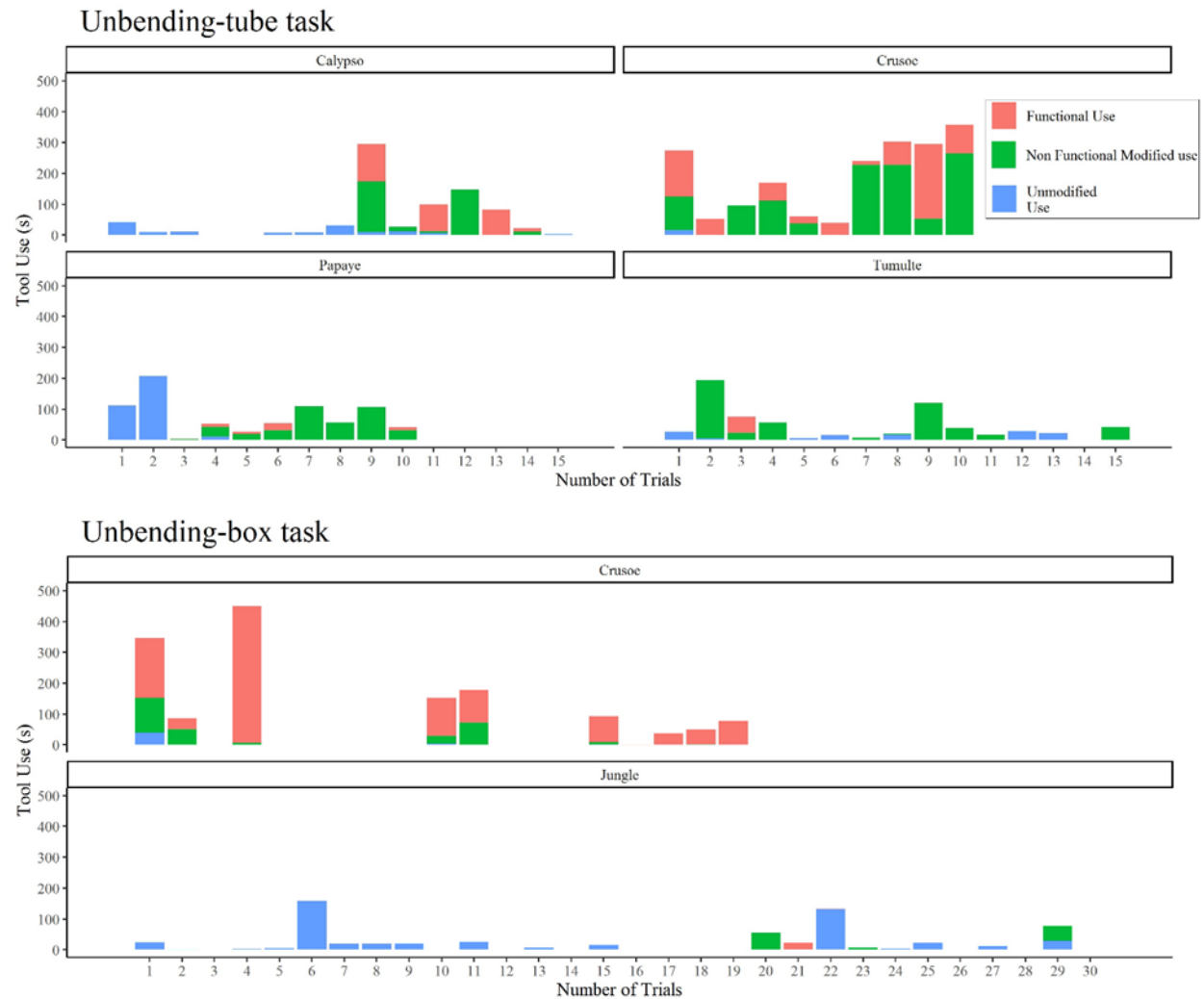
Both Crusoe and Papaye began the modification process immediately after picking up the tool, but Calypso probed before the modification (Figure 7). Subjects sometimes stopped at the first step of modification and used the non-functional modified tool (only one angle straightened) to probe the tube (see Figure 7). For subjects that solved more than one trial, the duration they spent probing with the non-

functional modified tool seemed to increase significantly with the number of complete modifications they had made (LMM, E = 0.61 (95% CI [-0.32, 0.90]), $p < .001$).

Three subjects (Crusoe, Calypso, and Papaye) were tested in the Unbending-tube task again a year later, but only Crusoe reproduced his performance. Both Papaye and Calypso failed four out of four trials whereas Crusoe solved two subsequent trials (out of two).

Figure 7

Tool Use of the Subjects with at Least One Solved Trial in Experiments 2 and 3

Discussion

In the Unbending-tube task, four out of ten subjects solved at least one trial, with two solving consistently from their first successful trial onward. Here we examine how their relatively complex modification process likely arose from simple mechanisms.

The successful crows in this task used a "shove-and-flip" technique for unbending, which resembled the one shown by Betty in her second unbending task. The most parsimonious explanation of its discovery is that it emerged accidentally while the crows attempted to probe with the folded tool. During such probing, one of the bamboo pieces would get caught on the rim of the apparatus, and while the crow kept forcing the tool inside the apparatus, that piece would get straightened. Next, after probing with such a half-modified tool, the subject would typically flip it around (flipping is a common behaviour, see Holzhaider et al., 2008), causing the remaining angle to get caught on the rim of the apparatus and becoming straightened, once again by forcing the tool inside the tube. The tool getting closer, and making more contact with the reward likely constituted a reinforcing signal for the instrumental learning of the behavioral sequence. Betty, in her second unbending task, similarly used the inside wall of the tube for straightening the tool and it remains unclear whether this occurred accidentally as a result of probing, as it might have been the case in our subject, though the authors posit that the behavior appeared "deliberate" (Weir & Kacelnik, 2006).

Two subjects, Papaye and Crusoe, quickly learned to apply this modification technique consistently, while Calypso and Tumulte appeared to have learned about it but only partially. Calypso could modify one angle, and only modified the other angle every other trial while Tumulte half modified her tool in 60% of her trials. In addition, they all interrupted the modification process to probe with the non-functional modified tool. This probing duration in the tube task unexpectedly appeared to increase rather than decrease as they gained experience with modifying the tool (though that appeared to apply mostly for Crusoe). It is possible that whatever cue they used to decide to flip their tool took longer to arise with the number of modifications, or it could be an effect of dealing with a small number of subjects with large variation between subjects.

It is likely that similar unintended cognitive hurdles as in Experiment 1 (i.e., the two-step heuristic approach) contributed to these errors. Whereas in the Bending-trunk task in Experiment 1, subjects could use unmodified tools to touch the reward, here the tool was only long enough to touch the reward after an initial modification. If the subjects indeed used a two-step heuristic approach as described by Hunt et al. (2006) and discussed earlier, our setup would be challenging as it would necessitate a third step, i.e., unbending the second angle that required suppressing their strong drive to probing (which by itself could be rewarding).

Considering the technique used by the successful subjects in our Unbending-tube task, unbending likely arose from trial-and-error learning in our subjects, despite the requirement of two unbending actions to render the tool functional. Because the subjects' behavior probably resulted from pushing the tool toward the reward, with the structure of the apparatus incidentally promoting the modifications, we planned our next experiment to prevent this "shove-and-flip" technique by adapting the width of the apparatus in order to offer less contact surface for levering the bents.

Experiment 3: Unbending Task with the Box Apparatus

Method

Subjects

We tested eight subjects with the box apparatus in October 2015 (Table 1). They were also wild-caught or hand-raised as well as housed as described in Experiment 2. Five out of those eight subjects had been previously tested with the Unbending-tube task as described in Experiment 2. Aigaios had been tested in the Bending-trunk task and Jungle in the Unbending-trunk task. One subject (Liane) had not been tested in any Unbending or Unbending study before. In accordance with the German Animal Welfare Act (Section V, Article 7; Borchert 1998), no specific research permission was required for this non-invasive behavioral study.

Materials

The box apparatus consisted of a rectangular box (29 cm deep) with a square opening (10 cm x 10 cm) at one side (Figure 1D). Two battens screwed to a board (26 x 30 cm) propped the box to a height of 8.5 cm. A small protrusion (1 x 2 cm) was attached to the center of the inner back side of the apparatus, onto which the reward was placed. The reward was one dead beetle larva (*Zophobas morio*) encased in a plastic capsule for easier retrieval.

This design aimed at preventing accidental modifications of the tool, which can easily occur when a bent tool is forced inside a narrow tube (see Figure 6). Importantly, the box apparatus was designed specifically to prevent the previously successful subjects in the Unbending-tube task from using the modification techniques they had developed. The tool used was the same as described in Experiment 2.

Experimental procedures

The subjects were habituated and tested using the same procedures described in Experiment 2. The main difference was that the subjects were never trained in any way, even if they failed to modify their tools within their first five trials. Instead, the maximum number trials increased from 15 to 30 to compensate for the increased difficulty.

Scoring and analysis

The scoring methods were identical to those described in Experiment 2. To assess inter-observer reliability, a second coder coded 10% of videos, which were randomly selected. The intra-class correlations (ICC) were calculated using the "icc" command from the "irr" package. The ICCs were good to excellent: duration of unmodified tool use (ICC = .74, $p < .001$), non-functional modified tool use (ICC = .78, $p < .001$), modified tool use ((ICC = .93, $p < .001$) and duration of tool modification (ICC > .99, $p < .001$).

We determined if the successful birds learned that both angles needed to be straightened. To do so, we investigated whether their use of non-functional modified tools decreased with the number of successful modifications, i.e., resulting in a functional tool. This number was for each trial, the number of successful modifications prior to the current trial. For example, if a subject modified tools in trials 1, 2, but not 3, the assigned modification numbers are 0 for trial 1 (no preceding tool modification), 1 for trial 2, 2 for trial 3, and still 2 for trial 4. We determined the correlation between non-functional modified tool use and the number of successful modifications with a Spearman's rank test.

We also used a Spearman's rank test to determine the correlation between the duration of tool modifications and increasing number of successful modifications for a particular successful subject. For all Spearman's rank tests, the *p*-values are reported with a Bonferroni correction for repeated tests.
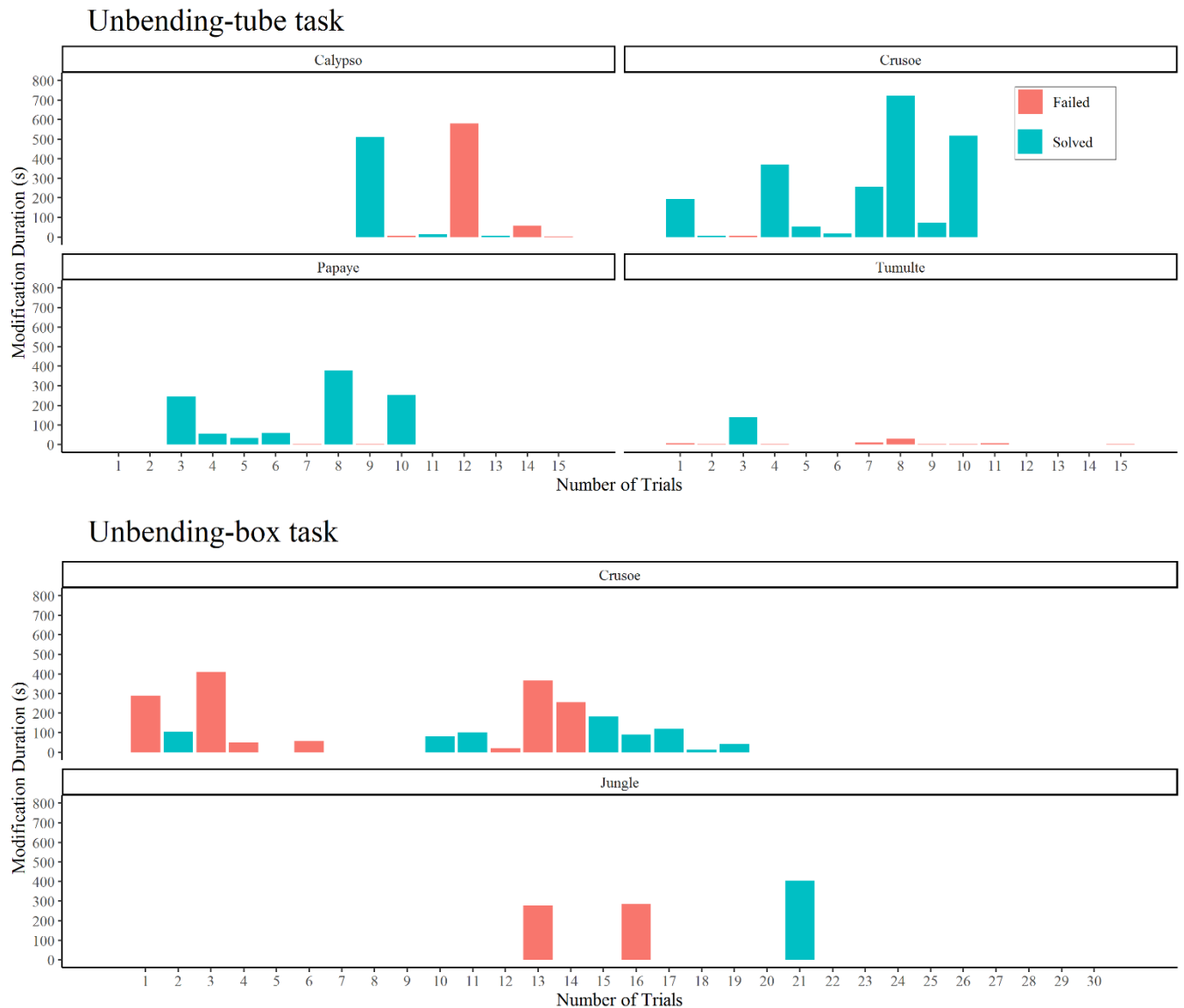
Results

All eight subjects reached habituation criterion and were tested. Two subjects, Crusoe and Jungle, solved the task (see Table 2); however, the latter solved it only a single time and once extracted a reward using a non-functional modified tool (he was a particularly large bird and had a further reach than anticipated). Neither Calypso nor Papaye, who had previously solved the Unbending-tube task, solved the Unbending-box task. Crusoe first solved the task on trial 2, then consistently throughout five trials from trial 15 onward. He produced his first functional tool in trial 1 (yet without reward extraction). In the following trials, he still probed with the unmodified tool before proceeding with modifying it two times (see Figure 7).

Whereas the design of the box setup aimed at preventing the "shove-and-flip" technique, Crusoe still used the wall of the box to unbend the tool in five trials (1, 2, 4, 10 and 11), but in later successful trials (15-19), he suddenly modified it by using a crevice in the aviary instead. The total duration of his modifications (μ = 108.1 ± 79.0 s) did not decrease significantly with the number of successful

modifications (Spearman: rho = -0.06, $p > .99$; see Figure 8). This crevice (in the rail of a sliding door) was situated 30 cm from the apparatus, and when modifying the tool, Crusoe turned his back on the apparatus and used it to completely straighten his tool. Calypso and Jungle also used narrow crevices in their respective testing chambers to modify the tool.

Figure 8

Total Modification Duration (Including Breaks) in Subjects with at Least One Solved in Experiments 2 and 3

While modifying the tool inside of the tube apparatus with the "shove-and-flip" technique (as described previously), subjects often probed the apparatus with the non-functional modified tool (only one angle straightened). This was the case for Crusoe, when he initially used the sides of the box as a lever to modify the tool, but his probing almost ceased when he later used the crevice for unbending the tool. However, he did not show a significant decline in probing with non-functional modified tools with the number of successful modifications (Spearman, rho = -0.37, $p = .80$; see Figure 7). In trials 15 and 18, he

still inserted a non-functional modified tool in the box but withdrew it swiftly as it was too short, and continued to modify it.

Given that the experiment was conducted in the same aviary, Calypso happened to use the same crevice as Crusoe to make functional tools in trials 10 and 11. However, while making a tool functional in trials 9–11, she only attempted to extract the reward with a functional modified tool in trial 10, but still failed to do so after probing intermittently for 53 s.

Jungle used an enrichment device consisting of a wooden box with various holes, and crevices attached to the wall of the experimental compartment to modify his tools, but often did so out-of-sight. He obtained a non-functional modified tool in nine trials (17–20, 22–24, 26, and 29), and functional tools in trials 13, 16, and 21, which he only used successfully in trial 21. He succeeded to extract a reward with a non-functional modified tool in trial 20 because he was a particularly large bird and managed to stretch far enough.

Discussion

Two out of eight subjects solved the Unbending-box task at least once, and one, Crusoe, was consistently successful.

Because the "shove-and-flip" technique appeared to be accidental as a result of forcing the tool toward the reward, it was meant to be prevented in the substantially widened box apparatus. Crusoe solved it and still managed to use this particular technique of shoving the bent part of the tool against the lip of the box. Importantly, however, this modification could not have resulted from forcing the tool toward the reward as it was never directed at the food in the inner box. Rather, Crusoe specifically pushed the tool against the wall of the box –not directed at the food- until both angles were straightened.

Three trials later, he used a crevice not forming part of the apparatus to achieve the same result and even moved and turned away from the box for that purpose. Given that Crusoe had participated in, immediately succeeded and consistently solved the previous tube unbending task, it seems likely that he transferred his prior experience of using the wall of the tube for unbending the tool to the new box task, even though the features of the new problem differed from the experienced one in visual appearance. Yet, undoubtedly, he had been previously reinforced for unbending by means of using a transparent rim as lever. To what extent he had, through that experience, extracted a functional understanding of the task from the previous one remains unclear.

General Discussion

We aimed at determining to what extent captive NCCs flexibly modify the shape of pliant materials through bending or unbending in order to solve novel problems in a controlled setting. We tested NCCs in one bending task and three unbending tasks and investigated possible underlying mechanisms by examining a) the path leading to their first modification, b) whether they replicated a successful modification as well as c) mistakes that would be indicative of the underlying mechanisms involved.

Three out of four individuals reached success criterion in the Bending-trunk task, one out of three in the Unbending-trunk task, two out of ten individuals in the Unbending-tube task and one out of eight in the Unbending-box task. This finding shows that the ability to bend but also the ability to unbend flexible material to form functional tools is not restricted to a single individual, Betty (Weir & Kacelnik, 2006).

The crows' predisposition and motivation to probe with tools into crevices (Kenward et al., 2005, 2006) seemingly played an important role in the mechanisms that underpinned the subjects' path to first success. It likely caused several subjects to stumble on the solution accidentally and likely also facilitated trial-and-error learning in subjects that acquired consistent solving only gradually (e.g., Mango in the Bending-trunk task). This predisposition coupled with such associative learning was probably also the reason several subjects developed a very similar two-step modification technique in the Unbending-tube task.

The types of techniques used by our subjects were largely consistent within each task and similar to the ones Betty had used in comparable tasks. This was probably the result of the similar task affordances and behavioral predispositions (Kenward et al., 2005, 2006) as well as motor constraints of NCCs for certain movements when handling flexible stick-like tools. In line with this, some of the bending techniques observed in the lab also paralleled those reported from wild NCCs, namely the beak technique while holding the tool underfoot, the beak technique while propping the tool in a hole, and the pushing technique where the distal end of the tool is pushed against a hard surface (Klump et al., 2015; Rutz et al., 2016).

In contrast to the bending behavior in the wild, which was applied on the stem of already hooked tools in order for its functional end to appear in the NCC's field of vision, the bending shown by Betty and our subjects was used to produce the functional hooked end itself. This shows that while the NCCs in the lab might have used the same techniques as their wild counterparts, they applied them flexibly in a different and most likely novel problem-solving situation that exceeded the natural context in complexity (i.e., creating a functional tool is arguably more complex than adjusting an already existing one).

Moreover, the NCCs kept in the lab also showed unbending, thus *diminishing* the curvature of hooked tools, in problem-solving tasks where a straight tool was required. Here, in the Unbending-tube task, one subject, Uék, employed the "beak" technique to unbend a hooked tool end, as well as another new technique, i.e., the "insertion" technique, whereas further subjects showed an unbending technique similar to the one shown by Betty when unbending ("shove-and-flip") in the two further unbending tasks (tube and box tasks). This indicated that if NCCs did possess predispositions to bending, most subjects did not apply them when counterproductive (except for Uék on one occasion for ergonomic adjustment), and in the case of Betty, even with a history of bending being rewarded for aluminum strips, she still did not bend them counterproductively in her unbending tasks (Weir & Kacelnik, 2006).

A combination of behavioral predispositions and learning has already been demonstrated in NCCs for both stick-tool use and Pandanus tool manufacture (Hunt et al., 2012; Kenward et al., 2005, 2006) and has also been shown in woodpecker finches (Tebbich et al., 2001), another habitual avian stick-tool user. The fact that our subjects could develop both unbending and bending techniques may point toward a predisposition of NCCs to modify the shape of flexible materials generally rather than bending specifically. Alternatively, it might indicate that unbending forms part of NCC's genetically predisposed behavioral repertoire as well, or that they display a behavioral flexibility that allows them to develop novel modification techniques, especially when challenged in experimental settings.

Novel problem solving in bending or unbending tasks has also been studied in non-human primates and children, as well as in one corvid and a parrot species (Anderson & Henneman, 1994; Beck et al., 2011; Cutting et al., 2011; Klüver, 1937; Laumer et al., 2017, 2018; Povinelli, 2000). Only one out of seven tested chimpanzees (*Pan troglodytes*) succeeded to unbend a tool (Povinelli, 2000), and in capuchins (*Cebus apella*), one out of three (Anderson & Henneman, 1994; Klüver, 1937) but the underlying mechanism remained opaque and, in the case of the chimpanzee, involved extensive scaffolding of her behavior (Povinelli, 2000).

In contrast to the tested chimpanzees, four out of five orangutans (*Pongo abelii*) could spontaneously unbend flexible tools, and two of them could also bend tools into hooks (Laumer et al., 2018). Most were successful from the first trial onwards and their actions appeared both spontaneous and goal directed. When the task necessitated steeper hooked tools, they were able to adjust their actions flexibly and create adequate tools (Laumer et al., 2018). Similarly, from 13 Goffin's cockatoos (*Cacatua goffiniana*), two spontaneously bent and another two spontaneously unbent tools and succeeded consistently thereafter (Laumer et al., 2017). Four rooks (*Corvus frugilegus*), a non-habitually tool using corvid species, have also been tested in a similar paradigm with a vertical tube containing food (Bird & Emery, 2009). They had previously been familiarized with wooden hooks and therefore were tested with pre-experience like the majority of our subjects. Three could manufacture hooks out of wire from their first trial and the fourth rook from his fourth trial onward. The authors argue that their behavior was a possible indication of "insight" as their manufacture was spontaneous without indication of some form of trial-and-error learning although the use of this term has been debated since (Kacelnik, 2009). Rutz et al. (2016) acknowledge this study as the most compelling case of spontaneous hook shaping in an experimental

context, given that a genetic predisposition is unlikely in this non-tool using corvid species. How spontaneous unbending in a problem-solving situation may be considered in NCCs, now known to bend flexible materials in nature frequently, depends on how commonly unbending of pliable materials may occur in the wild, and whether it is highly stereotypic or variable across individuals.

Our piloting of the role of pre-experience of tool shape in Experiment 1, albeit interesting in regards to Aigaios' results (Bending-trunk task), was too small for broader conclusions. Since we carried out the experiments, other publications made this point a feature of their methods. One out of three inexperienced orangutans and one out of 13 Goffin's cockatoos were able to manufacture hooks without pre-experience (Laumer et al., 2017, 2018) thus, were capable of truly innovating hooks as tools according to a tool innovation definition proposed by Cutting et al. (2011), because they could not have deduced the solution from a visual template. In children, pre-experience with the use of hooks increased significantly their ability to manufacture hooks from pliant material (Whalley et al., 2017), allowing children between 4 and 7 years old to perform nearly as well as 8-year-olds. Concerning spontaneous unbending, this has to date been achieved by only four orangutans, but by none of the 13 cockatoos. Overall, the effect of pre-experience on tool-modification tasks in non-human animals remains too little studied for drawing conclusive inferences.

In sum, our results of one bending and three unbending tasks spanning a total of 17 subjects seem to indicate that, for most successful subjects, instrumental learning underpinned their behavior. We therefore show that modifying pliant materials by bending and unbending is not just limited to a single individual NCC (Betty), but was present in at least six more tested subjects. To determine if pre-experience with functional tools affects performance in problem-solving tasks involving tool modification, or whether NCCs can invent tool shapes they have not used before, would require further adequately controlled studies with large sample sizes. Further, studies on wild NCCs should carefully monitor if any forms of unbending behavior are shown in both natural situations or in controlled experimental contexts, e.g., when operating bent sticks that obstruct their vision or disturb their grip against the cheeks. This will help to elucidate to what extent it is a natural behavior often encountered in the wild that may be a partially genetically predisposed tendency or rather an expression of their flexible problem-solving ability.

Author Contributions

Experiment 1. The design and apparatus was conceived by AvB. The experiments were carried out by JA with help by BM. Videos were coded by JA and MB but recoded by MvB. Data were analyzed by MvB with input by AvB, BM and ST. Experiment 2 and 3. The design and apparatus was conceived by MvB with input from AvB. The study was carried out by MvB and scored and analyzed by MvB with input by AvB and ST. The manuscript was written by MvB together with ST, AvB and BM. The manuscript has been seen and approved by all co-authors.

Acknowledgements

Data Accessibility

Raw data for all tasks can be found here.

References

Anderson, J. R., & Henneman, M. C. (1994). Solutions to a tool-use problem in a pair of Cebus apella. *Mammalia*, *58*(3), 351–362. https://doi.org/10.1515/mamm.1994.58.3.351

Beck, S. R., Apperly, I. A., Chappell, J., Guthrie, C., & Cutting, N. (2011). Making tools isn't child's play. *Cognition*, *119*(2), 301–306. https://doi.org/10.1016/j.cognition.2011.01.003

Beck, S. R., Cutting, N., Apperly, I. A., Demery, Z., Iliffe, L., Rishi, S., & Chappell, J. (2014). Is tool-making knowledge robust over time and across problems? *Frontiers in Psychology*, *5,* 1395. https://doi.org/10.3389/fpsyg.2014.01395

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*, *106*(25), 10370–10375. https://doi.org/10.1073/pnas.0901008106

Bluff, L. A., Weir, A. A. S., Rutz, C., Wimpenny, J. H., & Kacelnik, A. (2007). Tool-related cognition in New Caledonian crows. *Cognition and Behavior Reviews*, *2*, 1–25. https://doi.org/10.3819/ccbr.2008.20001

Borchert, J. (1998). Publication of the revision of the Animal Welfare Act of 25 May 1998. The Federal Minister of Food, Agriculture and Forestry. http://extwprlegs1.fao.org/docs/pdf/ger14753E.pdf

Brooks, M. E., Kristensen, K., van Benthem, K. J., Magnusson, A., Berg, C. W., Nielsen, A., Skaug, H. J., Maechler, M., & Bolker, B. M. (2017). glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *The R Journal, 9*(2), 378-400.

Chappell, J., & Kacelnik, A. (2004). Selection of tool diameter by New Caledonian crows *Corvus moneduloides*. *Animal Cognition*, *7*(2), 121–127. https://doi.org/10.1007/s10071-003-0202-y

Cutting, N., Apperly, I. A., & Beck, S. R. (2011). Why do children lack the flexibility to innovate tools? *Journal of Experimental Child Psychology*, *109*(4), 497–511. https://doi.org/10.1016/j.jecp.2011.02.012

Cutting, N., Apperly, I. A., Chappell, J., & Beck, S. R. (2014). The puzzling difficulty of tool innovation: Why can't children piece their knowledge together? *Journal of Experimental Child Psychology*, *125*, 110–117. https://doi.org/10.1016/J.JECP.2013.11.010

de Wit, S. & Dickinson, A. (2009). Associative theories of goal-directed behaviour: A case for animal-human translational models. *Psychological Research*, *73*(4), 463–476. https://doi.org/10.1007/s00426-009-0230-6

Emery, N. J., & Clayton, N. S. (2004). The mentality of crows: convergent evolution of intelligence in corvids and apes. *Science*, *306*(5703), 1903–1907. https://doi.org/10.1126/science.1098410

Emery, N. J., & Clayton, N. S. (2005). Evolution of the avian brain and intelligence. *Current Biology*, *15*(23), R946–50. https://doi.org/10.1016/j.cub.2005.11.029

Gamer, M., Jim Lemon, J., Fellows, I. & Singh, P. (2019). irr: Various coefficients of interrater reliability and agreement. R package version 0.84.1. https://CRAN.R-project.org/package=irr

Holzhaider, J. C., Hunt, G. R., Campbell, V. M., & Gray, R. D. (2008). Do wild New Caledonian crows (*Corvus moneduloides*) attend to the functional properties of their tools? *Animal Cognition*, *11*(2), 243–254. https://doi.org/10.1007/s10071-007-0108-1

Hunt, G. R. (1996). Manufacture and use of hook-tools by New Caledonian crows, *Nature,* 379, 249–251. https://doi.org/10.1038/379249a0

Hunt, G. R. (2014). New Caledonian crows' (*Corvus moneduloides*) Pandanus tool designs: Diversification or independent invention? *The Wilson Journal of Ornithology*, *126*(1), 133–139. https://doi.org/http://dx.doi.org/10.1676/13-085.1

Hunt, G. R., & Gray, R. D. (2002). Species-wide manufacture of stick-type tools by New Caledonian crows. *Emu*, *102*(4), 349–353. https://doi.org/10.1071/MU01056

Hunt, G. R., Holzhaider, J. C., & Gray, R. D. (2012). Prolonged parental feeding in tool-using New Caledonian crows. *Ethology*, *118*(5), 423–430. https://doi.org/10.1111/j.1439-0310.2012.02027.x

Hunt, G. R., Rutledge, R. B., & Gray, R. D. (2006). The right tool for the job: What strategies do wild New Caledonian crows use? *Animal Cognition*, *9*(4), 307–316. https://doi.org/10.1007/s10071-006-0047-2

Jelbert, S. A., Hosking, R. J., Taylor, A. H., & Gray, R. D. (2018). Mental template matching is a potential cultural transmission mechanism for New Caledonian crow tool manufacturing traditions. *Scientific Reports*, *8*(1), 1–8. https://doi.org/10.1038/s41598-018-27405-1

Kacelnik, A. (2009). Tools for thought or thoughts for tools? *Proceedings of the National Academy of Sciences of the United States of America,* 106, 10071–10072. https://doi.org/10.1073/pnas.0904735106

Kenward, B., Rutz, C., Weir, A. A. S., & Kacelnik, A. (2006). Development of tool use in New Caledonian crows: Inherited action patterns and social influences. *Animal Behaviour*, *72*(6), 1329–1343. https://doi.org/10.1016/j.anbehav.2006.04.007

Kenward, B., Weir, A. A. S., Rutz, C., & Kacelnik, A. (2005). Behavioural ecology: Tool manufacture by naive

juvenile crows. *Nature*, *433*(7022), 121–121. https://doi.org/10.1038/433121a

Klump, B. C., Sugasawa, S., St Clair, J. J. H., & Rutz, C. (2015). Hook tool manufacture in New Caledonian crows: Behavioural variation and the influence of raw materials. *BMC Biology*, *13*(1), 97. https://doi.org/10.1186/s12915-015-0204-7

Klüver, H. (1937). Re-examination of implement-using behavior in a cebus monkey after an interval of three years. *Acta Psychologica*, *2*, 347–397. https://doi.org/10.1016/S0001-6918(37)90014-3

Knaebe, B., Taylor, A. H., Elliffe, D. M., & Gray, R. D. (2017). New Caledonian crows show behavioural flexibility when manufacturing their tools. *Behaviour*, *154*(1), 65–91. https://doi.org/10.1163/1568539X-00003411

Laumer, I. B., Bugnyar, T., Reber, S. A., & Auersperg, A. M. I. (2017). Can hook-bending be let off the hook? Bending/unbending of pliant tools by cockatoos. *Proceedings of the Royal Society of London. Series B: Biological Sciences, 284*(1862), 20171026. https://doi.org/10.1098/rspb.2017.1026

Laumer, I. B., Call, J., Bugnyar, T., & Auersperg, A. M. I. (2018). Spontaneous innovation of hook-bending and unbending in orangutans (*Pongo abelii*). *Scientific Reports*, *8*(1), 16518. https://doi.org/10.1038/s41598-018-34607-0

Neldner, K., Mushin, I., & Nielsen, M. (2017). Young children's tool innovation across culture: Affordance visibility matters. *Cognition*, *168*, 335–343. https://doi.org/10.1016/j.cognition.2017.07.015

Neuringer, A., Kornell, N., & Olufs, M. (2001). Stability and variability in extinction. *Journal of Experimental Psychology: Animal Behavior Processes*, *27*(1), 79–94. https://doi.org/10.1037/0097-7403.27.1.79

Povinelli, D. (2000). *Folk physics for apes: The chimpanzee's theory of how the world works*. https://philpapers.org/rec/POVFPF

R Core Team. (2018). R: A language and environment for statistical computing [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/

Rutz, C., Sugasawa, S., van der Wal, J. E. M., Klump, B. C., & St Clair, J. J. H. (2016). Tool bending in New Caledonian crows. *Royal Society Open Science*, *3*(8), 160439. https://doi.org/10.1098/rsos.160439

Shettleworth, S. (2009). *Cognition, evolution, and behavior*. Oxford University Press.

St. Clair, J. J. H. & Rutz, C. (2013). New Caledonian crows attend to multiple functional properties of complex tools. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *368*(1630), 20120415. https://doi.org/10.1098/rstb.2012.0415

Taylor, A. H., Hunt, G. R., Holzhaider, J. C., & Gray, R. D. (2007). Spontaneous metatool use by New Caledonian crows. *Current Biology*, *17*(17), 1504–1507. https://doi.org/10.1016/j.cub.2007.07.057

Taylor, A. H., Knaebe, B., & Gray, R. D. (2012). An end to insight? New Caledonian crows can spontaneously solve problems without planning their actions. *Proceedings of the Royal Society B: Biological Sciences*, *279*(1749), 4977–4981. https://doi.org/10.1098/rspb.2012.1998

Taylor, A., Roberts, R., Hunt, G., & Gray, R. (2009). Causal reasoning in New Caledonian crows. *Communicative & Integrative Biology*, *2*(4), 311–312. https://doi.org/10.4161/cib.2.4.8224

Tebbich, S., Taborsky, M., Fessl, B., & Blomqvist, D. (2001). Do woodpecker finches acquire tool-use by social learning? *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *268*(1482), 2189–2193. https://doi.org/10.1098/rspb.2001.1738

Teschke, I., Wascher, C. A. F., Scriba, M. F., von Bayern, A. M. P., Huml, V., Siemers, B., & Tebbich, S. (2013). Did tool-use evolve with enhanced physical cognitive abilities? *Philosophical Transactions of the Royal Society B: Biological Sciences*, *368*(1630). https://doi.org/10.1098/rstb.2012.0418

von Bayern, A. M. P., Danel, S., Auersperg, A. M. I., Mioduszewska, B., & Kacelnik, A. (2018). Compound tool construction by New Caledonian crows. *Scientific Reports*, *8*(1), 1–8. https://doi.org/10.1038/s41598-018-33458-z

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*(22), 1965–1968. https://doi.org/10.1016/j.cub.2009.10.037

Weir, A. A. S., Chappell, J., & Kacelnik, A. (2002). Shaping of hooks in New Caledonian crows. *Science, 297*(5583), 981. https://doi.org/10.1126/science.1073433

Weir, A. A. S. & Kacelnik, A. (2006). A New Caledonian crow (*Corvus moneduloides*) creatively re-designs tools by bending or unbending aluminium strips. *Animal Cognition*, *9*(4), 317–334. https://doi.org/10.1007/s10071-006-0052-5

Whalley, C. L., Cutting, N., & Beck, S. R. (2017). The effect of prior experience on children's tool innovation. *Journal of Experimental Child Psychology*, *161*, 81–94. https://doi.org/10.1016/J.JECP.2017.03.009

Wood, S. N. (2011) Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society (B), 73*(1),3-36.