Department of Psychology
University of Strathclyde

Name:	Roy Hunter
Class:	36211 Perception & Cognition
Title of Coursework:	Concepts of Science Practical: The difference a scientific education makes in abstract and concrete problem solving.
Deadline:	Wednesday 14th April.

I affirm that this report is my own work, and does not include any unacknowledged material taken from another source.

Signed: Date:

Abstract

Research has shown that we have an intuitive grasp of mechanics which is not always in line with Newton's laws. McCloskey's (1983) 'naïve physics' model, diSessa's (1988) 'knowledge in pieces' model and Breedin and Cooke's (1994) 'constructed on the fly' model all contest the nature of this intuitive physics. Using Kaiser *et al*'s (1986) distinction between abstract and concrete problems, experiments using produced (drawn on paper) answers to analogous abstract and concrete problems tested the overall scores of physics-educated against non-physics-educated; concrete problems against abstract problems; and the significance of the order of presentation of abstract and concrete problems for the non-physics-educated. Participants with physics education scored higher than those without, the score for concrete problems was higher than for abstract, and there was no difference in score for order of presentation. This supports both diSessa's and Breedin and Cooke's models, but not McCloskey's, whilst paying heed to the criticisms raised by Ranney (1994).

Many of the fundamental laws of physics run counter to human intuition. Before Newton described his laws of mechanics in his *Philosophiae Naturalis Principia Mathematica* of 1687 (Hawking, 1988), it was widely accepted that a body naturally stays at rest until acted upon by an impetus force, whereupon it would move until the impetus force 'ran out' (McCloskey, 1983).

This misapprehension is entirely understandable given that the world is abundant in both friction and gravity, both of which are 'invisible' forces, but when Newtonian mechanics is applied to almost any problem it becomes apparent that Newton was correct. Friction is diminished by moving the body across a smoother surface, or enhanced by moving the body over a rougher surface, and the force of gravity is a constant 9.81 Newtons at sea level (NASA, 2004).

Michael McCloskey (1983) argued that, in the absence of formal scientific education, people form their own theory to rationalise their observations of things such as bodies, rest and motion. The impetus theory alluded to above is almost certainly such a theory, since it is not supported by empirical observation[1]. He called this phenomenon 'Intuitive Physics'.

McCloskey's experiments asked participants to predict the path of a ball rolling out of the end of a curved tube, or to drop a ball onto a target whilst walking. He found that many participants predicted that the ball would continue to roll out of the tube in a curved path, or that they should drop the ball directly over the target, contrary to Newtonian mechanics, but in line with their intuitive conclusions.

McCloskey's conclusions were that a single mental model, based on the impetus theory, was responsible for these misconceptions.

Andrea diSessa (1988) disputed McCloskey's argument, saying that instead of one theory, people would be more likely to have many fragmented ideas about

---

[1] It is worth noting that Newton's laws of mechanics have been observed to break down under empirical observation at both the quantum level, where a particle can exhibit the properties of being in two places at once; and the cosmological level, where matter and energy become interchangeable as a function of the speed of light (Bodanis, 2000), so even Newton's theories are not definitively correct.

how the world works, rather than one over-arching grand theory. He called this 'Knowledge in Pieces'. People might predict an unrealistic outcome to one problem, but come up with the correct conclusion to a very similar one.

diSessa used the example of apparently conflicting interpretations of dynamics, such as the difference between a car being pushed along by the continuous force of its engine and a ball being thrown from a hand in free, unsupported flight. He called these 'phenomenological primitives', or p-prims for short. diSessa's argument was that these p-prims were the basis of our pieces of knowledge, and that even though they may appear to contradict each other, they are the basis upon which a formal education in physics would be built.

The difference between McCloskey's and diSessa's conclusions has been characterised as the 'theory theory' versus 'no theory theory' argument.

Mary Kaiser *et al* (1986) attempted to probe this dilemma further by drawing a distinction between abstract and concrete problems in intuitive physics. An abstract problem like rolling a ball down a curved tube is something most people will never have actually done; a concrete problem like spraying water out of a coiled-up hosepipe is something which most people will have done or seen, and they will therefore be aware of the outcome. The important difference between them is that solutions to concrete problems can be learned from real life, but abstract ones must be deduced.

Kaiser *et al* found that, although the majority of participants answered both sets of problems correctly, there was a significant difference between the high number of correct solutions to the concrete problem coupled with a wrong answer to the abstract problem; and the lower number of correct solutions to the abstract problem coupled with a wrong answer to the concrete problem. The results from Kaiser *et al* suggested that we do think about problems in different ways, although their abstract and concrete problems were not analogous to each other. This casts doubt on McCloskey's 'theory theory', since a single, overarching theory model would be able to solve both abstract and concrete problems equally well.

Participants explained away the difference between their solutions to abstract and concrete problems using concepts like the length of time the ball has to ‘pick up curvature’, or the greater pressure of the water in the hose, or the physical differences between liquid water and a solid ball, all of which are mechanically unsound concepts for this context.

There was no significant difference for the order in which the problems were presented by Kaiser *et al*. This may have similar implications to those above for McCloskey’s theory; that the single theory model does not encompass learning about the abstract problem from solving the concrete problem, where such learning might be expected.

Cooke and Breedin (1994) found that with the right cues, the right answer could be elicited from participants who had little to no knowledge of physics. Left to formulate a response themselves, participants produced answers based on naïve theories, but given a multiple choice they can select the correct response with significantly greater accuracy. Cooke and Breedin concluded from this study that people’s judgements of dynamics are ‘constructed on the fly’ rather than being the result of overall theories or task-specific fragments.

There are methodological concerns with Cooke and Breedin’s experiments, since comparing produced answers with selected answers gives more scope for errors on the produced answers side than the selected. Although they tried to minimise the risk of such errors, they cannot be entirely dismissed. Despite the methodological aspects of their work, the theory is worth further investigation.

Ranney (1994) criticised most of the above for the polysemous nature of the language they use: no-one ties down a definition of what a theory is, what impetus is supposed to be, or how one misconception of dynamics is supposed to be more accurate than another misconception of dynamics.

Specifically, Ranney criticised the disparity between experimental materials in the different studies, and the assumption that all the experimental participants have the same version of impetus theory as their naïve model of mechanics. He argued that not only might one person have more than one theory of how things move, many different people may have many different impetus theories, and that the assumption of a single model of naïve physics interferes in the experimental process by compromising the null hypotheses of psychology experiments concerning naïve and educated physics.

If Ranney's concerns are to be taken seriously, then experimenters must be careful to avoid a null hypothesis involving a single model of naïve physics, experimental materials and methods must be consistent for both concrete and abstract problems, and the hypotheses and reporting must be worded in such a way that the theories being examined must be explicit.

Bearing Ranney's criticisms in mind, this experiment will test three hypotheses:

1. Participants with physics education will perform better in solving dynamics problems than participants without physics education.

2. Participants will perform better in concrete dynamics problem solving than in abstract dynamics problem solving.

3. Non-physics educated participants solving dynamics problems will exhibit a difference in their performance between those solving the abstract problems first and those solving the concrete problems first.

Method.

Design.

The three hypotheses were tested from one set of experimental data. Data concerning the participants' gender and their education in physics was collected, and three concrete problems and three abstract problems relating to gravity and forward motion were prepared.

Half of the participants performed the concrete tasks first, and half performed the abstract tasks first on a randomly selected basis.

Each of the concrete problems is analogous to one of the abstract problems. The three concrete problems described were to:

- Draw the path of a jet of water coming from a coiled hose;
- Draw the path of water, a canoe and a rock falling over a waterfall;
- Draw the path of a package dropped from an aeroplane to the ground.

The three abstract tasks described were to:

- Draw the path of a ball rolling out of the end of a coiled tube;
- Draw the path of a ball falling from an edge towards a surface;
- Draw the path of a ball falling from an elevated track to the ground.

The water and hose problem and the ball and tube problem are analogous; as are the waterfall and edge problems; and the aeroplane and track problems.

Participants were asked to draw the paths onto the questionnaire, and in the case of the second and third tasks in each category they also had to provide a landing site. There were sufficient visual and mathematical clues in the questionnaire to enable participants who remembered their physics to answer the questions accurately.

The analogous concrete and abstract problems were marked using the same criteria, and these marks were used as the basis of the analysis. An equal weight was attached to both types of problem.

Participants.

Approximately 120 psychology undergraduates from the University of Strathclyde participated as a requirement of their course. The participants' knowledge of physics was not known before the experiment, and they were not selected according to any other criteria. Their gender distribution is assumed to equal, although the sample for testing displays a slight female bias. The participants were aged between eighteen and approximately fifty.

Materials.

All participants completed a paper questionnaire (see appendix A). The front page of the document asked participants to fill in details of all their previous educational achievements, from which physics was selected, and their gender. Half of the questionnaires displayed the abstract problems first, and half displayed the concrete problems first. The problems were presented one on each page, so that possible 'interference' from other problems would not occur.

Procedure.

The experiment took place over six one-hour sessions in the late afternoon of a normal working day. The sessions were held in a computer laboratory at the University. Participants were given the questionnaire, which asked them to discuss the problems as a group for about fifteen minutes, but also told them that they did not have to reach a consensus before filling in their answers individually.

The group discussion was very informal and was not timed, moderated or controlled, and participants filled in the questionnaires at their leisure before handing them back to the experiment leader.

When the questionnaires were scored, one point was awarded for accuracy in predicting the path of the moving or falling object in each task, and a further point was awarded for accuracy in the object’s landing point in the second and third tasks in each category. This means that a maximum of five points were available for correct answers to the concrete problems; and a maximum of five points were available for correct answers to the abstract problems.

Experiment 1.

Method

The first hypothesis is one-tailed and it uses an independent samples design. The independent variable is the physics education or otherwise of the participant, and the dependent variable is their overall score. The data collected is on a ratio scale.

Results.

The mean score taken from a random sample of 30 participants (see appendix B) with physics education was 7.33 with a standard deviation of 1.345. The mean score for participants without physics education was lower, at 3.67 with a standard deviation of 1.397. The scores, illustrated in figure1, were normally distributed and on a ratio scale.

An independent samples t-test returned a value of 7.322 with 28 degrees of freedom, which is significant at the 1% level ($t(28)=7.322$, $p<0.01$). The calculations were performed using SPSS for Windows v11.5 (see appendix C).

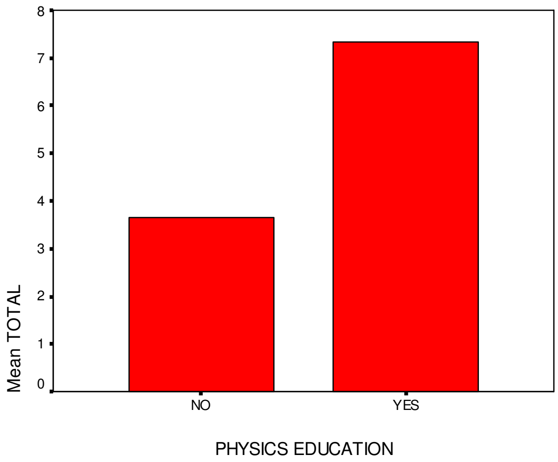

Figure 1.

Mean scores for participants with physics education and without physics education for all tasks.

Discussion.

The results of this experiment support the first hypothesis; that participants with physics education would perform better than participants without physics education. The scores of participants with a physics education were significantly higher than the scores of participants without physics education.

This supports all the various theories of intuitive physics, because none of these studies argue that intuitive processes are any better or more accurate than the more formal theories. As the hypothesis is concerned only with educated physics, it does not make any pre-judgement about the nature of naïve physics as noted by Ranney (1994).

Although this does not tell us much about the arguments within the field of naïve physics, it is useful to know that the experiment used in this study has produced an appropriate result for the most obvious distinction in the field: that educated participants out-perform those reliant on their intuition.

Experiment 2.

Method.

The second hypothesis is one-tailed and it uses a related samples design. The independent variable is the type of problem: concrete or abstract, and the dependent variable is the score for that type of problem. The data collected is on a ratio scale.

Results.

The mean score taken from a random sample of 30 participants' (see appendix B) performance in abstract tasks was 2.63 with a standard deviation of 1.245. The mean score for all participants' performance in concrete tasks was higher at 2.87 with a standard deviation of 1.306. The scores, illustrated in figure 2, are normally distributed and are on a ratio scale.

A related samples t-test returned a value of 1.157 with 29 degrees of freedom, which is significant at the 1% level (*t*(29)=1.157, *p*<0.01). The calculations were performed using SPSS for Windows v11.5 (see appendix C).

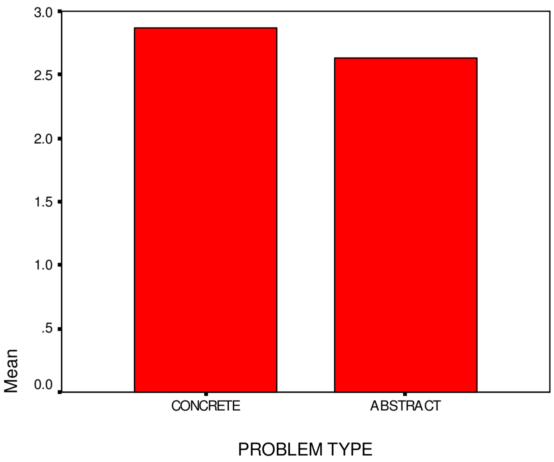

Figure 2.
Mean scores for all participants in abstract problems and concrete problems.

Discussion.

The results of this experiment support the second hypothesis; that participants would perform better in concrete problem solving than abstract problem solving. The scores for concrete problems were significantly higher than the scores for abstract problems. This finding is consistent with diSessa's (1988) arguments, as well as Kaiser *et al* (1986) and Cooke and Breedin (1994), but not McCloskey (1983).

The results are inconsistent with McCloskey's (1983) arguments, at least as far as diSessa's criticisms are concerned, because a unified theory of naïve mechanics should produce consistent results for both types of tasks, concrete and abstract.

The significant difference between abstract and concrete task results tends to support diSessa's 'knowledge in pieces' model, since to arrive at two different conclusions for what is essentially the same problem would indicate two different methods of arriving at that conclusion.

The results from Kaiser *et al* (1986) are also supported. It could be argued that since Kaiser *et al* used actual physical apparatus, and this experiment used drawings of apparatus, the abstract problems aspect of this experiment was better represented than in Kaiser's experiment.

To actually build the experimental apparatus may interfere with its abstract nature. The non-physical nature of the concrete problem apparatus should not be so much of a problem since the distinction between the two states – abstract and concrete - is related to the ability to draw from real-life experience. One can relate real-life experience to a two-dimensional drawing almost as easily as one can relate it to a three-dimensional object, but to give an abstract problem physical form may give too many clues away about the probable outcome.

Cooke and Breedin's (1994) 'calculated on the fly' model is supported in so far as it is not contradicted by this result, but this experiment fails to draw a distinction between their model and diSessa's.

As with hypothesis 1, the hypothesis does not make any judgement about the nature of naïve physics, and also the materials used were consistent for all tasks: the three abstract tasks were asking the same underlying questions as the three concrete tasks. The potential criticisms raised by Ranney (1994) have been to some extent addressed.

Experiment 3.

Method.

The third hypothesis is two-tailed and it uses an independent samples design. The independent variable is the order of presentation of the problems: abstract or concrete first; and the dependent variable is the overall score. The data collected is on a ratio scale.

Results.

The results shown below are taken from a random sample of 30 participants' scores (see appendix B).

The mean score taken from a random sample of 30 participants (see appendix B) who solved the concrete problems first was 1.88 with a standard deviation of 0.835. The mean score for participants who solved the abstract problems first was 1.57 with a standard deviation of 0.535. The scores are illustrated in figure 4. The scores, illustrated in figure 3, are normally distributed and are on a ratio scale.

An independent samples t-test returned a result of -0.824 with 13 degrees of freedom, which is not significant at the 5% level ($t(13)=-0.824$, $p>0.05$). The calculations were performed using SPSS for Windows v11.5 (see appendix C).

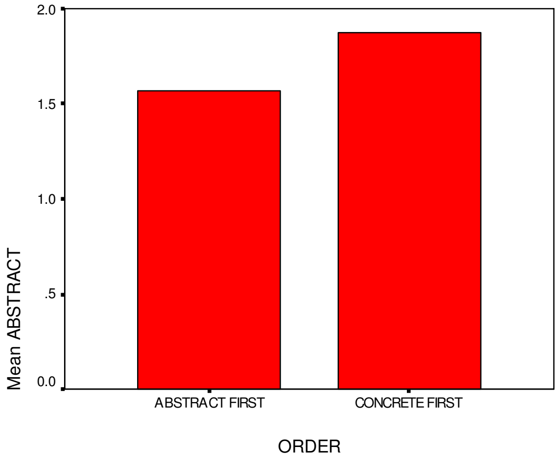

Figure 3.

Mean scores for non-physics educated participants for abstract problems by order of presentation.

Discussion.

The results of this experiment do not support the third alternative hypothesis, that there will be a difference between the performance of non-physics educated participants based on presentation order , so the null hypothesis, that there is no difference in performance due to presentation order, must be accepted.

This supports the findings of Kaiser *et al* (1986) as regards problem order. The implication of this result is that participants did not learn anything about the outcome of the abstract problems from performing the concrete problems.

As in the previous experiment, Breedin and Cooke's (1994) findings are not contradicted, but not sufficiently distinguished from those of diSessa (1988) to be supported. In order to support Breedin and Cooke, assumptions about the inevitability of learning from experience, and assumptions about the nature of naïve physics would have to be made. This is something that Ranney (1994) specifically warned against.

General Discussion

The results of this study take issue with McCloskey's (1983) model of intuitive physics. If we had a single intuitive model of dynamics, then it should produce similar performance levels for both abstract and concrete problems. The different levels of performance for these two types of task suggest that we approach both types of task in a different way as shown in the second experiment, and as concluded by Kaiser *et al* (1986).

Neither that experiment or the Kaiser *et al* (1986) study tells us much about the nature of these different ways of modelling dynamics, only that problems rooted in real-life experience are solved more readily than problems outwith our everyday experience. The obvious conclusion to be drawn from this is that we learn from our experiences.

The findings of the third experiment, along with the findings of Cooke and Breedin (1994), dispute that learning conclusion to some extent. If we learn from our everyday experiences, then we should also have learned from performing concrete problems (for which the score was significantly higher: experiment 2) when it comes to performing abstract problems.

This study does not draw any conclusions about the distinction between Cooke and Breedin's (1994) model and diSessa's (1988) model, but the experiment could be adapted to try and do so. A longitudinal study of abstract problem solving by non-physics educated participants could establish whether they keep making the same mistakes or not. If the same participants repeatedly make the same mistakes on the same problems, that would tend to support diSessa's (1988) argument that we have a concrete set of preconceptions about individual dynamic activities.

If participants did not make consistent mistakes, answering the same problems in different erroneous ways over the course of the study, that would tend to support Cooke and Breedin's (1994) argument that we theorise as we go and have no set pattern of naïve physics.

This study has examined the concept of naïve physics: the assumption or set of assumptions that people without a formal physics education use to try and predict the path of objects under the influence of forces like gravity. This study has not supported McCloskey's (1983) model of naïve physics, but by using a very similar set of experiments to Kaiser *et al* (1986): the distinction between abstract and concrete problems, has backed up the findings of diSessa (1988) and to some extent Cooke and Breedin (1994). The study took careful heed of the methodological concerns in this field raised by Ranney (1994) and has strived to avoid such errors.

The results of the first experiment support the idea that an education in physics is still the best way to solve dynamics problems accurately. The results of the second experiment support the 'plurality of theory' ideas, although they make no judgement about which model; diSessa's (1988) or Cooke and Breedin's (1994); is correct. The results of the third experiment tentatively support Cooke and Breedin's 'calculated on the fly' model, but to accept that conclusion would be to fall into the null hypothesis trap described by Ranney (1994). Further study into the consistency of mistakes made by naïve physicists may clarify the issue.

References

Bodanis, David (2000), *E=mc^2: A Biography of the World's Most Famous Equation,* Macmillan, London.

Cooke, Nancy J.; Breedin, Sarah D. (1994), Constructing naïve theories of motion on the fly, *Memory and Cognition* vol. 22 (4), p.474-493, Psychonomic Society, Austin TX.

diSessa, Andrea A. (1988), Knowledge in Pieces, in Forman, George & Pufall, Peter (eds.) *Constructivism in the Computer Age*, Laurence Erlbaum Associates, Hillsdale NJ.

Hawking, Stephen W. (1988), *A Brief History of Time,* Transworld, London.

Kaiser, Mary Kister; Jonides, John; Alexander, Joanne (1986), Intuitive Reasoning about abstract and familiar physics problems, *Memory and Cognition* vol. 14 (4), p.308-312, Psychonomic Society, Austin TX.

McCloskey, Michael (1983), Intuitive Physics, *Scientific American,* vol. 248 (4), p.114-122, Scientific American, New York NY.

National Aeronautics and Space Administration (NASA) (2004), *Space Station Biological Research Project* (accessed 2.03.04), http://brp.arc.nasa.gov/Science/Y_GBL/bsc_resrch.html.

Ranney, Michael (1994), Relative consistency and subjects' "theories" in domains such as naïve physics: Common research difficulties illustrated by Cooke and Breedin, *Memory and Cognition,* vol. 22 (4), p.494-502, Psychonomic Society, Austin TX.

Appendices

Appendix A: Questionnaire.

Appendix B: Scores and other criteria for participant sample.

Appendix C: SPSS calculation printout.

Appendix B: Participants’ scores and other data.

Participant	Physics education	Gender	Question order	Abstract score	Concrete score	Total score
1	P	F	A before C	5	4	9
2	P	M	A before C	4	4	8
3	P	M	A before C	3	4	7
4	P	F	A before C	3	3	6
5	P	M	A before C	5	4	9
6	P	M	C before A	3	3	6
7	P	F	A before C	4	3	7
8	P	F	A before C	3	4	7
9	P	F	A before C	2	3	5
10	P	F	C before A	3	5	8
11	P	M	C before A	2	4	6
12	P	M	C before A	5	4	9
13	P	M	C before A	4	4	8
14	P	F	C before A	3	3	6
15	P	F	A before C	4	5	9
16	NP	F	A before C	1	2	3
17	NP	F	A before C	2	3	5
18	NP	M	A before C	2	0	2
19	NP	M	A before C	2	4	6
20	NP	F	A before C	1	2	3
21	NP	F	A before C	2	1	3
22	NP	F	A before C	1	2	3
23	NP	M	C before A	2	2	4
24	NP	M	C before A	3	2	5
25	NP	F	C before A	2	2	4
26	NP	M	C before A	3	2	5
27	NP	F	C before A	1	0	1
28	NP	M	C before A	1	1	2
29	NP	F	C before A	2	3	5
30	NP	F	C before A	1	3	4

P: Physics education

NP: No Physics education

A before C: Abstract problems presented before Concrete problems

C before A: Concrete problems presented before Abstract problems