Neth, H., & Payne, S.J. (2002). Thinking by Doing? Epistemic Actions in the Tower of Hanoi. In W.D. Gray and C.D. Schunn (Eds.), Proceedings of the Twenty-Fourth Annual Conference of the Cognitive Science Society (pp. 691—696). Mahwah, NJ: Lawrence Erlbaum.

Thinking by Doing? Epistemic Actions in the Tower of Hanoi

Hansjörg Neth (NethH@Cardiff.ac.uk)
Stephen J. Payne (PayneS@Cardiff.ac.uk)
School of Psychology, Cardiff University
Cardiff CF10 3YG, Wales, United Kingdom

Abstract

This article explores the concept of epistemic actions in the Tower of Hanoi (ToH) problem. Epistemic actions (Kirsh & Maglio, 1994) are actions that do not traverse the problem space toward the goal but facilitate subsequent problem solving by changing the actor's cognitive state. We report an experiment in which people repeatedly solve ToH tasks. An instructional manipulation asked participants to minimize moves either trial by trial or only on the last three of six trials. This manipulation did not have the predicted effect on the trial-by-trial move counts. A second, device manipulation provided some participants with an "exploratory mode" in which move sequences could be tried then undone without affecting the criterion move count. Participants effectively used this mode to reduce moves on each trial, but there was no clear evidence that they used it to learn about the problem across trials. We conclude that there is strong evidence for one sub-type of epistemic action (acting-to-plan) but no evidence for a second sub-type (acting-to-learn).

Introduction

How do we learn to solve a problem? The most popular view within the Cognitive Science community is that we do so by solving the problem. Anzai and Simon's (1979) theory of 'learning by doing' marks a major breakthrough in research on learning through problem solving. They proposed an adaptive production system which mirrored the strategy transformations of a human participant as she solved the Tower of Hanoi (ToH) problem, and in so doing provided the impetus for many subsequent theories of the mechanisms by which problem solving leads to learning (e.g. Klahr, Langley & Neches, 1987).

All learning-by-doing accounts share the assumption that learning about a particular problem occurs as an automatic by-product of problem solving activity. However, in many problem solving situations learning may be more deliberate than the learning-by-doing account implies. We suggest that problem solvers may sometimes orient themselves to *learning goals* rather than *solution goals* (O'Hara & Payne, 1998; Trudel & Payne, 1995).

In relation to the ToH task, this position is encouraged by VanLehn's (1991) re-analysis of the original Anzai & Simon (1979) protocol, in which he notes that the participant was "acting like a scientist" (p. 16) and repeatedly suspended her problem solving activity to acquire new strategic knowledge.

Further general support for a deliberate learning mode nested within problem solving activity can be derived from the work of Kirsh and colleagues (1995, Kirsh & Maglio, 1994), who have explored a distinction between goal-directed *pragmatic actions* and *epistemic actions* whose primary purpose is to improve cognition by changing an agent's computational state. Although epistemic actions are not immediately goal-directed, they may improve subsequent performance through their cognitive effects.

The primary goal of this article is to seek experimental evidence for the use of epistemic actions in problem solving with the ToH puzzle. Identifying epistemic actions in ordinary problem solving activity is difficult, because they are only distinguished by their cognitive motivations and consequences rather than directly observable characteristics (and not all actions that do not successfully move toward the goal are epistemic!). We use two manipulations that may allow participants to utilise epistemic actions, and at the same time facilitate their detection. The first manipulation is *instructional*: participants were asked either to optimize their performance on every problem solving trial, or on trials 4, 5 and 6 of a series of six repeated problems. We hypothesize that delaying the enforcement of the performance criterion will encourage a learning orientation, and the use of epistemic actions, during the early first trials. The second manipulation is to provide *device support* that enables participants to separate pragmatic from epistemic actions. Thus, our computer-based version of ToH allowed participants to switch into an "exploratory mode" in which they could make move sequences that were later undone and were not counted towards the performance criterion.

These twin manipulations allow us to refine Kirsh's formulation of pragmatic and epistemic actions by distinguishing between two kinds of epistemic action: those that have only immediate within-problem effects (*acting-to-plan*) and those that have longer-term cognitive consequences (*acting-to-learn*). If the exploratory mode is used merely as an external support for look-ahead or planning, motivated by questions such as 'Is this a good sequence of moves?', we would regard such usage as acting-to-plan. On the other hand, if additional actions on earlier trials are shown to lead to better problem solving on later trials we would have evidence for acting-to-learn.

To anticipate our conclusions, we find strong support for acting-to-plan, but no decisive support for acting-to-learn.

Method

Participants

Forty-four Psychology undergraduates (with a mean age of 20.7 years) took part in the experiment to receive course credit. Participation was restricted to first year undergraduate students who reported no prior exposure to the task. All participants were familiar with graphical user interfaces and did not suffer from any perceptual or cognitive impairments.

Apparatus

The experiment used a graphical software version of the ToH problem which was programmed in Visual Basic 6 and displayed on a 17" screen. A disk could be transferred between towers by indicating its source and target locations using a drag-and-drop procedure. In case of an illegal move there was an auditory warning signal and the selected disk slid back to its original position. A counter showing the current number of pragmatic moves was displayed in the top right hand corner of the screen.

Materials

Participants had to solve a sequence of 5-disk ToH puzzles in the standard tower-to-tower version. To prevent improvements due to superficial rote memorization we used six simple isomorphs, which were created by systematically switching the source and target towers.

Design

As we wanted to test participants' *spontaneous* use of epistemic actions we did not want to specifically encourage them to explore the problem, but rather provide subtle opportunities that may be used or ignored.

The *instructional manipulation* consisted of two levels. Participants were either instructed to optimize their performance (i.e., minimize the number of pragmatic moves needed to solve the puzzle) on each of several problems, or asked to optimize their performance on the last three of six problems. Hence, whereas the first group of participants was implicitly discouraged from using epistemic actions by the instruction to be *performance oriented* throughout an unspecified number of trials, the second group was presented with an opportunity to be *learning oriented* in the first three of six trials.

The second experimental manipulation consisted in withholding or providing *device-support* for epistemic actions. Two different versions of the device were distinguished:

In the standard *pragmatic moves only* condition, each move of a disk on the screen counted towards the performance criterion of minimizing the number of (pragmatic) moves.

	Device-support	*Instruction*
1.	pragmatic mode only	'minimize on trials 1–6'
2.	pragmatic mode only	'minimize on trials 4–6'
3.	pragmatic+epistemic mode	'minimize on trials 1–6'
4.	pragmatic+epistemic mode	'minimize on trials 4–6'

Table 1: Overview of the two experimental factors and four groups.

In a second *pragmatic plus epistemic moves* condition two different device modes were introduced to the participants. Whilst having to solve the puzzle in so-called "solution mode", participants had the option to switch into an "exploration mode" at any point by pressing and holding down the Shift key. Whereas in both modes disks could be moved in an identical fashion, moves made in exploration mode were not added towards the total performance score and always reversed when switching back into "solution mode" by releasing the Shift key.

Note that the specific design of exporatory mode addresses the difficulty of detecting epistemic moves by effectively creating an *operational definition*: Since participants are aware of the mandatory reversal of all moves made in exploratory mode, entering the mode signals the use of epistemic moves.

One way of characterizing both the instructional and device manipulation is that they do not prevent learning by doing, but provide additional opportunities for *learning by not solving* the puzzle. A combination of both experimental factors yielded the four experimental groups shown in Table 1.

As each participant had to solve a total of six ToH puzzles the experiment employed a mixed design, with *device-support* and *instruction* as between-subjects manipulations and *trial* as a within-subjects factor.

Procedure

Each participant was assigned to one experimental group according to the order of arrival at the laboratory. After reading a generic description of the Towers of Hanoi puzzle participants were introduced to the graphical user interface. To demonstrate that they had understood the task constraints and to familiarize themselves with the user inteface they solved a simple two-disk version of the puzzle.

Participants then received their respective minimization instructions and were told that the experiment normally takes around 45 minutes regardless of their individual performance.

After each successful completion of a problem, participants received a brief message reminding them of their respective minimization instruction before starting the next trial.

On average, participants completed the experiment within 40 minutes.

Predictions

Our primary predictions refer to comparisons between and within experimental groups (rather than assuming a 2x2 factorial design; in particular Group 4 plays a subsidiary role in the study, and will only be analysed in relation to first-order findings).

The main predictions concern the number of pragmatic moves needed to solve the puzzle. As we expect all groups to learn throughout the course of the experiment, we predict a gradual reduction of the mean number of pragmatic moves required to solve the puzzle. This familiar practice effect constitutes the baseline which we expect to be modulated by the experimental factors of instructional goal orientation and device support.

If the instructional manipulation encourages members of Group 2 to invest additional moves in trials 1–3 and this in turn results in better learning, they ought to need fewer moves on trials 4–6. Thus, we predict an *interaction* of instruction and trial for Groups 1 vs. 2.

Next, if participants are spontaneously capable of using the exploratory device mode to improve their performance, Group 3 should need fewer pragmatic moves than Group 1 in all trials. Hence, we predict a *main effect* of device support on the number of pragmatic moves for Groups 1 vs. 3.

Our secondary predictions involve Groups 3 and 4 and address different possible motivations for epistemic moves:

If the exploratory device mode is primarily used for *learning* purposes (acting-to-learn) we should find an instant use of epistemic moves in both Group 3 and 4. If learning actually occurs, the frequency of epistemic moves should decrease over time. If, on the other hand, epistemic moves are used primarily for *online planning* within a trial (acting-to-plan) we expect a more opportunistic use due to the instructional manipulation. In this case, we expect the frequency of epistemic moves in Group 3 and 4 to display an *interaction* over trials.

Finally, if the use of epistemic mode is unselective, and predominantly due to *affordances* created by the design of our device we should find a constant use of epistemic moves throughout all trials and similar usage patterns in Groups 3 and 4.

Results

Numbers of Moves

As all groups were instructed to minimize the number of moves to solve the ToH puzzle our comparative analysis of their performance will be based on the number of pragmatic moves per trial.

Overall learning effects Before we consider the comparisons between individual groups, we will examine the expected overall effects of learning. Figure 1 displays the mean number of pragmatic moves for each group from trials 1 to 6. A mixed ANOVA of pragmatic moves with *group membership* as between-subjects factor and *trial* as within-subjects factor confirms the overall learning effect [$F(5,200)=16.3$, $p=.000$, $MSE=260.4$] and indicates that there were differences between groups [$F(3,40)=5.8$, $p=.002$, $MSE=489.2$], but the interaction between the two factors did not reach significance [$F(15,200)=1.5$, $p=.13$, $MSE=260.4$].

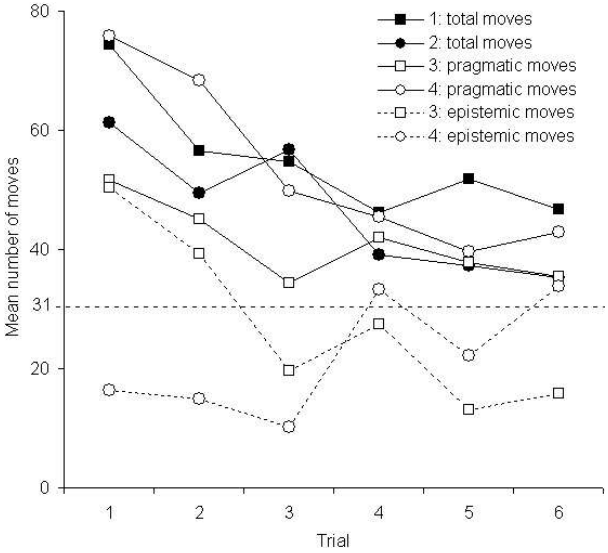

Figure 1: Mean number of moves for each of the four groups on each of six trials. For Groups 1 and 2 the number of pragmatic moves corresponds to the number of total moves, whereas Groups 3 and 4 had the option of making epistemic moves in addition to pragmatic moves. (Note: The minimum possible number of pragmatic moves to solve the task is 31.)

Whether an individual group has significantly improved over time can be assessed by conducting simple effect tests within groups by trial. These show that Groups 1, 2 and 4 significantly reduced their number of pragmatic moves over the course of the experiment. The means for Group 3 were low even at the first trials, suggesting that it did not improve significantly because it consistently performed at a high level.

On trials 4–6, in which all groups were instructed to minimize the number of pragmatic moves, both Group 2 and 3 outperformed Group 1 by taking fewer pragmatic moves ($p=.003$ and .008 respectively).

Thus, the overall results seem promising: With respect to the performance criterion Groups 1, 2 and 4 improved over time and Groups 2 and 3 managed to solve the ToH puzzle in fewer moves than Group 1 in the second half of the experiment.

Effects of Instruction (Groups 1 vs. 2) Our first specific prediction concerned Groups 1 and 2, neither of which had the epistemic device mode at their disposal, but differed in their instructions: Whereas Group 1 was instructed to minimize the number of moves in each trial, Group 2 was only asked to optimize their performance in

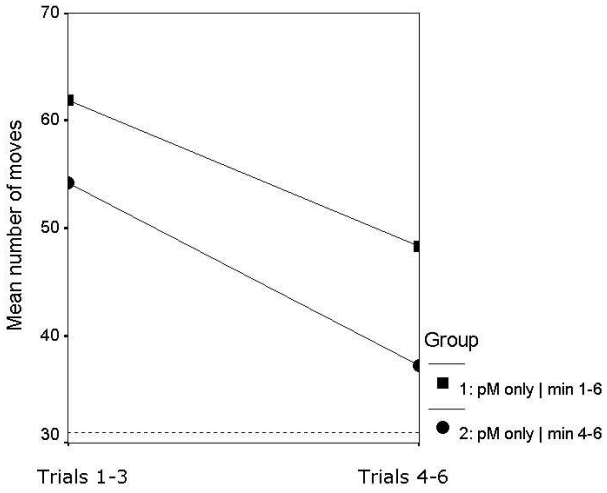

Figure 2: Mean number of moves for Groups 1 and 2 on trials 1–3 and 4–6. (Note: The minimum possible number of moves is 31.)

the last three of six trials. As the scope of this experimental manipulation juxtaposed trials 1–3 with trials 4–6 it is appropriate to collapse the data across each triple of trials by computing the respective means.

Figure 2 shows the mean number of moves for both groups on trials 1–3 and 4–6. It illustrates that there is no hint of the predicted crossover interaction [$F(1,20)=.44$, $p=.51$, $MSE=72.6$]. Instead, the predicted learning effects over both test halves [$F(1,20)=35.4$, $p=.00$, $MSE=72.6$] are combined with an unexpected main effect of group [$F(1,20)=8.1$, $p=.01$, $MSE=119.0$].

While successfully predicting that Group 2 would use fewer moves on trials 4–6 [$F(1,20)=8.9$, $p=.01$] it is immediately obvious that this advantage in performance cannot be attributed to its members using epistemic actions in the initial trials: they clearly have not invested *additional* moves on trials 1–3.

One plausible, if rather annoying explanation for this pattern of data, is that, by an accident of assignment, Group 2 might comprise more able problem solvers than Group 1. (We will briefly consider an alternative account below.)

Effects of Device Support (Groups 1 vs. 3) Groups 1 and 3 shared the same instructions (to minimize the number of moves on each trial) but differed in the options provided by the user interface (device). Specifically, members of Group 3 had the "exploratory mode" at their disposal which supported the use of epistemic moves.

In the overview of the number of pragmatic moves of all four groups we have already established that Group 3 performed better than Group 1 on trials 4–6. A mixed ANOVA of pragmatic moves by group membership and trial confirms the predicted main effect of group over all six trials [$F(1,20)=18.0$, $p=.000$, $MSE=359.9$]. Thus, Group 3 *consistently* performed better than Group 1 with respect to the criterion.

To interpret this difference in performance the number of epistemic moves of Group 3 has to be taken into account as well. (The number of epistemic moves is represented by dotted lines in Figure 1).

If we add the epistemic moves carried out by Group 3 to their pragmatic moves, Group 3 used more total moves on average than Group 1 (mean total moves=68.7 and 55.1 respectively), but this difference is not statistically significant [$F(1,20)=.11$, $p=.12$, $MSE=2165.3$].

This demonstrates that Group 3 spontaneously managed to use the device-supported option of epistemic moves to improve their performance with respect to the criterion. However, it leaves open exactly *why* and *how* members of Group 3 used the exploratory mode. We will address these issues after assessing the effects on solution latency.

Solution Times

Although participants had been told that the total experiment took a standardized length of time—hence could not assume that by being quick or slow they would alter the overall duration of their experimental session—their latencies to solve a problem can be used as an alternative indicator to assess their performance.

Overall effects As one might expect solution latencies over the course of the experiment decreased for all groups. An overall mixed ANOVA on the effects of group membership and trial on the total time required to solve each task yields a main effect of trial [$F(5,200)=16.9$, $p=.000$, $MSE=13352.7$] but no differences between groups. However, a significant interaction of the two factors [$F(15,200)=1.8$, $p=.038$, $MSE=13352.7$] drew attention to the possibility that different groups might have exploited time selectively to optimize their performance.

Subsequent simple main effect tests confirm that while the total solution times of Groups 1 and 3 significantly decreased over repeated trials, this was not the case for Groups 2 and 4. This suggests that the instructional manipulation had a selective effect on solution time, and in particular raises the possibility that the improved performance of Group 2 over Group 1 on trials 4–6 was caused in part by Group 2 exerting greater effort on these trials.

Effects of Instruction (Groups 1 vs. 2) The suggestion that Group 2 outperformed Group 1 in trials 3–6 by exerting extra effort (rather than the hypothesized investment of epistemic moves) is supported by an analysis of *move rates*, i.e. the number of moves made per second. Figure 3 shows the mean move rates of Group 1 and 2 over both test halves. A corresponding ANOVA on move rate by group and test half yields a highly significant interaction [$F(1,20)=18.3$, $p=.000$, $MSE=.002$].

If we interpret an increase in move rate (as seen in Group 1) as signalling the necessity of less effort per individual move, the *absence* of an increase in Group 2 suggests that its members invested relatively more effort in the second half of the experiment.

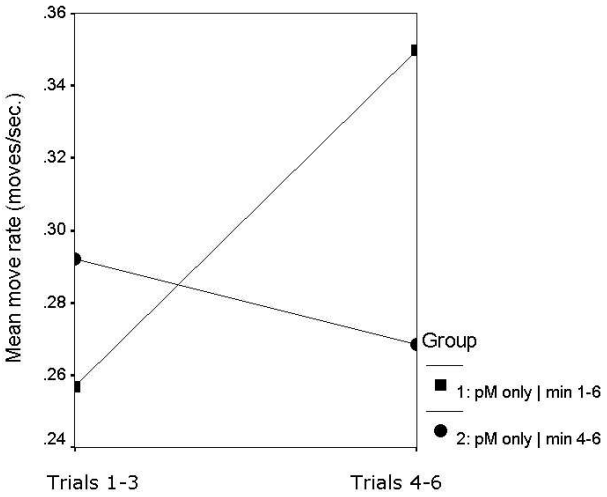

Figure 3: Mean move rates for Groups 1 and 2 on trials 1–3 and 4–6.

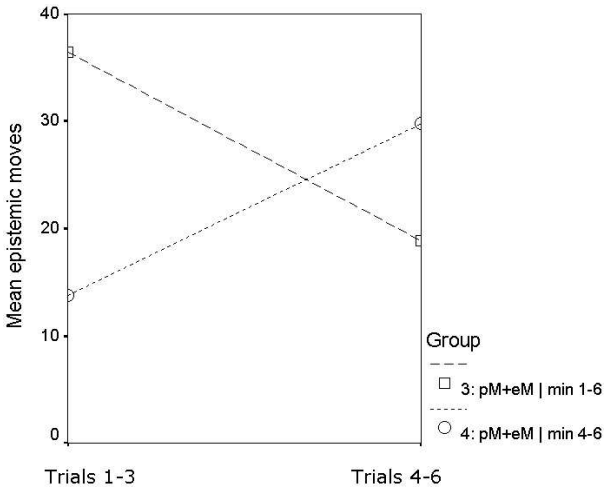

Figure 4: Mean number of epistemic moves for Groups 3 and 4 on trials 1–3 and 4–6.

Effects of Device Support (Groups 3 and 4)

One of the questions raised above was: *How* did Group 3 use epistemic moves to outperform Group 1? As the total solution times for Groups 1 and 3 did not differ [$F(1,20)=.38$, $p=.54$, $MSE=66525.2$] recourse to latency data does not resolve this issue.

Although the present experiment does not allow us to answer questions about possible causes and effects of device-supported epistemic moves conclusively, we can provide tentative evidence for some of the related issues:

- Did the use of epistemic moves actually lead to *better learning*? The fact that Group 3 *continued to use* epistemic moves until the last trials suggests that they probably did not learn more about the ToH puzzle than Group 1, but used the exploratory device mode as a tool to optimize their performance. However, our design allows for the alternative explanation that the continued use of epistemic moves might have been due to a conservative strategy.

- Did the use of epistemic moves become *more effective over time*? An index of the effectiveness of each epistemic move can be computed by dividing Group 3's mean savings of pragmatic moves (compared with Group 1) by the number of epistemic moves invested on each trial. As the six corresponding ratios (0.5, 0.3, 1.0, 0.2, 1.1, 0.7) do not show any obvious trend, we conclude that the use of exploratory mode did not become more effective over time.

- Were epistemic moves used *to learn* or *to plan*? Even without evidence for superior learning due to device-support of epistemic moves we can address the question of participants' *motivation* to use exploratory mode by comparing the usage patterns of Group 3 and 4. Figure 4 shows a clear interaction of group membership and test half on the mean number of epistemic moves [$F(1,20)=.5.4$, $p=.03$, $MSE=572.7$]. The same pattern can be observed when we consider the relative frequency of epistemic moves: Whereas the use of epistemic moves for Group 3 decreases over time, members of Group 4 increase their use of epistemic moves in the second half of the experiment. This suggests that exploratory mode was used opportunistically to meet the instructional constraints, i.e., for online-planning (acting-to-plan) on the current trial, rather than as a prospective investment into learning (acting-to-learn).

- Were epistemic moves used because they were available, i.e., was the usage of exploratory mode simply a task-demand like artifact, prompted by our device manipulation? The strategic use of epistemic moves observed by Group 4 attenuates this concern. Rather than being a mere device affordance, exploratory mode was used selectively to achieve online planning.

Discussion

Participants in the exploratory mode conditions spontaneously and effectively used the device-support to achieve a performance criterion, and in so doing they demonstrated capability for using epistemic actions to improve immediate performance.

However, both the observed unwillingness to invest additional moves in early trials and the selectivity of usage patterns suggest that participants were only willing to invest epistemic moves when they stood to gain an immedate benefit from so doing. There was no clear sign of increased learning through use of exploratory mode or willingness to use epistemic moves for learning purposes. Instead, the selective use suggests that epistemic actions were mainly serving the function of look-ahead (acting-to-plan) rather than learning prospectively about the ToH task (acting-to-learn).

Our instructional manipulation did not have the predicted effect. This may have been due to an unfortunate mismatch between the experimental groups, or it may be that our initial hypotheses about an interesting distinction between problem-solving and learning orientations are unfounded, at least for Tower of Hanoi. Perhaps more likely still is the possibility that our instructional manipulation was too subtle to invoke any change in orientation.

In Kirsh's writing on epistemic actions and related themes, which was one of the sources of inspiration for the current study, an additional concept is introduced by contrast with goal-directed behaviour, namely "complementary strategies" (Kirsh, 1995, 1996). It is not clear to us how precise a distinction Kirsh is promoting between "complementary" and "epistemic": indeed there is a hint in his writings of mere terminological evolution. Nevertheless, we suggest that there is an important distinction that might be sketched. As defined above, epistemic actions have their effect by modifying cognitive structures in the actor. By contrast, consider such example complementary strategies as moving coins in order to count them, or marking numbers in order to add them (Kirsh, 1995, 1996; Neth & Payne, 2001). Such operations work by modifying the problem so as to be more compatible with cognitive capabilities, rather than changing the cognitive state of the actor. We agree with Kirsh that complementary strategies of this kind are ubiquitous in human behaviour.

The case for ubiquity is less clear for epistemic actions. In this article we have sketched a distinction between two kinds of epistemic actions, actions-to-learn and actions-to-plan. We have found evidence for the latter, but none for the former.

One reason that acting-to-learn may be relatively less common than complementary strategies and than acting-to-plan, is the success of learning-by-doing. A second reason, ironically, is that acting may sometimes compete with learning. As shown by O'Hara and Payne (1998), and Trudel and Payne (1995), internalising problem solving activity and planning (doing more mental look-ahead or reflection) can itself increase learning in a problem solving context. For example, when exploratory learners had their interactions with a digital watch rationed, they learned more successfully how to use the watch (Trudel & Payne, 1995).

Despite these arguments, we are confident that, as defined in the introduction, actions-to-learn (i.e. actions that are *not* intended to solve the current problem but only to learn about the current problem) are indeed an important aspect of problem solving and learning. However, such actions may be less widely and spontaneously available and harder to study in conventional puzzle-solving domains.

Turning from the philosophical to the practical, one very concrete contribution of the current article is the idea of incorporating an exploratory mode, with instant undo, into the user interface. Undo functions are, of course, well-established and universally acknowledged contributors to device usability (although some thorny technical design issues are still debated). What is novel about our exploratory mode, we believe, is that it guarantees a very rapid return to particular user-chosen system states. It can accomplish this because the user makes a specific *commitment* to undoing subsequent actions. Although this might seem counter-indicated in mundane HCI contexts, we suggest related designs may be worth pursuing in any domain where people stand to benefit from "thinking by doing".

Acknowledgments

We would like to thank Will Reader for helpful comments and suggestions on an earlier draft. This research was supported by ESRC Research Studentship Award No. R00429934325 to HN.

References

Anzai, Y., & Simon, H.A. (1979). The theory of learning by doing. *Psychological Review*, *86*, 124–140.

Kirsh D. (1995). Complementary Strategies: Why we use our hands when we think. In J.D. Moore & J.F. Lehman (Eds.), *Proceedings of the Seventeenth Annual Conference of the Cognitive Science Society*. Hillsdale, NJ: Lawrence Erlbaum.

Kirsh, D. (1996). Adapting the environment instead of oneself. *Adaptive Behavior*, *4*, 415–452.

Kirsh, D., & Maglio, P. (1994). On distinguishing epistemic from pragmatic action. *Cognitive Science, 18*, 513–549.

Klahr, D., Langley, P. & Neches, R. (1987). *Production System Models of Learning and Development*. Cambridge, MA: MIT Press.

Neth, H. & Payne, S.J. (2001). Addition as interactive problem solving. In J.D. Moore, & K. Stenning (Eds.), *Proceedings of the Twenty-third Annual Conference of the Cognitive Science Society* (pp. 698–703). Mahwah, NJ: Lawrence Erlbaum.

O'Hara, K.P., & Payne, S.J. (1998). The effects of operator implementation cost on planfulness of problem solving and learning. *Cognitive Psychology*, *35*, 34–70.

Trudel, C.I., & Payne, S.J. (1995). Reflection and goal management in exploratory learning. *International Journal of Human-Computer Studies*, *42*, 307–339.

VanLehn, K. (1991). Rule acquisition events in the discovery of problem-solving strategies. *Cognitive Science, 15*, 1–47.