Note: this is a draft version of the manuscript published in Animal Cognition *(DOI 10.1007/s10071-006-0052-5). The original publication is available at http://www.springerlink.com.*

A New Caledonian crow (*Corvus moneduloides*) creatively re-designs tools by bending or unbending aluminium strips

ALEX A. S. WEIR (✉), ALEX KACELNIK
Department of Zoology, South Parks Road, Oxford, OX1 3PS, UK

e-mail: alex.weir@magdalen.oxon.org

This contribution is part of the special issue “Animal Logics” (Watanabe and Huber 2006).

Abstract

Previous observations of a New Caledonian crow (*Corvus moneduloides*) spontaneously bending wire and using it as a hook [Weir et al. *(2002)* Science 297:981] have prompted questions about the extent to which these animals ‘understand’ the physical causality involved in how hooks work and how to make them. To approach this issue we examine how the same subject (“Betty”) performed in three experiments with novel material, which needed to be either bent or unbent in order to function to retrieve food. These tasks exclude the possibility of success by repetition of patterns of movement similar to those employed before. Betty quickly developed novel techniques to bend the material, and appropriately modified it on four of five trials when unbending was required. She did not mechanically apply a previously learned set of movements to the new situations, and instead sought new solutions to each problem. However, the details of her behaviour preclude concluding definitely that she understood and planned her actions: in some cases she probed with the unmodified tools before modifying them, or attempted to use the unmodified (unsuitable) end of the tool after modification. Gauging New Caledonian crows’ level of understanding is not yet possible, but the observed behaviour is consistent with a partial understanding of physical tasks at a level that exceeds that previously attained by any other non-human subject, including apes.

Electronic Supplementary Material

Supplementary material is available in the online version of this article at http://dx.doi.org/10.1007/s10071-006-0052-5 and is accessible for authorized users. Movie clips are also available on the authors’ website: http://www.NewCaledonianCrow.com (navigate to Tool Use > Photos and Movies).

Key words

New Caledonian crows; tool use; tool modification; planning; folk physics

Note: this is a draft version of the manuscript published in Animal Cognition *(DOI 10.1007/s10071-006-0052-5). The original publication is available at http://www.springerlink.com.*

Introduction

The act of making functional artefacts is often thought to be especially revealing about cognitive processes, because it may require reference to both the representation of the problem and the expected future use of the artefact. However, this assumption is not always valid: the artefact maker might be simply following action rules acquired by the species through natural selection. For example, antlion larvae (*Myrmeleon crudelis*) build up their traps by stereotyped movements in sandy soil (Lucas 1982), but there is (probably) no problem representation other than in the rules encoded in the DNA of the actor. In other cases, individuals may learn through trial-and-error, or by observing others, what sequences of actions modify the artefact effectively, but again with no cognitive representation of the problem or planning the future use of the instrument. This process may be responsible for some green-backed herons (*Ardeola striata*) acquiring the behaviour of using bait to attract fish (e.g. Higuchi 1986; Higuchi 1988).

New Caledonian crows (*Corvus moneduloides*) are renowned for their highly sophisticated and diverse tool manufacture and use, both in the wild (Hunt 1996; Hunt and Gray 2002, 2003, 2004a, b; Hunt et al. 2006) and in captivity (Chappell and Kacelnik 2002; Weir et al. 2002; Chappell and Kacelnik 2004; Weir et al. 2004), but careful experimentation is necessary to determine whether this behaviour is cognitively different from cases in which the behaviours are acquired either as inherited rules or through shaping by reinforcement, as in the examples of antlion larvae or herons mentioned above. This is particularly pertinent in light of recent observations that basic tool use and manufacture ('tool-oriented behaviour') develops in New Caledonian crows even if they are reared in isolation (Kenward et al. 2005), and is preceded by stereotyped, apparently non-functional, 'precursor' behaviours, although tutoring (by human foster parents) does increase the frequency of tool-related behaviours in juveniles (Kenward et al. in press). This complex developmental path highlights the difficulty of establishing whether tool-oriented behaviour involves cognitive representation of the task, planning with foresight, and goal-directed problem solving (also see Watanabe and Huber 2006).

The creation or specific modification of novel objects to solve novel problems, by re-organizing individuals' own experience rather than following pre-programmed species-typical rules, would be good evidence for some of these faculties. However, it is hard in practice to test for this ability: it is unrealistic to expect any agent to be capable of solving entirely novel problems with objects or materials they have never had any experience with. Furthermore, inherited predispositions are likely to play a significant role in how the solution is reached, but this does not automatically imply that cognition is not involved. For example, human infants seem to inherit propensities which cause them to bang objects together (e.g. Thelen 1979, 1981), and through such actions learn about properties of objects (Lockman 2000), such as their hardness or flexibility. It is only as a result of the combination of such predispositions and trial-and-error learning that older humans are then able to, for example, use stones as hammers, but this does not mean that hammer use by adult humans does not involve goal-representation. Specific features of how the behaviour is acquired can help elucidate the underlying cognitive abilities. The more innovative the actions and the more specific, deliberate, and unusual the modification of the raw material, the more acceptable it becomes to hypothesise that the agent's behaviour is controlled by cognitive representation of a definite goal and the means of reaching it.

Several of these features were demonstrated by a New Caledonian crow ('Betty') studied by Weir and colleagues (2002). Betty spontaneously and repeatedly bent straight wire into hook-like shapes, and appropriately used these to retrieve a small bucket (containing food) from a vertical tube. Firstly, to a human observer Betty's actions appeared to be **deliberate**: the action used to bend the wire was highly distinct, bending did not occur as a result of general manipulation of the tool, and the bent wire was always used to retrieve the bucket immediately after modification. Secondly, they were **specific**: she never made anything similar when not facing the problem, and despite using several different techniques the final tools were all of similar shapes. Thirdly, they were **novel**: wire, a material that is suitable to this treatment because it is pliable and retains its shape after bending, does not seem to share these properties with materials regularly found in the species' habitat or familiar to that individual, and the actions used to bend the wire do not resemble other actions known to be performed by New Caledonian crows. Betty's performance has been widely cited as an example of unexpectedly complex cognition in birds (e.g. Defeyter and German 2003; Sterelny 2003; Emery and Clayton 2004; Emery et al. 2004; Griffin and Speck 2004; Ricklefs 2004; Jarvis et al. 2005), and has been claimed to provide evidence for animal insight (e.g. Butler et al. 2005; Zorina 2005).

However, no single experiment can definitively establish the level of comprehension of physical causality ('folk physics') by an organism. Experience is necessary both for learning appropriate actions and for making inferences about physical laws, and it is often nearly impossible to map precisely the level of abstraction at which experience is recruited to solve each specific new problem. Establishing the distinction between behavioural control by procedural rules and problem-solving by application of "high level" (*sensu* Povinelli 2000) logical inferences requires subjects to be tested on a series of tasks: if subjects systematically find new solutions to transformations of the problems they face without prolonged periods of trying old routines with random modifications, an attribution of understanding and planning becomes justified. Here we used two such modifications. The first required a functional transformation of the potential tool similar to the one already observed (making a hook) but in a different material, so that different movements were necessary to achieve a similar end product. This could serve to exclude the possibility of 'low level' (non-cognitive) control. The second required shaping this new material in the opposite direction (unbending a bent object), in a task similar to some explored in primates (reviewed below).

Inferences about the level of 'understanding' (vs. 'shaping') that lead to changes in behaviour can be informed by the time course of acquisition of solutions to novel problems. Given the complexity of our tasks, we speculate that a slow and gradual increase in proficiency would indicate that the subject relies on within-task trial-and-error learning. In contrast, immediate or step-wise acquisition would suggest that, at a minimum, the subject generalises from concepts formed during earlier experience in related tasks. The need to adjust multiple aspects of behaviour in a sequence means that in a small number of trials the amount of behavioural variation is unlikely to provide sufficient data for a trial-and-error process to reach the required solutions.

In summary, the greater the understanding of the problem (namely, the degree to which abstracted general principles play a role), the greater should be the flexibility shown to produce novel transformations in line with new demands.

Control over tool shape in primates and birds

There have been few experimental investigations into the cognitive control animals have over the shape of manufactured tools, both because few animals naturally make tools at all, and because no wild vertebrates apart from humans and New Caledonian crows make tools with a precisely-determined final shape (Hunt 1996, 2000; Hunt and Gray 2004a; reviewed in Kacelnik et al. 2006). Chimpanzees modify the shape of probing tools and leaf sponges to make them functional, but the final shape does not seem to be controlled in detail and the modification involves relatively non-specific actions such as chewing the ends of the sticks to make brushes for termite-dipping (Sugiyama 1985, 1997; Sanz et al. 2004), removing leaves and twiglets from twigs to make probing tools (Goodall 1986), or chewing and crumpling leaves up to make sponges (McGrew 1992; Sugiyama 1997; Tonooka 2001). In all of these examples, the final tool shape is either inherent in the structure of the raw material (e.g. the petiole of a leaf), or is non-specific (a crumpled ball of leaves; a chewed end of a stick), whereas the stepped-cut pandanus tools made by New Caledonian crows (Hunt 1996; Hunt and Gray 2004b), and perhaps to a lesser extent, the hooked-stick tools they 'craft' (Hunt and Gray 2004a), involve the imposition of precise shape onto an unstructured substrate.

To our knowledge, the only studies that have examined the performance of non-human animals in tasks requiring tools to be shaped by bending or similar actions are one series of experiments involving chimpanzees (*Pan troglodytes*; Povinelli et al. 2000b), and two less detailed experiments with capuchin monkeys (*Cebus* spp.; Klüver

1937; Anderson and Henneman 1994). Their findings are described in more detail below, because they relate very closely to the experiments we report in this paper.

Povinelli's group tested seven juvenile chimpanzees on a task where they could obtain an apple by inserting a tool through a small hole. They were provided with a piece of flexible piping, either straight (where both ends could be pushed through the hole), or bent into S- or C-shapes (which could not be pushed through the hole). Consequently, when the tool was S- or C-shaped, subjects had to modify one of the ends of it in order to use it effectively. In the first experiment (Povinelli et al. 2000b Experiment 24), two apes each modified the tool in one of four test trials (interspersed with probe trials with straight tools), but one of them (Jadine) did not modify it sufficiently, and the other (Kara) only used the unmodified end; neither successfully retrieved the apple.

Povinelli and colleagues subsequently (in Experiment 25) demonstrated the solution to the subjects: in each trial, the tool was initially C-shaped, but in the first trial in each session the experimenter unbent both ends of the tool, and gave the straight tool to the subjects; in the second trial, the experimenter unbent only one end; and in the third (test) trial, only the unmodified C-shaped tool was presented. Three apes (Kara, Candy, and Jadine) did modify the tool on five test trials, but again never retrieved the apple with it, and on two occasions first directed the unmodified end towards the apparatus. Moreover, they had some difficulty even in trial 2: their success rate dropped from 100% (on trial 1) to 60%, mainly because they preferred (69% of the time) to orient the hooked end of the tool towards the apparatus. In Experiment 26, the researchers attempted to scaffold their subjects' responses to an even greater extent, by explicitly training them to bend the tool in the experimental chamber (but in the absence of the probing apparatus). When the sequence of trials from Experiment 25 was repeated after the scaffolding, the subjects still preferentially tried to insert the hooked end of the tool in trial 2 (64% of the time), but in the test trials one subject (Jadine) did consistently modify the tool with her hands and use it to retrieve the apple. Three other subjects also modified the tools, but two of their modifications appeared to occur incidentally as they poked at the apparatus, and one used the modified end as a handle, attempting to probe with the unmodified (and ineffective) end; none managed to retrieve the apple as a result.

Klüver's (1937) and Anderson and Henneman's (1994) investigations into tool shaping were less detailed, and provided conflicting results. Klüver tested how one captive capuchin monkey performed on over 300 problems, one of which required a circular wire to be unbent to obtain food; although the subject did (apparently unintentionally) open the wire slightly, this was not enough to retrieve the food. In one of eight tasks presented to two captive capuchins by Anderson and Henneman, the subjects were provided with loops of soldering-wire, which had to be straightened out to fit into a honey-dipping apparatus. The male subject did straighten the tool a few seconds after the start of each trial, and used it immediately afterwards in the honey-dipping apparatus, whereas the female rarely contacted the wire and never straightened it.

In summary, from three studies investigating whether non-human primates could unbend tools, two individuals (one chimpanzee and one capuchin) did deliberately straighten the tools. However, the chimpanzee was only successful after extensive scaffolding of her response by the experimenter, and considering her performance on other similar tasks, the authors concluded that she "came to understand some very specific features of the tool configuration that was necessary to solve the tool-insertion problem, as opposed to reasoning about an abstract conception of 'shape'" (Povinelli et al. 2000b, p. 295). These negative results should not, though, be taken as evidence that chimpanzees are incapable of developing an understanding of the relevant physical concepts and modifying the shape of tools accordingly: all seven subjects involved in the experiments were relatively young at the time of testing, and had been reared in unnatural conditions (Anderson 2001; Hauser 2001; Whiten 2001; Machado and Silva 2003). Regarding the capuchin subject, Anderson and Henneman (1994) do not provide enough details of the initial acquisition of the wire-unbending to allow conclusions to be drawn as to the extent of 'mental representation' involved in its behaviour. However, the wire was described as "soldering-wire", which is so flexible that it might unbend without any deliberate, goal-directed attempt to modify it—the unbending might have happened solely as a consequence of grabbing the wire to use it.

A few experiments have tested whether non-human animals will modify tool shape using other techniques, not involving bending or unbending. For example, Visalberghi and colleagues found that chimpanzees, bonobos (*Pan paniscus*), an orangutan (*Pongo pygmaeus*), and capuchin monkeys would remove transverse cross-pieces from a piece of dowelling, which allowed them to insert the dowelling into a tube to push food out (Visalberghi and Trinca 1989; Visalberghi et al. 1995). However, in all the experiments the subjects made errors (such as inserting the cross-pieces after removing them, rather than the now-functional tool), and in Visalberghi et al. (1995) the cross-pieces were inserted into the dowelling in such a way that they would fall out if the dowelling was rotated; a similar experiment by Povinelli and colleagues (Povinelli et al. 2000b, experiment 27) found that chimpanzees had a general tendency to modify tools if they were not instantly successful, even if the modification served to make the tools less functional. It therefore seems likely that the tool modification shown was not specific to the task, but the result of a general tendency to disassemble tools when possible. One of the capuchins ("*Cm*") in Visalberghi and Trinca's (1989) study did, from the first trial, modify a reed that was too wide to fit into a tube by biting and hitting hit it forcefully, and used one of the smaller pieces as a tool. However, in later trials the subject frequently attempted to use the *larger* splinter that was still too thick, in addition to continued attempts to insert the intact reed, suggesting that the tool modification seen was not related to the task, but instead a result of non-specific tool manipulation.

Two bird species have also been tested in tool modification tasks. Tebbich and Bshary (2004) tested five woodpecker finches (*Cactospiza pallida*) on the H-stick task described above, and a similar task involving natural tools (dry twigs of *Scutia spicata* with thorns projecting near each end). Three of their subjects learned to remove the transverse sticks in the first task after 14-21 trials (the other two never did so), and all apart from one removed the thorns from the natural tools from the first trial with them (this experiment immediately followed the H-stick one). However, all subjects continued to make frequent errors throughout the experiments, including continuing to attempt to insert the unmodified tool. In contrast, two captive New Caledonian crows made tools of significantly narrower diameters when they needed to push them through narrower holes (Chappell and Kacelnik 2004, Experiment 2), and 27 of their 30 successful tools were of the appropriate diameter before they were first used.

To our knowledge, therefore, there is at present no conclusive evidence that any non-human animal apart from the New Caledonian crow is able to solve tasks requiring them to bend or unbend tools, or to solve tasks requiring precise control over tool shape, and nothing is known of the process by which such behaviours may be acquired.

Here, through three modifications of a problem that had previously led to hook-making by bending wire (Weir et al. 2002), we explore whether our subject succeeds in developing novel solutions and we examine the process by which her behaviour changes when the situation is modified. Although predicting precise behaviour in novel situations is impossible, we would expect that an agent whose behaviour is guided by comprehension of causal relationships between objects would make ‘relevant’ modifications to the tool from the start (i.e. before any specific actions are rewarded), whereas one reliant solely on improving performance adaptively though its consequences would gradually converge to a suitable sequence of actions as a function of its own sporadic successes.

Note: this is a draft version of the manuscript published in Animal Cognition *(DOI 10.1007/s10071-006-0052-5). The original publication is available at http://www.springerlink.com.*

Experiment 1: bending novel material

In this experiment we introduced a novel material while maintaining the original problem of obtaining food using a hook. The experiment used the same subject (Betty) as the previous wire-bending study (Weir et al. 2002) and addressed three inter-related questions:

1. What did the subject know about the relationship between tool shape and success at retrieving the bucket (i.e. did she understand that hook-like structures are necessary / most efficient)?
2. What did she understand about the link between modification technique and tool shape (i.e. the specific effect(s) her actions had upon the resulting shape of the tool)?
3. To what extent was she aware of the connection between (1) and (2) above, namely, her manipulation of the tool, and the efficiency with which it achieved its goal?

Since Betty had no *a priori* way of knowing that the new material required a different technique to modify it, we expected that initially she would tackle it with similar movements to those employed with the original wire. However, following the discovery that the new material did not respond well to that treatment, the process of acquiring new techniques would differ considerably if it were guided by success of random variations or directed by pre-figuring the functional outcome. Shaping random behaviour by its consequences until an appropriate chain of actions is acquired should demand a very large number of trials, since initially only the last actions would be proximate to the reward. In general, when training an animal in an arbitrary operant task involving sequences of behaviour, reward must influence intermediate stages rather that feedback being received from final success alone (e.g. Mackintosh 1994; Schwartz et al. 2002). For example, Epstein (1984) showed that pigeons spontaneously solve an analogous problem to Köhler's chimpanzees' (Köhler 1925) of pushing a box underneath a reward and standing on it to reach a manipulandum that could yield a reward, but only because they had been explicitly trained over many trials on each part of the sequence separately.

Methods

Subject and housing

The history of our subject (Betty) was described in Chappell and Kacelnik (2002) and Weir et al. (2002). At the time of the present experiments she was housed with four other crows (with one of whom she had paired and mated earlier in the year) with permanent access to indoor (4.00 × 2.50 × 2.50 m high) and outdoor aviaries (2.80 × 2.80 × 2.50 m high). The accommodation contained many perches of varying widths and heights. Plastic children's toys provided environmental enrichment, and tree

Note: this is a draft version of the manuscript published in Animal Cognition *(DOI 10.1007/s10071-006-0052-5). The original publication is available at http://www.springerlink.com.*

branches provided sources for tool-making. Drinking and bathing water were permanently available. The crows were fed *ad libitum* on soaked cat biscuits (Go-Cat®), an insect and fruit mix (Orlux® Universal and Orlux® granules), peanuts, and mealworms. They were encouraged to use tools regularly by making some of their preferred food otherwise inaccessible: mealworms were placed in holes drilled into tree stumps, and occasionally pieces of pig heart were placed in clear acrylic tubes that were left in the aviaries.

Experimental room

Experiments took place in a separate testing room (2.00 × 2.80 × 2.50 m high), which communicated with the indoor aviary via two adjacent openings (160 × 180 mm high) with hanging ‘bob-wires’, one serving as an entrance and the other as an exit. Partitions inside the testing room prevented the birds seeing the table where experiments were carried out until they had entered the testing room. The ‘entry’ bob-wires could be locked by means of a custom-built system, whereas the ‘exit’ ones were permanently open. By locking the entrance after a subject had entered the testing room, it was possible to test birds individually without trapping them in the testing room since they were always free to leave. For a period of several months before experiments began, the birds were accustomed to enter and leave the testing room at will by regularly provisioning it with favoured food and signalling the presence of food (or an experiment) by means of a red LED inside the regular aviary compartment.

While voluntary participation with freedom to leave the room at will ensures that subjects are motivated and unstressed (stress can impair performance in cognitive experiments; e.g. de Kloet et al. 1999), it also has the disadvantage of reducing control over which bird from the group would enter. Consequently, for many sessions Betty was temporarily isolated in the main indoor aviary compartment. The experimental apparatus were placed on a table (1.00 × 1.15 × 1.00 m high) placed against a darkened translucent acrylic wall which, with illumination only on the bird’s side, served as one-way observation window.

Apparatus

The apparatus consisted of a small bucket (made from an empty film canister) with a plastic ‘handle’ attached by sticky Gaffa® tape at opposite ends of the rim, placed at the bottom of a vertical transparent tube (as described in Weir et al. 2002). The bottom of the tube was fixed to a small ceramic bowl and secured in the centre of a plastic feeding tray using sticky tape. A brick in the tray next to the tube immobilised the whole set up and served as a stand to allow the bird to probe inside the tube.

Thin strips of aluminium with rounded blunt corners were provided as material from which to make tools (usually 90 mm long and 3.5 or 5.0 mm wide; see Table S1 for precise dimensions in each trial). The strips were easily pliable but could only be

Note: this is a draft version of the manuscript published in Animal Cognition *(DOI 10.1007/s10071-006-0052-5). The original publication is available at http://www.springerlink.com.*

bent in one plane, due to their rectangular cross-section. The strip was either placed horizontally on top of the tube (Trials 1-26), or lodged in a hole in a wooden block (10 × 6 × 6 cm deep) fixed to the table (Trials 27 onwards). A new strip was provided for each trial.

Table 1 Experimental timetable

Date	Experiment	Trials
20 August 2004	1	1–12
23 August 2004	1	13–14
25 August 2004	1	15–23
27 August 2004	1	24–25
14 September 2004	1	26–31
14 September 2004	2	1–3
14 September 2004	3	1–2
28 February 2005	1	32–34
28 February 2005	3	3–4

This table shows the date each trial was carried out, for all experiments. See the ‘Procedure’ section of each experiment for more details.

Procedure

No pre-training was given, since Betty was already familiar with the apparatus. She had been presented with the apparatus and ordinary wire several times since the experiments reported in Weir et al. (2002), but had not been exposed to wire in the 6 months preceding this experiment. She had, however, been presented with the apparatus and straight, stick tools (with which she could retrieve the bucket) in the 3 weeks preceding this experiment. Trials were performed between 20 August 2004 and 28 February 2005. Normal food was removed from the aviary 1-2 hours before experiments began, and was replaced immediately after each session. Before each trial the bucket, containing a small piece of pig heart (0.5 ± 0.1 g) and/or a waxmoth larva (the reward was varied to maintain motivation), was positioned in the apparatus. The experimenter then unlocked the entry bob-wires and switched on the LED signalling to the subject the beginning of an experiment. All trials were videotaped through the observation window using a mini-DV camcorder (Canon DM-MV550i or Canon XL1); the final shape of the tool was also videotaped against a standard background, and all modified strips were numbered and retained for later analysis.

Trials were terminated either 10 minutes after the subject first picked up or dislodged the aluminium strip, or earlier if the subject left the testing room. A “trial” was only scored if the subject interacted with the apparatus. Trials for experiments 1, 2, and 3 overlapped to some extent, as described in Table 1. Notice that the last three trials of experiment 1 took place after a gap of five months in testing. The variation in the number of trials on each day is primarily due to the voluntary participation, since on some days Betty entered the testing room more frequently than on others.

Scoring and analysis

All scoring was done from videotapes. Each trial was summarised descriptively, and the following measures were recorded:

- Success (whether or not food was obtained)
- Trial duration (interval between first contact with apparatus and food extraction, excluding time when the subject was not interacting with the apparatus or tool)
- Duration of probing with unmodified tool (probing defined as one end of the tool inserted into the tube)
- Whether or not the strip was modified
- Latency between first interaction with the apparatus and first modification of the strip
- Method of modification, with the following components:
    - place where the modification took place
    - position along the strip that was modified (proximal, middle or distal respect to the place where the strip was held in the beak)
    - modification technique (either ‘twist’, where the tool was held at an angle part of the way along it and twisted around the beak; or ‘bend’, where the tool was held at one end in line with its main axis, and bent by moving the beak up and towards the tool)
- Length of time spent modifying the strip (‘tool crafting time’, defined as the length of time from the first moment the tool started to bend until the last, excluding interruptions)
- The end of the modified tool first used to probe for food
- Duration of probing with each end of the modified tool
- Whether and on how many occasions the tool was turned around
- Final shape of the tool (photographed). We designed a score for quality of the final product. Although neatness of design to a human observer does not unequivocally imply functional quality, this score is used to examine the process of acquisition. We gave each final tool a score of either 1 or -1 (the positive score for the better shaped tools), according to the following criteria: a positive was scored if the tool had a bend of more than 90° within 1/3 of the strip from either end and was not grossly distorted, (e.g. helical or ‘knotted’ on the modified end), and a negative otherwise.

The main indicators of performance we were interested in can be grouped into three loose categories. The first one is trial duration—i.e. how long until the subject retrieved

Note: this is a draft version of the manuscript published in Animal Cognition *(DOI 10.1007/s10071-006-0052-5). The original publication is available at http://www.springerlink.com.*

the food (did she become quicker over time?). The second category reflects the subject's understanding of what kind of tool is necessary: the length of time probing with the tool before modifying it (did the subject recognise that the unmodified tool was not appropriate?), and how long the subject spent probing with the unmodified end of the tool (did she understand that she needed to *use*, rather than just *make*, a hook?). The third category measures aspects of the modification itself: the length of time the subject spent modifying the material (did she become more skilful over time?), the technique she used (did this change across trials?), and the final shape of the tool (did this become more regular and hook-like across trials?).

To examine changes in performance over time, latency and duration measures (in seconds) were natural-log transformed (0.1 was added to all values before transformation, to eliminate errors due to zero values) and used as the dependent variables in separate general linear models (GLMs). Experimental day (a number from 1–5, shared by all trials carried out on the same day) and trial-within-day, as well as the interaction between them, were used as continuous explanatory variables. Residual plots were visually inspected to check that the assumptions of normality of error, homogeneity of variance, and linearity were satisfied. Due to non-orthogonality, if the interaction was not significant the model was re-fitted without the interaction and it is these results that are reported.

Additionally, a GLM was used to assess whether time spent probing with the modified end of the tool in successful trials was related to the hook 'score' (as a categorical variable) and, to test if tool design improved with practice, a Spearman Rank Correlation was used to examine whether cumulative hook score increased across trials. The other measures are only presented graphically and not statistically analysed, since formal analysis would not be any more informative than visual inspection.

Trials were classified into one of six categories, depending on two dimensions: the tool used ("own" tool—not the aluminium strip, but instead a twig or feather brought from the main aviary; "modified" or "unmodified" aluminium strips) and whether or not the subject successfully retrieved food ("success", "failure"). Different analyses use different subsets of trials: for example, analysis of the latency before modifying the tool includes only "modified" trials (successful and unsuccessful combined).

Note: this is a draft version of the manuscript published in Animal Cognition *(DOI 10.1007/s10071-006-0052-5). The original publication is available at http://www.springerlink.com.*

Results

Betty adapted quickly to the new material, and started to modify and use the strips to retrieve the bucket from the third trial. She retrieved the bucket on 25 of the 34 trials using the strips; out of the 9 remaining trials, in 4 she retrieved the bucket using a twig or feather she brought into the testing room with her, rather than the metal strip, and in the other 5 she dropped the metal strip irretrievably into the tube or behind the brick (Fig. 1). She developed a technique for modifying the new material that differed from that she had previously used with garden wire: she acted on the *proximal* end of the tool (i.e. the end held in her beak), whereas with wire she usually bent the *distal* end of it by wedging the tip and pulling sideways from the proximal end, levering the wire around the tube or other objects. Her general performance and detailed modification of the tool are discussed in the next two sections (her detailed behaviour in each trial is described in Table S1, and the photographs of the final shape of each tool she modified are shown in Fig. S4).

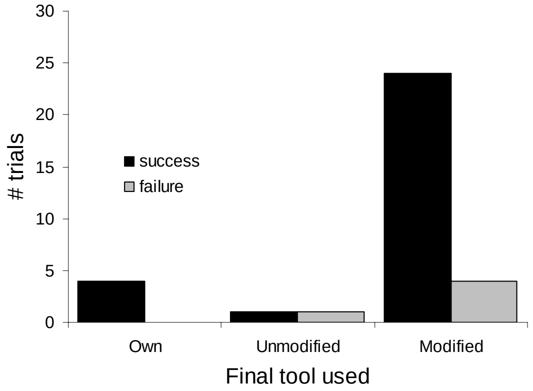


Fig. 1 Success retrieving the bucket with different tools. When more than one tool was used in a trial (for example, if the subject attempted to use the aluminium strip, but then finally retrieved the bucket with a twig tool), only the last tool is counted. "Own" means that the subject used a twig or feather tool brought in from the aviary, and "Unmodified" and "Modified" refer to the aluminium strip tools. "Success" (filled bars) means that the subject retrieved the bucket, whereas "Failure" indicates that she did not (these trials were normally terminated by the subject dropping the tool into the tube or another inaccessible location).

Overall performance

Betty first modified the strip and successfully retrieved the bucket with it on Trial 3, and thereafter modified it on all but two trials (on Trial 1 she ignored the aluminium strip and retrieved the bucket with a feather shaft; on Trial 2 she probed for 2 seconds with the unmodified strip and then dropped it irretrievably into the tube). The duration of successful trials halved between Trials 3 and 4, and again between 4 and 7 (in Trial 5 she succeeded with the unmodified strip, and in Trial 6 she twisted it into a 'helical' shape (see Fig. S4 tool 6) but then dropped it irretrievably into the tube). There was a statistically significant interaction between the time to bucket retrieval (Fig. 2a) across days and within days (day*trial: $F_{1,20} = 4.85$, $p = 0.040$, successful trials only), which was due to the fact that on the first day the time to success fell much more steeply with trial number than on the other days. This effect was dependent on the first successful trial with the modified strip (Trial 3): excluding this trial, the interaction was not significant (day*trial: $F_{1,19} = 2.86$, $p = 0.107$), and the model without the interaction

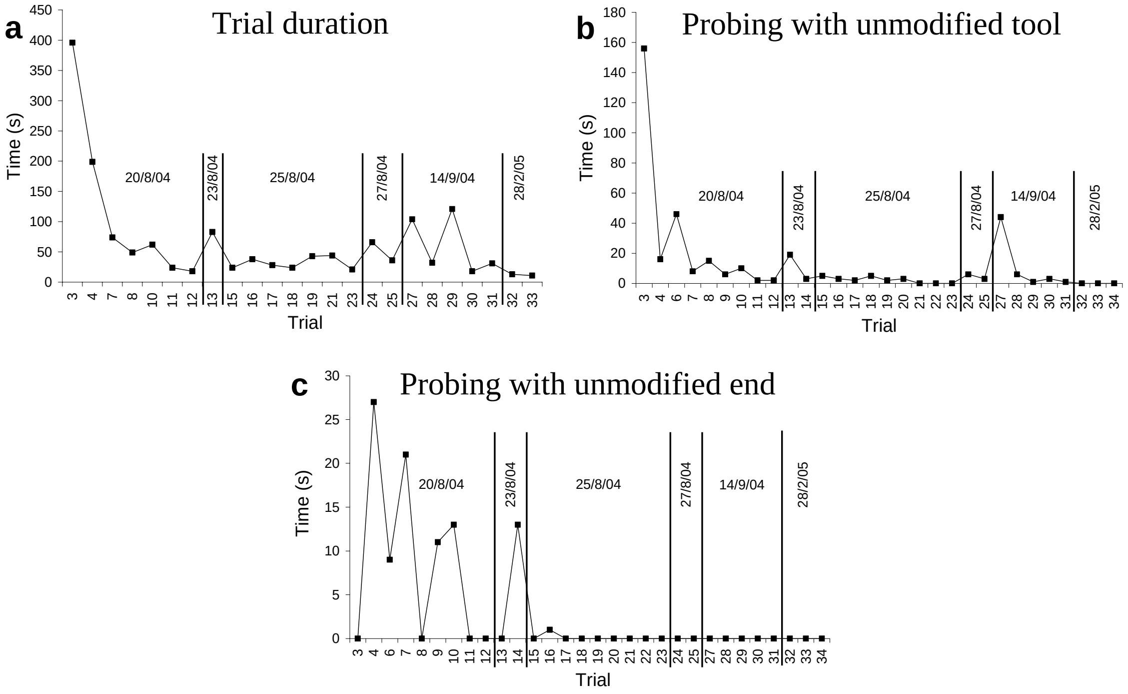


Fig. 2 **a** Duration (time from first contacting the apparatus to retrieving the bucket) of successful trials in Experiment 1. **b** Duration of probing with the unmodified tool on each trial. **c** Duration of probing with the unmodified end of the tool (after modification). In all graphs, vertical lines and annotations show the date the trials were carried out. Note that 24 trials are shown in **a**, since this includes only successful trials using a modified tool, whereas 30 trials are shown in **b** and **c**, since these include all trials where the tool was modified.

showed that trial duration fell across days and within days (day: $F_{1,20} = 6.15$, $p = 0.022$; trial: $F_{1,20} = 5.10$, $p = 0.035$). Note that both of these effects wane if the second successful trial with a modified tool (Trial 4) is also excluded (day: $F_{1,19} = 3.12$, $p = 0.093$; trial: $F_{1,19} = 3.08$, $p = 0.095$), implying that trial duration quickly reached a floor value after the first two trials.

Betty probed for the bucket before modifying the tool on most trials. The latency until she first modified the tool and the length of time she spent probing with the unmodified tool were closely correlated (Pearson correlation $r = 0.946$, $n = 30$, $p < 0.001$), so only the latter was analysed. The length of time spent probing with the unmodified tool (Fig. 2b) dropped by a factor of 10 between the first two successful trials (Trials 3 and 4), and decreased significantly across and within days, with no significant interaction (day: $F_{1,27} = 36.38$, $p < 0.001$; trial: $F_{1,27} = 25.43$, $p < 0.001$; day*trial: $F_{1,26} = 0.06$, $p = 0.803$), unaffected by excluding Trial 3 (day: $F_{1, 26} = 26.57$, $p < 0.001$; trial: $F_{1,26} = 21.71$, $p < 0.001$; day*trial: $F_{1,25} = 0.01$, $p = 0.926$). This means that at the start of each day's trials Betty tended to probe for longer with the unmodified tool than she did at the end of that day's trials, and overall the duration of such probes decreased as the experiment progressed. In total, there were only six trials where Betty probed for longer than 10 seconds with the unmodified tool, and the median duration of such probing was 3 seconds (mean = 12.2 ± 5.4 s SE). Frequently, she did not actually make contact with the handle of the bucket in these probes with the unmodified tool—it often appeared as if she was looking into the tube while holding the tool, rather than actually probing for the bucket (see detailed descriptions for each trial in Table S1).

Since it was the proximal end of the strip (i.e. the end held in the beak) that Betty modified, she had to turn around the tool in order to use the modified end. After modifying the tool, she first probed with the *unmodified* end for five of the first 10 trials, and on two more trials later in the experiment (Fig. 2c). The duration of probing with the unmodified end of the tool appeared to decrease up to the point where she started consistently turning the tool around (Fig. 2c), but the number of trials with non-zero durations is too small for statistical analysis. Across all 34 trials, she turned the tool around from holding the modified end to holding the unmodified end on 30 occasions, whereas she turned it the other way only twice (and on four occasions she turned the unmodified tool around).

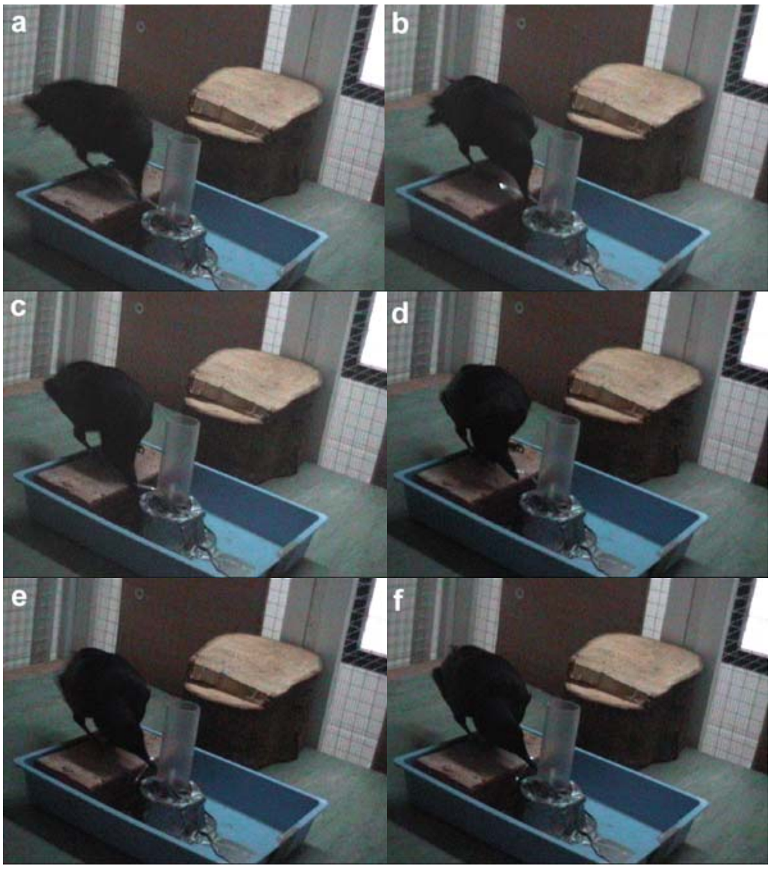
One further point of interest is the comparison between Trials 31 and 32, which were 167 days apart (see Table 1), during which time Betty had had no exposure to the material. Despite this long gap, her performance was indistinguishable: on both trials she modified the tool before probing with it, and turned it around immediately after modification.

Details of tool modification

The first time Betty modified the aluminium strips (Trial 3) is of the most interest in terms of how she reacted to this new material (see Video S5). As described in Table S1, she probed 9 times for the bucket with the unmodified strip (raising it almost all the way to the top of the tube once, and half-raising it several times), often poking the strip at the base of the tube in between probes (subjectively, it seemed as if she was trying to wedge the tool by inserting its end into the tape as she had previously done with wire, but since the metal strip has a larger cross-sectional area it did not puncture the tape). As she persisted the pokes became more vigorous, so that after 3.5 minutes the strip bent slightly in the middle, although the bending itself did not seem ‘deliberate’. She carried on probing for the bucket and poking the strip at the tape where she had in the past wedged the wire (once causing it to bend slightly more again) until 6.25 minutes into the trial, at which point she again pressed the distal end against the tape, but this time grasping the proximal end nearer the middle of the tool with her beak slightly sideways, and twisting her head so that the metal bent around her beak (Fig. 3 and Video S5). This is an action she had never performed with the wire, nor, to our knowledge, in any other context. The movement caused the strip to twist into a large loop (see tool 3 in Fig. S4), which she then picked up (by the modified end) and inserted into the tube, but almost immediately she let it drop. Thanks to the loop she could still reach it, and picked it out of the tube, dropped it onto the tray, picked it up again by the unmodified end, and used the twisted end to successfully retrieve the bucket.

Note: this is a draft version of the manuscript published in Animal Cognition *(DOI 10.1007/s10071-006-0052-5). The original publication is available at http://www.springerlink.com.*

Fig. 3 Stills from the video of Trial 3 of Experiment 1 (Video S5), showing the moment Betty first 'deliberately' bent the new tool (using the "wedge-twist" technique). In **a–b** she moves her grip on the tool further down its shaft, and then in **c–d** she twists her head around, bending the tool in the process. In **e–f** the resulting bend in the tool is visible.



In the following trial (Trial 4), she probed twice (for 16 seconds) with the unmodified tool, and then pressed the end of the tool against the tape at the base of the tube, and made a twisting head movement as in Trial 3. This caused the tool to twist a little, but perhaps because the distal end was not firmly wedged, the bend was far smaller than in the previous trial. She carried on probing with the unmodified end for 30 seconds, interspersed with another poke-twist movement, before turning the tool around and probing a further 7 times (for 1.5 minutes) with the modified end, interspersed with three apparent poke-twist episodes, none of which modified the tool substantially. She eventually succeeded in getting the bucket, but the final tool was not modified very much from the original, and the modification attempts were clumsy and did not appear to be precisely controlled.

The amount of time Betty spent modifying the tool (Fig. 4a) decreased across days but not within days, and there was no interaction between the two (day: $F_{1,27} = 17.13$, $p < 0.001$; trial: $F_{1,27} = 0.49$, $p = 0.489$; date*trial: $F_{1,26} = 2.22$, $p = 0.148$). Her modification technique changed across trials: in early trials, she used the 'twist' technique described above, but from around Trial 17 she predominantly used a 'bend' technique (e.g. Trial 32, shown in Video S6). Although hard to score formally, this transition seems to have been quite gradual: from Trial 12, she began to twist her head sideways less, and instead started pushing the tool away from her while raising the end

Note: this is a draft version of the manuscript published in Animal Cognition *(DOI 10.1007/s10071-006-0052-5). The original publication is available at http://www.springerlink.com.*

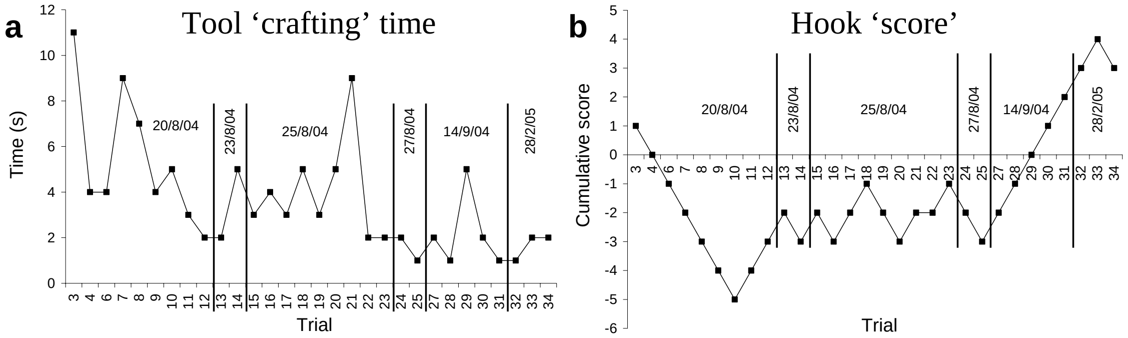


Fig. 4 **a** Length of time Betty spent modifying tools (from the first moment the tool started to bend until the last) in Experiment 1. **b** Cumulative hook 'quality score', by trial. The score increased by 1 if the hook on that trial met several criteria for good design (see Methods for details), and decreased by 1 if it did not. In all graphs, vertical lines and annotations show the date the trials were carried out. Note that since only trials where tools were modified are included, only 30 trials are shown in each graph.

of her beak, resulting in a bend rather than a twist. The effect of the different techniques is apparent in the final tool shapes (Fig. S4). Fig. 4b shows how the 'score' (see 'Scoring and analysis' above) of the resulting tool shape changes across trials. A Spearman Rank Correlation shows that hook score increased significantly across trials ($r_S = 0.508$, $p = 0.005$), although from visual analysis it seems that apart from a period from Trials 10 to 13, she only started making consistently 'well-shaped' hooks from Trial 27. However, as we pointed out before, how good a hook looks to us does not necessarily correspond to what is functionally best: Betty was able to retrieve the bucket with almost all of the tools she made, regardless of how neatly shaped they were, and there was no relationship between hook score and time spent probing (excluding Trial 4 which is 4.5 standard deviations away from the mean: $F_{1,21} = 0.35$, $p = 0.562$).

Discussion

From the third trial with the new material, Betty began to modify the aluminium strips, and was generally successful at retrieving the bucket. There was a striking drop in the length of time she spent probing with the unmodified tool (and, correspondingly, in the latency until she modified the tool) between Trials 3 and 4—namely, immediately after her first successful modification. Such sudden and large changes in performance have been attributed to 'insight'-like processes (e.g. Köhler 1925; but see Spence 1938). An alternative explanation, more parsimonious in this case, is that the subject had initially not 'discovered' that the new material was pliant, nor how to modify it. This conclusion is borne out by detailed analysis of her behaviour on the first trial involving modification: before successfully bending it, she repeatedly treated the strip as she had previously treated wire, but failed to achieve the same result as she could not wedge the strip by piercing the tape. Having discovered that the new material could be modified, she was subsequently much quicker to attempt to manipulate it (although note that on Trial 4 she modified it less effectively than on Trial 3, and consequently it took her longer to retrieve the bucket with the modified tool than on any other trial), and the length of time she spent crafting the tool decreased across trials.

Once she had learned about the properties of the new material, was her behaviour uniquely attributable to 'instant' understanding and appropriate behaviour thereafter? Two lines of evidence suggest that this is not the case. Firstly, although the time she spent probing with the unmodified tool rapidly decreased across trials, she still nearly always attempted to probe for the bucket before modifying the material. This does not necessarily imply a lack of understanding, because there might be a cost (e.g. effort or discomfort) to modifying it, and she was once (in Trial 5) successful with the unmodified tool so she might have perceived it as being 'worthwhile' probing without the hook, but it also precludes an explanation based purely on insight. Secondly, for 5 of the first 7 trials in which she modified the tool, her first probes after bending it were with the unmodified end. The duration of these probes dropped rapidly during these trials, and from Trial 11 she consistently turned the tool around before using it.

Was her behaviour, therefore, suggestive of improvement due solely to reinforcement of random successful actions? Again, the results appear to suggest otherwise. Firstly, she learned very quickly how to effectively modify the tool, even though she had to use completely different techniques from those used with wire. In fact, from the sixth trial onwards, she only once spent more than 5 seconds crafting the tool. In addition, the 'hook-ness' of her tools rapidly improved and became more regular (see Fig. 4b and Fig. S4), although even towards the end of the experiment there were the occasional malformed ones. This is despite the difficulty of modifying this kind of material with a beak as her only manipulative appendage, and the fact that the modification techniques she used are unlike any known actions used by wild crows, or by her in other circumstances. As argued earlier, such rapid acquisition would be highly unlikely to occur in an agent reliant solely on reinforcement learning, and if anything Betty's previous experience with wire should have retarded the speed with which she learned about this new material, due to interference (e.g. Wilson et al. 1985). Betty retained perfect performance after more than 5 months with no exposure to the material, but this, while impressive, is not helpful to gauge her level of cognitive processing; she could remember either the appropriate associations or the appropriate insights.

How can we then set bounds as to what she 'understands' about the task? It seems likely that she understands *aspects* of the task, but combines this with trial and error guided by reinforcement. She appears to understand the relationship between her actions and the resulting tool shape, since she was able to develop novel modification techniques very rapidly. It also seems that she understood or had previously learned the need for hook-like shapes to retrieve the bucket, since she fairly consistently produced suitable shapes from the fourth trial (of those where she modified the tool). It is not clear how to work out a 'null hypothesis' for the likelihood of producing hook-like shapes versus all other shapes from random manipulation of the material, but just from the diversity of shapes Betty produced it is clear that there are several possibilities, and there are obviously many possible shapes that she never made (some of which would not even have fitted into the tube), yet she produced far more of the usable hook-like tools than of any other. However, it seems that it would be premature to attribute understanding of *why* a hook was needed, because if she understood this well she would never have probed with the wrong end of the tool after modifying it—which in the first few trials she sometimes did for over 10 seconds. She quickly learned to turn the tool around after modifying it (which, incidentally, she only did 4 times with unmodified tools, suggesting that she recognised it was only worthwhile turning around modified tools), but an agent who truly understood *why* they needed a hook should never probe with the wrong end of the tool. It should be noted here, though, that chimpanzees tested by Povinelli and colleagues only correctly re-oriented a hooked tool on 6 of 28 trials (4 per subject), which was the same as the frequency with which they reoriented a straight tool (Povinelli et al. 2000a, Experiment 16 conditions E and G).

Experiment 2: Unbending for tool shape

In this experiment we used a task in which repeating the successful actions from the previous problem would lead to failure. We presented a strip of aluminium as in Experiment 1, but now the strip was bent at both ends and the task required inserting a tool through a narrow hole. The experiment began after Trial 31 of Experiment 1 (see Table 1), so the subject was now familiar with the aluminium strips and how to manipulate them. This experiment is similar to Experiments 24–26 carried out with chimpanzees by Povinelli and colleagues (2000b).

Methods

Apparatus

The apparatus was one Betty was already familiar with from experiments by Stephen Barlow (unpublished). It was constructed from 5 cm diameter Rotastack® components made for pet rodent housing (see Fig. 5a). The tubing formed a ‘cross’ shape; the upper arm and one of the horizontal arms of the cross were blocked by solid (red) end-caps; the other horizontal arm had a (red) end-cap with a 7 mm diameter hole drilled into it; for Trial 3, the vertical arm had an open semi-transparent section of tubing attached to it (as shown in Fig. 5a). The reward (a small piece of pig heart or a waxworm, as in Experiment 1) was placed in a small plastic cup, inside the horizontal arm of the apparatus, behind the perforated end cap. The task was similar to that described in Chappell & Kacelnik (2004): to retrieve food, a tool had to be inserted and the food cup had to be pushed along the tube to make it fall out of the vertical pipe.

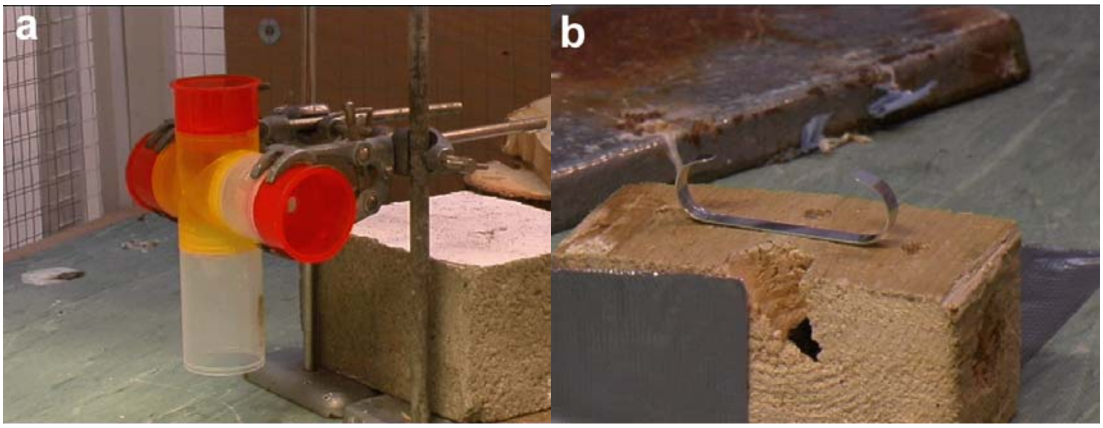


Fig. 5 Equipment setup for Experiment 2. **a** The apparatus for the experiment; the food cup (not visible) is in the horizontal arm facing towards the camera, just behind the red end-cap with the hole in. A tool can be inserted through the hole in the end-cap, to push the cup along the horizontal tube and make it fall out of the vertical tube. **b** The unmodified tool provided for Experiment 2. The tool is a similar strip of aluminium to that used in Experiment 1, except that the both ends have been bent into small hooks. Without modification, the tool does not fit through the hole in the end-cap of the apparatus.

In each trial a strip of aluminium (90 mm long × 5 mm wide × 1 mm deep), similar to those described in Experiment 1, was provided as a potential tool. Both ends of the strip were bent into small hooks using a metal rod of 15 mm diameter as a template; in its modified form, the tool was 60 mm long (see Fig. 5b), and the hooks on each end prevented it from fitting through the hole in the end-cap (above). The tool was placed on top of the wooden block described in Experiment 1, which was fixed to the table about 30 cm from the front of the apparatus.

Procedure

Before testing began, the subject was presented with several trials with the apparatus and either straight, rigid tools or straight aluminium strips. On every one of these familiarisation trials, Betty immediately picked up the tool and poked it through the hole, retrieving the food within seconds of the trial starting, and without modifying the tool in any way.

Food deprivation, rewards, participation of the subject, data recording, trial duration, and all such details were as described for Experiment 1. All scoring was done from videotapes. Each trial was summarised descriptively, and only informal analysis was carried out, since only three trials were performed, for reasons that will be apparent in the next section.

Results

Betty successfully retrieved the food on all three trials (described in detail in Table S2). On the first trial Betty got the food without using a tool by pecking hard at the end cap, which caused the food cup to fall out of the vertical arm. In the second and third trials, she modified the strip by squeezing together one (Trial 2) or both (Trial 3) ends. On

Note: this is a draft version of the manuscript published in Animal Cognition *(DOI 10.1007/s10071-006-0052-5). The original publication is available at http://www.springerlink.com.*

Trial 2, she used the unmodified (so still hooked) end to retrieve the food by inserting the strip into the vertical arm from underneath the apparatus, and hooking the food cup out. Consequently, in Trial 3 a semi-transparent vertical tube was added to the vertical arm to make this impossible (see Fig. 5a). In this trial, she squeezed together one end and picked up the tool holding this end. She spent 4 seconds attempting to insert the unmodified (hooked) end of the strip through the hole, and then turned the strip around (squeezing together the other end in the process) and poked the first flattened end through the hole, dislodging and thereby retrieving the food.

Discussion

Although Betty quickly got the food in all three trials, trial 3 was the only one where she performed the task as intended by us. On this trial she obtained the food by squeezing together both ends of the strip and inserting one end through the hole in the end-cap, thereby solving the problem of “spontaneously modifying the tool to allow it to fit through the hole”. However, this modification may have been aimed at making it easier to pick up the strip, as demonstrated by the fact that she also squeezed together one end in Trial 2 where she used this squeezed end to hold the tool and solved the task using the hook through a route we had not anticipated. Moreover, on Trial 3 she initially tried to insert the *unmodified* end of the tool into the hole, although she very quickly turned the tool around and used the modified end correctly.

These three trials do not, therefore, provide evidence that she specifically modified the tool with a view to its future functionality, but they are compatible with this possibility. While probing with the ‘wrong’ end of the tool indicates lack of comprehension, the duration of this action was so quickly replaced by a functional action that it could be interpreted as Betty instantly ‘understanding’ what was required. Turning the tool around was the appropriate but not the only possible response; she did not rotate it around the axis of the tool shaft, as she has done in other tasks (when it can be appropriate), nor did she persist in trying to use the incorrect end.

It is also informative to compare Betty’s performance with that of chimpanzees tested on a similar tool insertion problem, given tools that had a straight end (that could be inserted into the apparatus) and an end that could not be inserted (of two different designs). All seven subjects showed a strong preference for attempting to insert the “impossible” ends of the tools, and very rarely turned the tools around (Povinelli et al. 2000c, Experiment 12): out of 56 trials (8 per subject, 7 subjects), subjects only succeeded in getting the food 3 times (despite many successful trials using straight tools previously). As mentioned in the Introduction, the chimpanzees had a similar preference for probing with the impossible end of the tool in experiments 25–26 (Povinelli et al. 2000b), although in these experiments they do appear to have turned the tools around more frequently (the exact number of reorientations is not reported, but 69 / 64%

[Experiment 25 / 26 respectively] of first attempts were with the impossible end, yet the chimpanzees managed to retrieve food eventually on 61 / 80% of trials). In this context, Betty's response of turning the tool around almost instantly is impressive, even if not equivalent to a human-like understanding.

The experiment was discontinued because of the impossibility of separating the flattening of the tool for ergonomic reasons from flattening aimed at facilitating the insertion of the tool. Our next task was designed to overcome this difficulty.

Experiment 3: Unbending for tool length

Although in Experiment 2 our subject succeeded in modifying the original material as required to make it functional (making it thinner so as to pass through a narrow hole), the cognitive implications are hard to elucidate because she may have achieved this while attempting to make the tool easier to hold, with the tool narrowing occurring as a by-product. To avoid this difficulty, we now supplied an instrument that was more difficult to squeeze together, and changed the functional need, with the required modification being to make the tool longer, rather than narrower. The experiment began after Trial 3 of Experiment 2 (see Table 1).

Methods

Apparatus

The apparatus was a horizontal tube made from clear Perspex (30 cm long, 4 cm in diameter), mounted in a wooden stand with the centre of the pipe 12 cm high above the table (identical to that used in Chappell and Kacelnik 2002). A piece of pig heart was placed inside the tube (10 cm from the tube entrance for Trial 1, 13 cm for Trials 2–4).

A strip of aluminium (90 mm long × 5.0 mm wide × 1.0 mm deep), bent into a broad U-shape (as shown in Fig. 6 column 1), was provided as a potential tool. In Trial 1, the ends of the U were 2.5 cm apart and the two arms were almost parallel to each other (the angle between them being 5°), and the tool was 4 cm long from the ends to the apex of the U-bend. The U-bend was made broader for Trials 2–4, bringing the ends 5.5 cm apart (and the angle between the arms to 62°), and the tool was 3.4 cm long from the ends to the apex of the bend. The tool was placed on the wooden block described for Experiment 1, which was fixed to the table about 30 cm from the opening of the tube.

Procedure

No training was given, since the subject was already familiar with the apparatus and material (as used in Experiments 1 and 2). Food deprivation, rewards, subject participation, data recording, and trial definitions were as described for Experiment 1. Four trials were carried out, as described in Table 1. To ensure that the subject was still familiar with the properties of the tool on Trials 3 and 4 (which occurred five months after the first two), she was given three trials with the straight aluminium strip and the well/bucket apparatus immediately before Trial 3 (see Table 1).

Results

The subject successfully retrieved the food on 3 of the 4 trials (described in detail in Table S3). On the first trial, she managed to squeeze together the ends of the tool to create a flattened, straight tool 4.5 cm long (Fig. 6 tool 1). Although the meat was 10 cm inside the tube, she just managed to reach and retrieve it by inserting her head and neck into the entrance of the tube. For this reason, on subsequent trials we made the U-shape broader and positioned the meat further inside the tube.

On Trial 2, Betty tried for 1.5 minutes to get the reward by probing inside the tube with the unmodified U-shaped tool, but did not succeed and never showed any ‘deliberate’ attempt to modify the tool (although, presumably as a result of repeated probing attempts inside the tube, at the end of the trial the tool was flatter than at the beginning: the ends were 7.5 cm apart, with an angle between them of 75°; see Fig. 6 tool 2). On Trials 3 and 4, however, Betty did modify the tool and retrieve the reward. Both trials involved a similar modification technique, which occurred several minutes into the trial: in the middle of a bout of probing in the tube, she raised her head and beak (still holding one end of the tool) in a distinctive and unusual manner, causing the shaft of the tool to bend backwards against the lip of the tube (see Videos S7 and S8). On Trial 3 this resulted in a bend backwards of ~40° (Fig. 6 tool 3) and a tool 8.5 cm long, while on Trial 4 it was ~25° (Fig. 6 tool 4) and 8.0 cm long. While there is no absolute proof that the behaviour was ‘deliberate’ and planned, this action had never been seen before with this or other materials. On both trials she was able to get the meat with the modified tools.

Unfortunately, Betty died before we were able to complete any more trials.

Discussion

On three of the four trials, the subject modified the tool and retrieved the food. On the first trial, the modification could have been related to picking up and holding the tool, rather than elongating it, but the other two modifications appeared to be specifically problem-related: they made the tool longer, as required to solve the task, and seemed ‘deliberate’, because the action was not one she had performed before. Moreover, in no trial did she modify the strip by bending or twisting it, which were the actions associated with success in Experiment 1, suggesting that she recognised that those behaviours would have been inappropriate. However, she probed for some time with the unmodified tools, and this did not decrease along these few trials. In fact, she probed for longer with the unmodified tool in Trial 4 than in Trial 3, although she had already experienced the result of the modification. As in other tests, her behaviour was neither

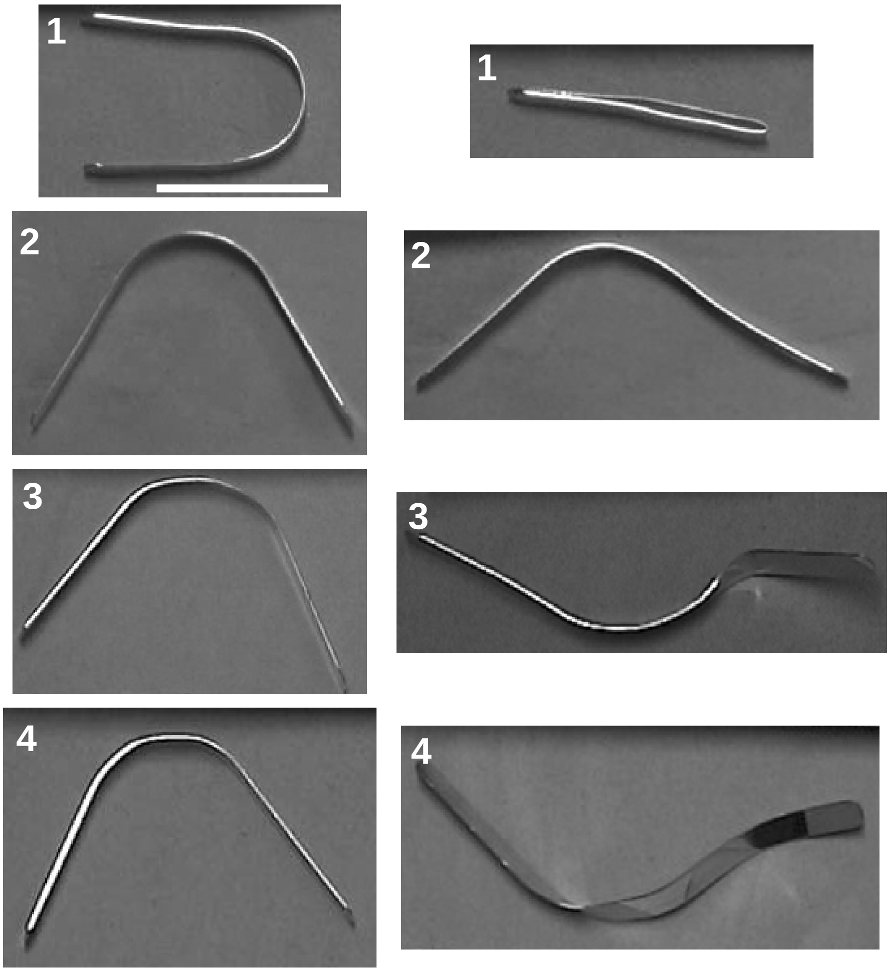


Fig. 6 Before (*left column*) and after (*right column*) tools from Experiment 3. The number in the top corner shows the trial, and the scale bar is 3 cm.

consistent with a complete human-like understanding of the problem nor with random manipulation of the raw materials.

It is again interesting to compare Betty's performance with that of the chimpanzees studied by Povinelli and colleagues (2000b, Experiments 24–26) (see 'Introduction' for a description of the experiments and overall results; the unbending experiments by Klüver (1937) and Anderson & Henneman (1994) are not reported in enough detail for comparison). In the first 56 test trials (Experiments 24 and 25: 7 subjects, 8 trials per subject) where the chimpanzees had to modify the tool to succeed, there were 7 instances where the tool was modified, but either the modification did not straighten the tool sufficiently, or the tool was not used appropriately after modification; no subject was successful in retrieving the reward. After explicit training in bending (note: not unbending) the tool (Experiment 26), one subject (Jadine) did appropriately modify and use the tool (although she still first directed the unmodified end at the apparatus on one trial). There are many differences between the situation for the chimpanzees and for Betty in the present experiment that limit the possibility of direct comparisons (including differences in the task—unbending for length rather than width), but it is interesting to note that Betty's modifications occurred after far fewer trials than Jadine's, and without any explicit training by the experimenter. However, Betty had had experience bending the material in other tasks, which might have given her an advantage.

General discussion

Our experiments aimed to address the question of how much New Caledonian crows understand about the nature of physical interactions between their tools and their target objects, and in particular whether they can create novel tools by modifying materials appropriately without shaping by direct reinforcement. Our target was to elucidate whether they can use trial-unique judgement to plan specific modifications to tools. Tool crafting and using in this species is the result of a complex interrelation between inherited and learned behaviour programs (Kenward et al. 2005; Kenward et al. in press), and aspects of tool manufacture may be culturally transmitted (Hunt and Gray 2003); however, this is of course also true for human tool making, and does not shed light on the level of cognitive processing at which the problems are dealt with. A human may learn by reinforcement to make a specific movement or to 'think about the problem' in order to solve novel problems and our goal is to know where along such a continuum the behaviour of this species should be placed.

We built on the observations of spontaneous wire-bending by one individual: Betty (Weir et al. 2002). In the previous study she appeared to create a tool of a shape she had seen before (a hook) by employing novel movements addressed appropriately to an unusual material, justifying the working hypotheses that she understood the functionality required of her tool and also that she could plan her movements in order to achieve a specific final shape of tool. Here, we were interested in how Betty would adapt to the introduction of a new material with different mechanical properties, and whether she would modify it in different and specific ways when faced with tasks that required different tools. In three experiments, Betty had a high level of overall success (she only failed to get the food on 7 of 41 trials), adapted very quickly to the new material, and was able to modify the tools in different ways depending on the task requirements.[1]

However, examination of the details of her performance showed that while her innovative behaviour cannot be accounted for purely by reinforcement of specific actions, it is not yet justified to assume that she possesses a full, human-like understanding of each task and that she uses it to plan and direct her behaviour (although whether the full understanding that humans presumably have of the task would reveal itself by perfect first-trial performance is unclear, since humans often

---

[1] It is also interesting to note that Betty managed to "outwit" the experimenters on several occasions, obtaining the food using techniques that had not been anticipated. While these observations do not necessarily shed light on the questions this study set out to address, they do illustrate the flexible, innovative nature of this individual's approach to solving problems.

make mistakes despite such understanding). There are three general points we would like to make from these results.

Firstly, the fact that a subject does not behave in the way we presume someone using the logic of an adult human would does not necessarily mean that the subject does not understand the task. This is because our presumptions about the behaviour of agents that do understand the problems can be mistaken. For instance, it is tempting to assume that if Betty were aware of the physical principles involved in the functioning of hooks she would never probe with the wrong end of a tool she has modified into that shape. However recent experiments have (re)emphasized the fallibility of intuition and introspection for making such predictions. Silva and colleagues (Silva et al. 2005) presented adult humans with both a physical and a diagrammatic ‘trap-tube’ task (Visalberghi and Limongelli 1994), which has been used to assess means-end understanding in several primate and avian species. In this task, subjects are presented with a horizontal transparent tube containing a reward, with a ‘trap’ in the middle: if the food is pushed (or pulled) over the trap, it falls into it and the subject cannot retrieve it. A critical test for whether subjects learn about the causal properties of the task is how they respond when the tube is inverted, so that the trap is now facing upwards and therefore food is not lost if passed just under it. The argument has been made that if subjects understood gravity, they would not avoid the trap as they do when the trap faces downwards. Since most non-humans continue to avoid the inverted trap, they are often assumed to lack this understanding (e.g. Visalberghi and Limongelli 1994; Reaux and Povinelli 2000; but see Tebbich and Bshary 2004). However, in Silva et al.’s (2005) experiments with humans that certainly do understand the role of gravity, the subjects continued to avoid a trap after it had been inverted on over 90% of trials, while reporting that they understood that it was no longer effective. As Silva and colleagues point out, it is critical to test how humans perform on tasks that they do understand before interpreting a non-human animal’s failure as evidence for lack of understanding.

Secondly, some progress can be made by comparing behaviour of members of different species in comparable tasks. This is not easy because cross-species comparisons are beset by technical difficulties and because there are relatively few such experiments. In our case, the closest comparisons can be made with experiments conducted by Povinelli’s group (Povinelli 2000). Compared to their chimpanzees, Betty seems to have learned more quickly and been generally more successful. We cannot rule out non-cognitive explanations for this disparity (such as differences in motivation, or specific inherited predispositions for the tasks involved), but taken at face value, this bird seems to outperform our closest relatives, who are often considered to be the most intelligent non-humans. The observed advantage may, though, be restricted to a narrow class of tasks: tool use develops spontaneously in New Caledonian crows reared in isolation (Kenward et al. 2005), appears to be genetically well-canalised (Kenward et al.

in press), and is very widespread in the wild (Hunt and Gray 2002, 2003). These birds might, therefore, have specific cognitive adaptations that make them particularly good at learning and possibly reasoning about tasks involving physical interactions between solid objects, but perhaps not extraordinary at other equally difficult logical tasks. In contrast, tool use in chimpanzees may be a product of more generalised learning and reasoning processes, since it seems to be strongly culturally influenced (e.g. Whiten et al. 1999, 2001; Whiten 2005; Whiten et al. 2005), and takes a long time for individuals to learn (e.g. Biro et al. 2003; Hirata and Celli 2003; Lonsdorf et al. 2004; Lonsdorf 2005, 2006).

Finally, questions about understanding are frequently posed as all-or-none options (either the subject fully understands the causal nature of the task—the "high-level model", in Povinelli's terminology—or is simply following procedural rules, with no causal understanding at all), but we believe that this probably represents a false dichotomy. It seems clear from the results presented in this paper that 'intermediate' levels of understanding are possible (see also Watanabe and Huber 2006)—perhaps reflecting the degree to which the individual can apply concepts learned in one situation to other, causally-similar but perceptually-different, situations. We suspect that progress might come when we can replace terms such as understanding (which we feel compelled to maintain for the time being) by precise hypotheses about the operations the subjects make in the course of generating solutions to novel problems. This can be achieved both by further work using similar paradigms in different species (importantly, including humans), and by incorporating the tools of artificial intelligence. Instead of asking whether our subject understands a given physical principle, it might be helpful to determine the minimum specifications that need to be incorporated into a robot to achieve a similar level of generalisation and creativity (although we acknowledge that similarities between the way animals and a model perform do not demonstrate that the same mechanisms are involved in each). If, as we suspect, it is impossible for a robot equipped exclusively with associative learning algorithms to solve these tasks with a similar amount of experience, then it would be justified to speculate about the higher cognitive functions that may need to be invoked.

The goal of the present paper was to further explore the cognitive processes underlying the abilities of Betty, potentially an exceptional individual. The use of a single subject to date means that we do not know if Betty's abilities are representative of New Caledonian crows. Experiments are presently being run on other individuals to address this question.

Acknowledgements

We are grateful to Jackie Chappell, Ben Kenward, Christian Rutz, Lucas Bluff, Jo Wimpenny, and Stephen Barlow for many discussions about experimental design and interpretation. David Wilson looked after Betty through her life in Oxford with affection and efficiency. Thanks to Charlotte Burn and three anonymous referees for helpful comments on the manuscript. David Palmer and Tony Price helped design and build the experimental room and its entrance-control mechanism. The experiments described here comply with all applicable laws of the United Kingdom, where they were carried out. This work was funded by a scholarship from the Wellcome Trust to AASW and BBSRC grant BB/C517392/1 to AK.

References

Anderson JR (2001) [Review of *Folk physics for apes: The chimpanzee's theory of how the world works*, by Daniel Povinelli]. Anim Behav 61: 1042–1043 DOI 10.1006/anbe.2001.1694

Anderson JR, Henneman MC (1994) Solutions to a tool-use problem in a pair of *Cebus apella*. Mammalia 58: 351–361

Biro D, Inoue-Nakamura N, Tonooka R, Yamakoshi G, Sousa C, Matsuzawa T (2003) Cultural innovation and transmission of tool use in wild chimpanzees: evidence from field experiments. Anim Cogn 6: 213–223 DOI 10.1007/s10071-003-0183-x

Butler AB, Manger PR, Lindahl BIB, Arhem P (2005) Evolution of the neural basis of consciousness: a bird-mammal comparison. BioEssays 27: 923–936 DOI 10.1002/bies.20280

Chappell J, Kacelnik A (2002) Tool selectivity in a non-mammal, the New Caledonian crow (*Corvus moneduloides*). Anim Cogn 5: 71–78 DOI 10.1007/s10071-002-0130-2

Chappell J, Kacelnik A (2004) Selection of tool diameter by New Caledonian crows *Corvus moneduloides*. Anim Cogn 7: 121–127 DOI 10.1007/s10071-003-0202-y

de Kloet ER, Oitzl MS, Joels M (1999) Stress and cognition: are corticosteroids good or bad guys? Trends Neurosci 22: 422–426

Defeyter MA, German TP (2003) Acquiring an understanding of design: evidence from children's insight problem solving. Cognition 89: 133–155 DOI 10.1016/S0010-0277(03)00098-2

Emery NJ, Clayton NS (2004) The mentality of crows: convergent evolution of intelligence in corvids and apes. Science 306: 1903–1907 DOI 10.1126/science.1098410

Emery NJ, Dally JM, Clayton NS (2004) Western scrub-jays (*Aphelocoma californica*) use cognitive strategies to protect their caches from thieving conspecifics. Anim Cogn 7: 37–44 DOI 10.1007/s10071-003-0178-7

Epstein R, Kirshnit CE, Lanza RP, Rubin LC (1984) Insight in the pigeon: antecedents and determinants of an intelligent performance. Nature 308: 61–62

Goodall J (1986) The chimpanzees of Gombe: patterns of behavior. Harvard University Press, Cambridge, MA.

Griffin DR, Speck GB (2004) New evidence of animal consciousness. Anim Cogn 7: 5–18 DOI 10.1007/s10071-003-0203-x

Hauser MD (2001) Elementary, My Dear Chimpanzee. Science 291: 440–441 DOI 10.1126/science.1057937

Higuchi H (1986) Bait-fishing by the green-backed heron *Ardeola striata* in Japan. Ibis 128: 285–290

Higuchi H (1988) Individual differences in bait fishing by the green-backed heron *Ardeola striata* associated with territory quality. Ibis 130: 39–44

Hirata S, Celli ML (2003) Role of mothers in the acquisition of tool-use behaviours by captive infant chimpanzees. Anim Cogn 6: 235–244 DOI 10.1007/s10071-003-0187-6

Hunt GR (1996) Manufacture and use of hook-tools by New Caledonian crows. Nature 379: 249–251 DOI 10.1038/379249a0

Hunt GR (2000) Human-like, population-level specialization in the manufacture of pandanus tools by New Caledonian crows *Corvus moneduloides*. Proc R Soc Lond B 267: 403–413 DOI 10.1098/rspb.2000.1015

Hunt GR, Gray RD (2002) Species-wide manufacture of stick-type tools by New Caledonian Crows. Emu 102: 349–353 DOI 10.1071/MU01056

Hunt GR, Gray RD (2003) Diversification and cumulative evolution in New Caledonian crow tool manufacture. Proc R Soc Lond B 270: 867–874 DOI 10.1098/rspb.2002.2299

Hunt GR, Gray RD (2004a) The crafting of hook tools by wild New Caledonian crows. Proc R Soc Lond B (Suppl) 271: S88–S90 DOI 10.1098/rsbl.2003.0085

Hunt GR, Gray RD (2004b) Direct observations of pandanus-tool manufacture and use by a New Caledonian crow (*Corvus moneduloides*). Anim Cogn 7: 114–120 DOI 10.1007/s10071-003-0200-0

Hunt GR, Rutledge RB, Gray RD (2006) The right tool for the job: What strategy do wild New Caledonian crows use? Anim Cogn, DOI 10.1007/s10071-006-0047-2

Jarvis ED, Güntürkün O, Bruce L, Csillag A, Karten H, Kuenzel W, Medina L, Paxinos G, Perkel DJ, Shimizu T, Striedter G, Wild. JM, Ball GF, Dugas-Ford J, Durand SE, Hough GE, Husband S, Kubikova L, Lee DW, Mello CV, Powers A, Siang C, Smulders TV, Wada K, White SA, Yamamoto K, Yu J, Reiner A, Butler AB (2005) Avian brains and a new understanding of vertebrate brain evolution. Nat Rev Neurosci 6: 151–159 DOI 10.1038/nrn1606

Kacelnik A, Chappell J, Weir AAS, Kenward B (2006) Cognitive adaptations for tool-related behaviour in New Caledonian Crows. In: Wasserman EA, Zentall TR eds) Comparative cognition: experimental explorations of animal intelligence. Oxford University Press, Oxford, pp 515–528

Kenward B, Weir AAS, Rutz C, Kacelnik A (2005) Tool manufacture by naive juvenile crows. Nature 433: 121 DOI 10.1038/433121a

Kenward B, Rutz C, Weir AAS, Kacelnik A (in press) Development of tool use in New Caledonian crows (*Corvus moneduloides*): stereotyped action patterns and social influence. Anim Behav

Klüver H (1937) Re-examination of implement-using behavior in a *Cebus* monkey after an interval of three years. Acta Psychologica 2: 347–397

Köhler W (1925) The mentality of apes. Harcourt Brace, London.

Lockman JJ (2000) A perception-action perspective on tool use development. Child Development 71: 137–144 DOI 10.1111/1467-8624.00127

Lonsdorf EV (2005) Sex differences in the development of termite-fishing skills in the wild chimpanzees, *Pan troglodytes schweinfurthii*, of Gombe National Park, Tanzania. Anim Behav 70: 673–683 DOI 10.1016/j.anbehav.2004.12.014

Lonsdorf EV (2006) What is the role of mothers in the acquisition of termite-fishing behaviors in wild chimpanzees (*Pan troglodytes schweinfurthii*)? Anim Cogn 9: 36–46 DOI 10.1007/s10071-005-0002-7

Lonsdorf EV, Eberly LE, Pusey AE (2004) Sex differences in learning in chimpanzees. Nature 428: 715–716 DOI 10.1038/428715a

Lucas JR (1982) The biophysics of pit construction by antlion larvae (*Myrmeleon*, Neuroptera). Anim Behav 30: 651–664

Machado A, Silva FJ (2003) You can lead an ape to a tool, but …: A review of Povinelli's *Folk physics for apes: The chimpanzee's theory of how the world works*. J Exp Anal Behav 79: 267–286

Mackintosh NJ (1994) Animal learning and cognition. Academic Press, San Diego ; London.

McGrew WC (1992) Chimpanzee material culture: Implications for human evolution. Cambridge University Press, New York, NY.

Povinelli DJ (ed) (2000) Folk physics for apes. Oxford University Press, Oxford, UK.

Povinelli DJ, Reaux JE, Theall LA, Giambrone S (2000a) The rope, hook, touching-stick, and related problems: the question of physical connection. In: Povinelli DJ (ed) Folk physics for apes. Oxford University Press, Oxford, UK, pp 206–253

Povinelli DJ, Reaux JE, Theall LA, Giambrone S (2000b) The question of tool modification. In: Povinelli DJ (ed) Folk physics for apes. Oxford University Press, Oxford, UK, pp 271–296

Povinelli DJ, Theall LA, Reaux JE, Giambrone S (2000c) The tool-insertion problem: the question of shape. In: Povinelli DJ (ed) Folk physics for apes. Oxford University Press, Oxford, UK, pp 173–205

Reaux JE, Povinelli DJ (2000) The trap-tube problem. In: Povinelli DJ (ed) Folk physics for apes. Oxford University Press, Oxford, UK, pp 108–131

Ricklefs RE (2004) The cognitive face of avian life histories: The 2003 Margaret Morse Nice Lecture. Wilson Bull 116: 119–133

Sanz C, Morgan D, Gulick S (2004) New insights into chimpanzees, tools, and termites from the Congo basin. Am Nat 164: 567–581

Schwartz B, Wasserman EA, Robbins SJ (2002) Psychology of learning and behavior. 5th edn. W.W. Norton, New York, NY ; London.

Silva FJ, Page DM, Silva KM (2005) Methodological-conceptual problems in the study of chimpanzees' folk physics: How studies with adult humans can help. Learn Behav 33: 47–58

Spence KW (1938) Gradual versus sudden solution of discrimination problems by chimpanzees. J Comp Psychol 25: 213–224

Sterelny K (2003) Charting control-space: Comments on Susan Hurley's 'Animal action in the space of reasons'. Mind Lang 18: 257–265

Sugiyama Y (1985) The brush-stick of chimpanzees found in south-west Cameroon and their cultural characteristics. Primates 26: 361–374

Sugiyama Y (1997) Social tradition and the use of tool-composites by wild chimpanzees. Evol Anthropol 6: 23–27

Tebbich S, Bshary R (2004) Cognitive abilities related to tool use in the woodpecker finch, *Cactospiza pallida*. Anim Behav 67: 689–697 DOI 10.1016/j.anbehav.2003.08.003

Thelen E (1979) Rhythmical stereotypies in normal human infants. Anim Behav 27: 699–715

Thelen E (1981) Rhythmical behavior in infancy: An ethological perspective. Dev Psychol 17: 237–257

Tonooka R (2001) Leaf-folding behavior for drinking water by wild chimpanzees (*Pan troglodytes verus*) at Bossou, Guinea. Anim Cogn 4: 325–334 DOI 10.1007/s100710100110

Visalberghi E, Trinca L (1989) Tool use in capuchin monkeys: distinguishing between performing and understanding. Primates 30: 511–521

Visalberghi E, Limongelli L (1994) Lack of comprehension of cause-effect relations in tool-using capuchin monkeys (*Cebus apella*). J Comp Psychol 108: 15–22

Visalberghi E, Fragaszy DM, Savage-Rumbaugh ES (1995) Performance in a tool-using task by common chimpanzees (*Pan troglodytes*), bonobos (*Pan paniscus*), an orangutan (*Pongo pygmaeus*), and capuchin monkeys (*Cebus apella*). J Comp Psychol 109: 52–60

Watanabe S, Huber L (2006) Animal logics: decisions in the absence of human language. Animal Cognition, DOI 10.1007/s10071-006-0043-6

Weir AAS, Chappell J, Kacelnik A (2002) Shaping of hooks in New Caledonian crows. Science 297: 981 DOI 10.1126/science.1073433

Weir AAS, Kenward B, Chappell J, Kacelnik A (2004) Lateralization of tool use in New Caledonian crows (*Corvus moneduloides*). Proc R Soc Lond B (Suppl) 271: S344–S346 DOI 10.1098/rsbl.2004.0183

Whiten A (2001) Tool tests challenge chimpanzees. Nature 409: 133 DOI 10.1038/35051638

Whiten A (2005) The second inheritance system of chimpanzees and humans. Nature 437: 52–55 DOI 10.1038/nature04023

Whiten A, Horner V, de Waal FBM (2005) Conformity to cultural norms of tool use in chimpanzees. Nature 437: 737–740 DOI 10.1038/nature04047

Whiten A, Goodall J, McGrew WC, Nishida T, Reynolds V, Sugiyama Y, Tutin CEG, Wrangham RW, Boesch C (1999) Cultures in chimpanzees. Nature 399: 682–685 DOI 10.1038/21415

Whiten A, Goodall J, McGrew WC, Nishida T, Reynolds V, Sugiyama Y, Tutin CEG, Wrangham RW, Boesch C (2001) Charting cultural variation in chimpanzees. Behaviour 138: 1481–1516 DOI 10.1163/156853901317367717

Wilson B, Mackintosh NJ, Boakes RA (1985) Transfer of relational rules in matching and oddity learning by pigeons and corvids. Q J Exp Psychol B 37: 313–332

Zorina ZA (2005) Animal intelligence: Laboratory experiments and observations in nature. Zool Zhurnal 84: 134–148