# Choice and goal-directed behavior in preschool children

Ulrike M. H. Klossek · Shan Yu · Anthony Dickinson

Published online: 21 April 2011

**Abstract** Preschool children (3–4 years old) were trained to perform two actions to gain different outcomes, in the form of video clips from different cartoons, before one of these outcomes was devalued by noncontingent exposure. The effect of outcome devaluation was subsequently assessed in an extinction test by giving children the opportunity to perform both actions in the absence of any outcomes. When the two actions were trained concurrently, performance during the test was modulated by outcome value and children showed a preference for the action trained with the currently valued outcome. By contrast, when each action was trained separately on different trials, test performance was insensitive to outcome devaluation. These effects of the training schedules are interpreted in terms of dual-process theories of action control.

**Keywords** Instrumental performance · Schedule of reinforcement · Human learning · Goal-directed behavior · Habitual responding

Evidence from studies in rodents and adult humans has suggested that at least two distinct learning processes mediate instrumental behavior: response–outcome (R–O) and stimulus–response (S–R) learning (Dickinson, 1994; Dickinson & Balleine, 1993; Tricomi, Balleine, & O'Doherty, 2009; Valentin, Dickinson, & O'Doherty, 2007). S–R learning represents the formation of an association between representations of the instrumental response and antecedent stimuli under the influence of a reinforcing outcome, whereas R–O learning involves encoding the instrumental contingency between the response and the outcome. As a consequence, the selection and initiation of subsequent instrumental behavior depends, in the former case, on antecedent stimuli rather than expected consequences. By contrast, the associative representations formed during learning in the latter case enable the learned behavior to be regulated by its consequences; specifically, by representations of the R–O contingency and the current incentive value of the outcome.

In the domain of animal learning, the outcome revaluation procedure has been commonly employed as a behavioral assay to determine whether selection and initiation of an acquired response is mediated by a representation of the outcome. Adams and Dickinson (1981) trained different groups of rats to press a lever that produced food pellets as an outcome. In addition, a different food outcome was delivered independently of the action. Following training, one of these outcomes was devalued by aversion conditioning outside the instrumental training context. When the contingent rather than the noncontingent outcome had been devalued, the rats performed significantly fewer leverpresses in a subsequent extinction test.

More recently, outcome devaluation procedures have also been used in studies on human learning. In adults, Valentin et al. (2007) showed that after being allowed to consume a soft drink to satiation, their participants showed a significant reduction in pressing a response button that had caused intraoral delivery of this drink during training, while they continued to perform buttonpresses previously reinforced by an alternative soft drink. Using animated visual stimuli as outcomes, Klossek, Russell, and Dickinson (2008) trained preschool children to manipulate icons on a touch-sensitive display in order to obtain video clips from different children's cartoons. Following training, one of the video outcomes was devalued through repeated exposure. In a subsequent choice test, 2- to 4-year-old children performed the response associated with the non-devalued training outcome significantly more often than

U. M. H. Klossek (✉) · S. Yu · A. Dickinson
University of Cambridge,
Cambridge, UK
e-mail: umhk2@cam.ac.uk

that trained with the now devalued outcome. In both cases, postdevaluation performance was assessed in extinction—that is, in the absence of outcome delivery—so the selection and initiation of the instrumental behaviors in question must have been controlled by knowledge about the relevant R–O relationships acquired during training, in combination with an evaluation of the current desirability of the available outcomes.

Whether or not an instrumental response is controlled by a representation of the R–O relationship and the current incentive value of the outcome depends on the conditions of training. In a second study by Adams (1982), different groups of rats received moderate versus extensive training (i.e., 100 vs. 500 reinforced leverpresses), followed by outcome devaluation and a subsequent extinction test. On test, only the moderately trained group showed reduced leverpressing for the devalued outcome; the extensively trained animals responded to the same extent, regardless of whether or not the outcome associated with leverpressing was devalued. More recently, Tricomi et al. (2009) demonstrated that overtraining also increases resistance to outcome devaluation in human adults.

In contrast to these findings, which demonstrated that extended training produced behavior that was autonomous of the current value of the outcome, Colwill and Rescorla (1985, 1988) reported that when animals were given a choice between two or more responses during training that produced different food outcomes on a concurrent schedule, their performance remained sensitive to the current value of the outcomes despite extensive training.

One explanation for the differential effect of extensive training on single-action versus choice training schedules was put forward by Adams (1982), who suggested that early in training, responding was controlled by a representation of the sensory properties of the outcome, but that with extensive exposure to the outcome, animals might increasingly come to encode its affective rather than its sensory properties. Because outcome devaluation is mediated by the sensory properties of the outcome, responding following extensive training should therefore be impervious to outcome devaluation. Holland (2004) proposed a similar explanation by suggesting that animals trained on single-action versus concurrent schedules might acquire different representational content about the outcome. Specifically, he argued that early in training, the outcome representations governing instrumental performance were more detailed in both cases, including both sensory and motivational properties. With extended training on a single-action schedule, maintaining detailed outcome representations would become increasingly irrelevant to the task, with the result that less detailed outcome representations, which encoded only more general, affective properties of the outcome, would come to control responding in the course of extended training. The experience of alternative outcomes that differ in their sensory properties on concurrent training schedules, on the other hand, should encourage processing of the sensory properties of the outcome, and therefore maintain detailed outcome representations that would remain susceptible to subsequent manipulations affecting the sensory properties of the outcome. The finding that training different responses concurrently with the *same* outcome did not maintain sensitivity to outcome value after extended training provided support for this account (Holland, 2004, Exp. 2).

Recently, Kosaki and Dickinson (2010) reexamined whether the resistance-to-overtraining effect conferred by concurrent schedules could be explained in terms of the differential outcome exposure maintaining detailed outcome representations. They trained different groups of rats on single-action and concurrent schedules. As in the study by Adams and Dickinson (1981), the single-action groups were trained to leverpress for one food outcome while receiving a different food outcome independently of leverpressing. In the concurrent group, the different food outcomes were contingent on performing responses on different levers. Critically, both groups received extensive training while experiencing a similar number of outcomes of each type. If exposure to multiple outcomes per se produced sustained sensitivity to outcome value, then both groups should have shown reduced responding for an expected, currently devalued outcome on test. However, this was not the case; Kosaki and Dickinson found that only the concurrently trained animals were sensitive to outcome value, and consequently, they argued that the choice contingencies arranged by the concurrent schedules maintained sensitivity to outcome value.

Although exposure to multiple outcomes per se did not appear to maintain sensitivity to outcome value in the study by Kosaki and Dickinson (2010), their findings leave open the possibility that instrumental outcomes are processed in a different way from events that are not contingent on the individual's own actions. Because only the concurrent group was exposed to two different *action-produced* outcomes, we cannot be certain that the contingent and noncontingent outcomes in the single-action group received equivalent processing, and consequently, whether the training conditions provided animals in the single-action group with the same outcome experience as those in the concurrently trained group. If the noncontingently delivered food outcomes were, for instance, represented primarily in terms of their affective properties to start with, or were subject to a rapid shift in processing bias from sensory to affective attributes over trials, then exposure to contingent and noncontingent food outcomes might not maintain processing of the sensory attributes of the contingent outcome in the same way that a second, response-contingent outcome would.

One way to examine this possibility would be to compare the training of two concurrent R–O relationships with single-action training of the same two R–O relationships on separate trials. In this way, both single-action and choice trained groups would experience two response-contingent outcomes during training, thereby ensuring that any differences in sensitivity to outcome value between the groups could not be attributed to potential differences in the processing of noncontingent and contingent outcomes during single-action training. If exposure to multiple, response-produced outcomes preserves processing of their sensory features, and if this is the principal reason why choice training maintains sensitivity to outcome value, then both groups should continue to process detailed outcome representations and remain sensitive to outcome value following devaluation of one of the training outcomes. On the other hand, if the critical feature of concurrent training for maintaining sensitivity to outcome value is that it provides a choice between actions that lead to different outcomes, as Kosaki and Dickinson (2010) suggested, then choice rather than single-action training should be more likely to maintain sensitivity to outcome devaluation.

The aim of the present experiment was to evaluate these contrasting predictions in preschool children, using the touch-response paradigm developed by Klossek et al. (2008). Two groups of preschool children between 3 and 4 years of age received either single-action or choice training. During choice training, two response options were available on every trial, and children could choose freely which action to perform. In the single-action group, the two R–O relationships were trained separately, so that only a single response option was available on any given trial, but both R–O contingencies were experienced in pseudorandom alternation across different trials.

Sensitivity to the current outcome value was subsequently assessed in the same way as in the original study by Klossek et al. (2008), by devaluing one of the two training outcomes and subsequently testing children's propensity to perform the two actions in an extinction test, during which no outcomes were presented. The devaluation procedure was based on current knowledge about the general dynamics of visual habituation in infants and young children (see, e.g., Schöner & Thelen, 2006) and exploited the novelty preference at short delay that is usually observed after massed presentations of a visual stimulus. Klossek et al. found this procedure to be effective in devaluing the animated scenes used as outcomes in the touch-response paradigm. After repeatedly presenting one of the video outcomes in the absence of the opportunity to perform the trained responses, Klossek et al. established that children between 18 and 48 months preferred the nonexposed video, by demonstrating that children preferentially chose to perform the response that yielded this outcome rather than the exposed one.

If experiencing multiple R–O relationships per se establishes sensitivity to outcome value, then it should not matter whether these experiences occur across different trials with only one available response option or during a series of choice trials. In the touch-response paradigm, all trials end with the presentation of an outcome and simultaneous removal of any response manipulanda. Therefore, only one R–O relationship can be sampled on a given trial in both the single-action and choice groups. Because the difference between the training trials experienced by these groups was confined to the presence of the second response manipulandum on choice trials, the training experience in the two groups should otherwise have been maximally similar. However, if the opportunity to choose between two outcomes during training enhances encoding of the outcome and its association with the relevant action, we anticipated that the choice group would show a greater devaluation effect than the single-action group.

## Method

### Participants

Forty-two children (24 boys, 18 girls) recruited from preschools and day nurseries in the Cambridge (U.K.) area who were between 3 and 4 years of age took part in the study. Informed, written parental consent was obtained prior to testing. Of the children, 21 were assigned to the single-action training group (9 girls, 12 boys), and another 21 were assigned to the choice training group (9 girls, 12 boys). The mean ages of the choice and single-action groups were 44.2 (*SD* 6.5) and 43.1 (*SD* 6.5) months, respectively [$t(40) = 0.54$, $p = .6$].

### Apparatus and stimuli

The task was run on a laptop computer (DELL XPS M1330) connected to a 17-in. touch-sensitive monitor (resistive, Hyundai G70TR; 1,280×800). The software used for controlling stimulus presentation and recording of responses was written and compiled using Microsoft Visual Basic Professional 6.0. Two 9×7.5 cm squares displaying a red versus a green butterfly icon, respectively, were the target areas for children's touch responses. Each of these response areas appeared consistently at the same distance from the center on either the left or the right side of the display. During single-action training, only one response area was shown on every trial, whereas during choice training, both response areas were visible side by side on every trial. Six 11-s video clips featuring popular children's characters from two different series (three clips from Cartoon A, three clips from Cartoon B) were used as outcomes. All clips were in color and included music and other sounds and vocalizations produced by the characters involved, but had no explicit verbal content.

## Procedure

Children were seated at a table within easy reach of the touch-sensitive monitor. Sessions took place in a familiar room in the children's regular day nursery. Each session began with a brief warm-up period during which the child was introduced to the investigator and the apparatus. While a child was completing the activity, the child's carer and the investigator sat on either side of the child.

*Pretraining* Children in both groups completed identical brief pretraining sequences, which consisted of two demonstration trials and two single-action warm-up trials. On the first pretraining trial, only the response area on the right side of the display was shown on the screen. The experimenter directed the child's attention toward the butterfly icon and then proceeded to demonstrate how to perform touch responses on the response area using the index finger of her right hand, triggering a presentation of a video clip from Cartoon A. On the next trial, the same butterfly picture appeared again in the same location, and the child was encouraged to perform a touch response. The child's first touch on the butterfly icon immediately produced another video clip from Cartoon A. After the child had completed this first warm-up trial, the display showed the other butterfly icon on the left side on the screen. As on the first trial, the investigator directed the child's attention toward the butterfly icon and then demonstrated how to start up a video clip from Cartoon B by performing touch responses directed at the butterfly image. Following the video presentation, the same icon appeared again on the following trial, and it was the child's turn to touch the icon and trigger another clip from Cartoon B. In each group, the position (left vs. right) of red and green butterfly icons and the assignment of Cartoons A and B to the two responses were counterbalanced across participants, so that each of the eight possible combinations of these variables was experienced by at least 2 children per group.

*Choice training* In the choice group, both butterfly icons were displayed side by side on every training trial, and children could freely choose which actions to perform on a given trial. Throughout choice training, the number of responses required in order to obtain a video clip on a given trial varied randomly between one and five for each action. Trials ended whenever the number of responses on a given icon matched the criterion set for that action on that trial. Until then, responses could me made on both icons and were recorded, but did not have any other programmed consequences. The purpose of using a variable response requirement during training was two-fold. First, it enabled the calculation of response rates during training, which would not have been possible if every response had produced an outcome. Second, training on such a partial reinforcement schedule should have established performance that was more resistant to extinction at the time of testing. In our previous study, following training that required only a single response to obtain a video presentation, many children ceased to respond almost instantly when their actions were not followed by an immediate outcome in the extinction test.

On each trial, the first response to meet the criterion caused the butterfly icons to disappear and the video clip assigned to the criterion response to be displayed in the center of the screen, leaving only an approximately 6-cm-wide margin on all sides where a uniform, light gray background was shown. Following the termination of the video clip, the butterfly icons were again presented and the response schedule for each action reinstated. The training phase ended when at least 10 video clips of each type had been obtained or a time limit of 450 s had been reached.

*Single-action training* Following the warm-up trials, the single-action training group completed 20 training trials (10 with each outcome type). On each of these trials, only one response area was displayed, until the child's touch responses caused a video clip to play. The number of actions required to produce an outcome on each trial varied between one and five. Each time a video outcome was shown, the butterfly picture disappeared and the video display appeared in the same format as for the choice group. All children completed the same pseudorandom sequence of left (L) and right (R) trials (R L L L R R L R L R R R L R R L R L L R L).

*Outcome devaluation* Both groups then experienced the same outcome devaluation procedure: The butterfly pictures from the training display disappeared and, following a brief interval of approximately 5 s, during which only the gray background was visible, the display background color changed from gray to blue. The video display then appeared, and the three video clips from one of the two cartoon series were repeated five times in sequence with a 3-s interval between clips. Because R–O assignments and icon locations were counterbalanced across participants, this meant that in each group, approximately half of the children experienced devaluation of the video outcome associated with left responses ($N$ = 10) and the remaining half of the children experienced devaluation of the video outcome associated with right responses ($N$ = 11).

*Extinction test* After outcome devaluation, the video display vanished, the background display color changed back to gray, and after a brief interval of approximately 5 s, the butterfly icons reappeared. Unknown to the children, both response areas were now deactivated for a period of 1 min, so that touch responses did not cause any video outcomes.

## Results

The mean training duration was 389 s in the choice group, which was significantly longer than that in the single-action group, which took on average 35 s less (354 s) to complete training, $t(40) = 2.3$ $p = .031$. The mean numbers of outcomes obtained were similar for the single-action and choice groups. The choice group chose a mean of 10.1 (*SD* 2.7) of the to-be-devalued outcome and 9.5 (*SD* 2.2) to-be-valued outcomes, which closely matched the 10 outcomes of each type that were earned by the single-action group.

As Fig. 1 illustrates, response rates in the single-action group were higher than in the choice group during training, but none of the groups showed a significant preference for one of the two responses during training. A Group × Response Type (to be devalued, to be valued) mixed ANOVA of the training rates confirmed that the training rates in the single-action group were significantly higher than the rates in the choice group [$F(1, 40) = 85.2$, $p < .001$]. Importantly, however, the effect of group did not interact with that of response type ($F < 1$), and the responses that produced to-be-valued and to-be-devalued outcomes were performed at similar rates during training ($F < 1$).

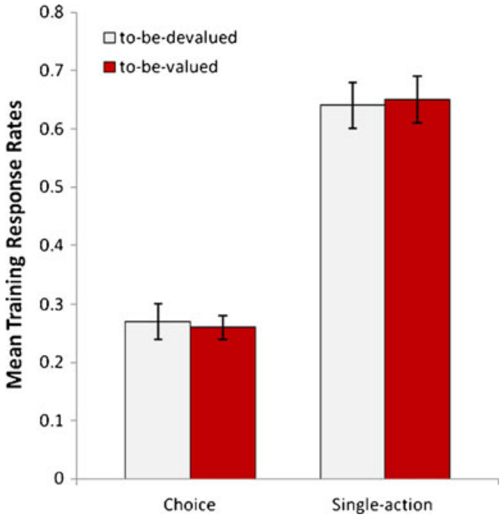

**Fig. 1** Mean response rates per second during training for the choice and single-action groups. Error bars represent the standard errors of the means

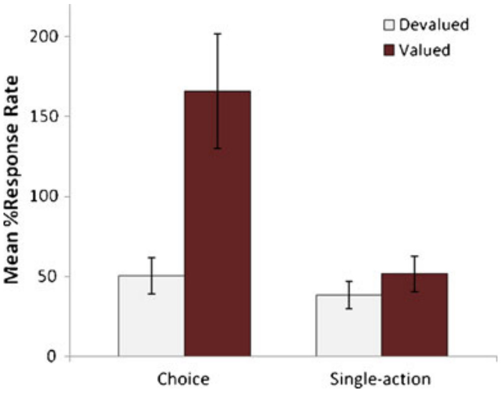

**Fig. 2** Mean percentage response rates for the choice and single-action groups during the postdevaluation extinction test. Error bars represent the standard errors of the means

Given that there was no competition between different response alternatives in the single-action group, the finding that single-response training produced higher response rates is not surprising: On the choice schedule, responding for one outcome always reduced the opportunity to perform the alternative response, which was not the case in the single-action group. A comparable difference was observed by Kosaki and Dickinson (2010), who reported consistently lower response rates for rats that were trained on two responses concurrently than for rats that were only presented with a single response option.

Because the response rates varied between the groups during training, the rates during the extinction test were expressed as a percentage of the absolute training rate in order to minimize the contribution of between-child variance. Figure 2 shows these mean percentage response rates for the two responses during the postdevaluation extinction test. During the extinction test, response rates were not constrained by the video presentations in the same way as during training. As a result, it was possible for the relative response rates on test to exceed 100%, which is the reason Fig. 2 suggests an overall marked increase in responding for the valued outcome in the choice group.

The results showed that the test performance of the choice group was sensitive to outcome devaluation, in that these children performed the response whose training outcome had been devalued less than the response whose outcome had not been devalued. In contrast, responding in the single-action group was insensitive to outcome devaluation, as children performed both actions at comparable rates. This description of the data was supported by a Group × Response Type (valued vs. devalued) mixed ANOVA. As is usual with this paradigm (see Klossek et al., 2008), the variance of the relative rates increased with the mean, and the data were therefore square-root transformed prior to analysis. Homogeneity of the error variances for the transformed data was confirmed by Levene's test for both valued

[$F(1, 40) = 2.0$, $p = .16$] and devalued [$F(1, 40) = 1.6$, $p = .21$] responding. Analysis of these data revealed a significant effect of Response Type [$F(1, 40) = 9.2$, $p = .004$] and, more importantly, a significant Group × Response Type interaction [$F(1, 40) = 4.3$, $p = .044$]. Planned analyses of the simple main effects confirmed that the choice group showed a significant preference for the valued response during the test [$F(1, 40) = 13.1$, $p = .001$], whereas no significant effect of response type was found for the single-action group ($F < 1$).

It is conceivable that the apparent insensitivity of the single-action group to outcome value reflected a decrement in stimulus generalization from the displays shown on training to those shown on the test trials, because this group had not been exposed to choice trials prior to the test. To examine this possibility, we compared the absolute response rates for the action trained with the valued outcome between the two groups during the extinction test. The mean rates of .35 (*SE* .06) and .32 (*SE* .06) valued responses per second for the choice and single-action training groups, respectively, did not differ reliably [$F(1, 40) = 0.19$, $p = .67$]. This finding suggests that the generalization decrement from training to test did not affect children in the single-action group to a greater extent than children in the choice group when a choice procedure was employed during testing.

## General discussion

The present findings join those of previous investigations that have demonstrated an ability to choose actions flexibly in pursuit of currently valued goals in children 2 years of age and older (Kenward, Folke, Holmberg, Johansson, & Gredebäck, 2009; Klossek et al., 2008). Our results also replicate the observation that choice training in which two responses yield different outcomes produces performance that is sensitive to current outcome value, which has been demonstrated in animal studies (e.g., Colwill & Rescorla, 1985, 1988), as well as in preschool children (Klossek et al., 2008). Moreover, as in previous studies, the present findings showed that choice-trained responding was sensitive to outcome devaluation even when an equivalent amount of single-action training had produced behavior that was autonomous of the current value of the outcome (Holland, 2004). In accord with the findings of Kosaki and Dickinson (2010), exposing the agent to two different outcomes throughout training does not appear to be the critical aspect of choice training that is responsible for producing this differential sensitivity to outcome value, because our choice and single-action groups experienced very similar outcome exposures.

The present study therefore extends our understanding of choice training in two respects. First, it establishes the critical role of choice training in maintaining goal-directed behavior not only in rats, but also in humans. Second, it determines the aspect of choice training that enhances sensitivity to the current value of the outcome. Kosaki and Dickinson (2010) matched outcome exposures during training by contrasting choice training with single-action training in which the alternative outcome was presented noncontingently. What the present results show is that choice training favors sensitivity to outcome value relative to single-action training in which *both* outcomes are response contingent. This finding reinforces the conclusion that choice training does not maintain sensitivity to outcome value simply by providing exposure to different outcomes throughout training.

We cannot be certain in the present study whether a transition from sensitivity to outcome value, to behavioral autonomy occurred in the single-action group, because we have not examined the effect of varying single-action training in the present paradigm. Kenward et al. (2009) did report that 2-year-old children were sensitive to current outcome value following single-action training, but procedural differences obviate generalization between the two paradigms. Therefore, it is possible that the single-action procedure used in the present study established performance that was autonomous of the current value of the outcome from the start of training.

Dual-process accounts of instrumental behavior offer alternative accounts of variations in sensitivity to outcome value with training. These accounts assume that this variation arises from the fact that instrumental behavior can be controlled by more than one process. Two such accounts have been put forward. Daw, Niv, and Dayan (2005) proposed a competing-systems account whereby different systems generate value predictions for available response options through qualitatively different computational processes, namely temporal difference learning, or caching, and iterative tree search. *Caching* involves predicting the probability of future reward based on an average value that reflects accumulated past experience, without encoding the identity of the outcome. Behavior controlled by this computation is therefore unable to adapt immediately to any changes in current outcome value.

Instrumental action sensitive to and regulated by its consequences is controlled by the system that uses tree search to generate value predictions. The tree-search mechanism implements a forward model that works out short-term value predictions for the immediate consequences of each action in the form of an iterative search, which explores possible future states associated with available response alternatives. Because the tree-search mechanism generates instant-by-instant predictions of a specific outcome value, instrumental behavior can be adapted immediately to changed circumstances—for instance, following outcome revaluation. Arbitration between these systems is realized in terms of uncertainty reduction, so that the system generating the more certain value prediction controls behavior.

Different learning situations and task demands will therefore tend to favor control by one system versus the other. Task simplicity and extensive training, for instance, promote control by the caching system, whereas under conditions of greater task complexity, the tree-search mechanism tends to generate more certain value predictions and will therefore control behavior. In the present study, this dual-process account would explain the results in terms of the operation of these different processes in the two groups. During single-action training, the caching system should generate more certain value predictions and take over control of behavior, rendering performance insensitive to the change in current outcome value in the postdevaluation extinction test. Conversely, presenting different response options simultaneously on choice training trials should preferentially engage behavioral control by the tree-search process. As a result, participants in the choice group could adapt their performance flexibly to postdevaluation changes in outcome value on test.

An alternative dual-process account of behavioral autonomy was offered by Dickinson (1985, 1989). This account proposes that behavior is conjointly controlled by S–R and R–O associations that are formed during instrumental learning, with the R–O process mediating sensitivity to changes of outcome value in a devaluation test. In contrast to the dual-systems account of Daw et al. (2005), there is no explicit action selection process, and indeed, the model assumes that the influence of the S–R and R–O associations summate in controlling performance. This model explains the effects of overtraining in terms of two processes: First, extensive experience of rewarded actions will strengthen relevant S–R associations through a standard reinforcement process (Hull, 1943; Thorndike, 1911). Second, training conditions that provide the experience of a strong behavior–outcome correlation promote the formation of a strong R–O association, which ensures that performance remains sensitive to the current value of expected consequences. Conversely, training conditions that fail to provide the subjective experience of a strong behavior–reward correlation, such as tasks that discourage behavioral variation or arrange a weak behavior–outcome correlation, such as interval schedules (Dickinson, Nicholas, & Adams, 1983), will not sustain a strong R–O association, leaving responding controlled by S–R associations, and therefore autonomous of the current value of the outcome.

Choice training sustains control by the R–O learning process, and therefore sensitivity to outcome value, because it ensures continued experience of the R–O contingency. For example, on trials when the children in the present study chose the left response, they received the associated cartoon outcome, and on trials when they chose the other action, they experienced that the left-associated outcome did not occur. Therefore, choice training exposed the children to the differential likelihoods of the associated outcome following the left response relative to the absence of the response, thereby ensuring experience of the full R–O contingency within a common stimulus context throughout choice training. By contrast, during single-action training, the children never experienced the absence of the associated outcome in the absence of the response in the stimulus context provided by the relevant butterfly stimulus, and therefore did not experience the full contingency necessary for R–O learning. Consequently, according to this version of dual-process theory, responding in the single-action group would have to be mediated by S–R learning.

Perhaps more problematic for Dickinson's (1985, 1989) dual-process account is the observation by Tricomi et al. (2009) that the performance of adult humans was insensitive to outcome devaluation following extended training on a nominal choice procedure. However, although two possible response options were presented on every trial, participants were instructed at the beginning of every trial which response to carry out. Therefore, each response was performed within the separate and distinctive stimulus context provided by the instructions, thereby making the effective training contingencies more similar to those of the single-action group in the present study.

To conclude, the present study confirms that giving an agent a free choice between actions that yield different outcomes promotes goal-directed action control. Moreover, the findings show that the critical feature of the choice contingencies that prevent the development of behavioral autonomy cannot be reduced to the experience of the two R–O options on their own.

## References

Adams, C. D. (1982). Variations in the sensitivity of instrumental responding to reinforcer devaluation. *Quarterly Journal of Experimental Psychology, 34B*, 77–98.

Adams, C. D., & Dickinson, A. (1981). Instrumental responding following reinforcer devaluation. *Quarterly Journal of Experimental Psychology, 33B*, 109–121.

Colwill, R. M., & Rescorla, R. A. (1985). Instrumental responding remains sensitive to reinforcer devaluation after extensive training. *Journal of Experimental Psychology: Animal Behavior Processes, 11*, 520–536.

Colwill, R. M., & Rescorla, R. A. (1988). The role of response-reinforcer associations increases throughout extended instrumental training. *Animal Learning & Behavior, 16*, 105–111.

Daw, N. D., Niv, Y., & Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nature Neursoscience, 8*, 1704–1711.

Dickinson, A. (1985). Actions and habits: the development of behavioural autonomy. *Philospohical Transactions of the Royal Society B, 308*, 67–78.

Dickinson, A. (1989). Expectancy theory in animal conditioning. In S. B. Klein & R. R. Mowrer (Eds.), *Contemporary learning theories: Pavlovian conditioning and the status of traditional learning theory* (pp. 279–308). Hillsdale: Erlbaum.

Dickinson, A. (1994). Instrumental conditioning. In N. J. Mackintosh (Ed.), *Animal learning and cognition. Handbook of perception and cognition series* (pp. 45–79). San Diego: Academic.

Dickinson, A., & Balleine, B. W. (1993). Actions and responses: The dual psychology of behaviour. In N. Eilan, R. McCarthy, & B. Brewer (Eds.), *Spatial representation* (pp. 276–293). Oxford: Blackwell.

Dickinson, A., Nicholas, D. J., & Adams, C. D. (1983). The effect of the instrumental training contingency on susceptibility to reinforcer devaluation. *Quarterly Journal of Experimental Psychology, 35B*, 35–51.

Holland, P. C. (2004). Relations between Pavlovian–instrumental transfer and reinforcer devaluation. *Journal of Experimental Psychology: Animal Behavior Processes, 30*, 104–117.

Hull, C. L. (1943). *Principles of behavior*. New York: Appleton.

Kenward, B., Folke, S., Holmberg, J., Johansson, A., & Gredebäck, G. (2009). Goal directedness and decision making in infants. *Developmental Psychology, 45*, 809–819.

Klossek, U. M. H., Russell, J., & Dickinson, A. (2008). The control of instrumental action following outcome devaluation in young children aged between 1 and 4 years. *Journal of Experimental Psychology: General, 137*, 39–51.

Kosaki, Y., & Dickinson, A. (2010). Choice and contingency in the development of behavioral autonomy during instrumental conditioning. *Journal of Experimental Psychology: Animal Behavior Processes, 36*, 334–342.

Schöner, G., & Thelen, E. (2006). Using dynamic field theory to rethink infant habituation. *Psychological Review, 113*, 273–299. doi:10.1037/0033-295X.113.2.273.

Thorndike, E. L. (1911). *Animal intelligence: Experimental studies*. New York: Macmillan.

Tricomi, E., Balleine, B. W., & O'Doherty, J. P. (2009). A specific role for posterior dorsolateral striatum in human habit learning. *European Journal of Neuroscience, 29*, 2225–2232.

Valentin, V. V., Dickinson, A., & O'Doherty, J. P. (2007). Determining the neural substrates of goal-directed learning in the human brain. *Journal of Neuroscience, 27*, 4019–4026.

## Author's note

This research was supported by a grant from the UK Biotechnology and Biological Sciences Research Council.

We would like to thank Jim Russell for his generous support and provision of testing facilities.