The Development of the Action Concept Inventory

Lachlan McGinness

A thesis submitted for the degree of
Bachelor of Science with Honours in Physics of
The Australian National University

October, 2014

Declaration

This thesis is an account of research undertaken between February 2014 and October 2014 at The Research School of Physics, The Australian National University, Canberra, Australia.

Except where acknowledged in the customary manner, the material presented in this thesis is, to the best of my knowledge, original and has not been submitted in whole or part for a degree in any university.

Lachlan McGinness
October, 2014

Acknowledgements

Firstly I would like to thank my supervisor Craig Savage, and John Aslanides, who developed the Relativity Concept Inventory here at the ANU two years ago. Without them laying the foundation, the basic methods of inventory development and analysis, it would not have been possible for me explore inventory development and statistics to the level I did in this thesis.

I would like to thank all the Action Experts who helped me along the way, especially our local action experts, Joe Hope, John Close, Bob Dewar, David Williams and Ben Buchler. A special thanks to the international experts Edwin Taylor, Chris Gray, Slavomir Tuleja and Eric Poisson who reviewed the inventory and suggested changes for its improvement. I would especially like to thank Chris Gray for helping me understand the generalised Hamilton and Maupertuis Action principles.

I would like to thank all my friends who have supported me, by listening to me discuss the project. A special thanks to Katrina Rivera, Aziah Williamson, Joe Hope, Craig Savage, Peter Smith and Joseph McGinness who took time to read my work and make suggestions.

Finally I would like to thank all the physicists at the ANU who have contributed to giving me the understanding of physics that I have today. Two years ago if you had asked me what I liked about physics I would have told you that I loved how you could apply it to the world around us and make sense of everyday examples. I never would have anticipated that I would be writing a thesis about one of the most theoretical branches of physics, Action physics. Through a joint effort you have worked together to bring me to a point where I can understand this theory. Even more I realise that if you choose to believe Action physics, it radically undermines the way you view the universe and that it bridges everything in physics together.

Abstract

An Action Concept Inventory was created to measure student understanding of the principles that underpin the principle of stationary action. International Action experts were consulted to create a list of 10 concepts that underpin the principle of least action.

This list was then used to create the action concept inventory, an eighteen question multiple choice conceptual test which measures student understanding of the principle of least action. 77 ANU first year students studying action physics took the Action Concept Inventory as both pre-test and post-test. 61 physics honours students, PhD students and academics from the ANU also completed the action concept inventory.

This inventory was validated through student interviews, expert review and using statistical analysis. In addition to one on one think aloud interviews, students were interviewed in pairs, this made the interview process more comfortable for the students and decreased the frequency with which students misinterpreted questions.

It was found that 15 of the 18 questions were valid measures of students' understanding of action concepts, while three of the questions needed to be revised before the inventory is published. A number of 'Explore All Paths' misconceptions were discovered and as a result it is recommended that one more question is added to the inventory to probe understanding of this concept.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Physics education is extremely important for Australia's future [1]. Introductory physics education has enormous potential to provide students (and therefore the Australian public) with analytical and statistical skills and knowledge about the scientific process. In this thesis I discuss a new approach to physics education called 'Action Physics'.

Physics has been taught a certain way for more than 60 years, mainly for historical reasons [2]. The introductory physics curriculum and textbooks mainly focus on teaching students Newton's laws, electromagnetism and solving algebraic equations. Force is emphasized, while principles that are more useful to a wider range of fields outside of physics, such as the scientific method, statistics and even energy are underemphasised and sometimes completely forgotten.

Physics education studies have shown that traditional methods of teaching are not effective and in response there has been a movement towards interactive learning (see [3] [4] and references therein). Many physics students have benefited from the revolution about 'how we teach'. However there has yet to be a revolution in 'what we teach'. Physics education researchers have not answered the question:

> *Should we continue teaching introductory physics through Newton's laws or is there a better way?*

1.1 A Call to Action

In 2003 an MIT professor, Edwin Taylor, published a paper "A Call to Action", where he outlined Action Physics as a new way of teaching introductory physics [5] [6] and encouraged other physics instructors to teach Action to their students. Action Physics is based on two scalar quantities, energy and Action.

Before we can evaluate whether teaching physics through Action is more effective than teaching physics through Newton's laws, the most effective ways of teaching Action needed to be established (for a true comparison). Many physicists have published papers about the resources and techniques they developed to teach Action Physics [5, 7, 8, 6, 9] (see also references therein) and they claim these resources are effective. However there is no formal instrument for measuring student understanding of Action, therefore it is impossible for instructors to have a objective universal evaluation of their instruments and teaching.

In order to find the most effective ways to teach Action we need a way of measuring student understanding of Action. Once we have a measurement device we can use this to determine the best way of teaching Action. In this thesis I discuss the development and validation of the Action Concept Inventory (ACI), the measurement tool that physics education researchers will use to measure students' understanding of Action and evaluate their teaching methods.

1.2 Teaching Physics Through Action

Action physics is based around the Principle of Stationary Action:

The classical (observed) path is the one for which the Action is stationary

Where the Action (more specifcally Hamilton's Action) is defined as:

$$S = \int_A^B L\,dt \tag{1.1}$$

Where S is the Action, A and B are the start and end events, L is the Lagrangian and t is time. For most simple classical mechanics systems the Lagrangian is given by $L = K - V$ where K is the kinetic energy, V is the potential energy.

The Principle of Stationary Action is a bridging concept that creates the link between classical mechanics, quantum mechanics and special relativity [5, 8]. From the principle of stationary Action it is possible to derive Lagrangian mechanics (the Euler-Lagrange equation) an alternate formulation of classical mechanics that is equivalent to Newton's Laws [10, 11, 12].

Historically, Action physics was mainly taught in the context of Lagrangian mechanics. It was reserved for later year or graduate students because of the complexity of the maths involved. However with new mathematical software like Mathematica, Matlab and Maple, the ability to do complex mathematics is no longer necessary to understand Action concepts. It is now possible to show students the concepts behind the Principle of Stationary Action and use software to tackle the difficult mathematics. This allows them to simultaneously develop an understanding of programming, a widely applicable skill in today's technological society.

Introductory physics is no longer only for physicists [13]. There are a wide variety of people who need physics knowledge, scientists from other disciplines, politicians and decision makers need to have a solid understanding to make choices about science funding and other contemporary issues such as renewable energy. Action physics is based on energy rather than force, a more important concept in other areas of science especially chemistry and biology [14]. Teaching energy gives the instructor the opportunity to focus on contemporary issues such as renewable sources of energy.

For students continuing on to become physics researchers, energy and Action will be far more important than force and Newtons laws. Action is the link which brings quantum mechanics, general relativity and classical mechanics together as one cohesive picture [5]. Action also emphasises constraints and symmetry which are some of the most central ideas in physics. Fundamental physics is formulated in terms of Lagrangians, not forces [15], and Action is used in other scientific research [16, 17, 18, 11] including biology [19], chemistry [20] and areas of physics as diverse as quantum mechanics [21, 22, 23] and fluid flows [24].

1.3 Measurement in physics education

Like all areas of physics, one of the biggest problems in physics education is measurement. A conceptual test to measure student understanding should have the same properties as any other measuring device in physics, for example a ruler to measure length. It should be valid, reliable, accessible and preferably easy to use.

Physics education researchers have developed measurement techniques which display all of these properties [25], frequently called concept inventories. Concept inventories are multiple choice tests which measure students' conceptual understanding of a given topic. In 1992 Hestenes, Wells and Swackhamer published the first concept inventory, the Force Concept Inventory [26]. The Force Concept Inventory (FCI) is a 29 question multiple choice test which probes students' understanding of key concepts in Newtonian physics.

The Force Concept Inventory was very successful and the original paper has been cited over 300 times. Concept inventories are now accepted by the physics education community as the standard method of measuring student understanding. The process for developing concept inventories has progressed a long way since the Force Concept Inventory was created in 1992 and there are now standard practices that allow reliable inventories to be developed [25].

1.4 Concept Inventory Development

In 2010, physics Nobel Laureate Carl Wieman and Wendy Adams reviewed the concept inventory literature and outlined the most effective methods for creating concept inventories [25]. A summary of these steps is shown in Figure 1.1 specifically outlining the development of the Action Concept Inventory.
The first stages establish topics that are important to teachers through the creation of a concept list. In this project the development of the concept list is extremely important because the key Action concepts have never been identified in the literature. This concept list will determine what concepts need to be tested and can be used by future Action instructors as a guide of topics to focus on.

The draft concept list was initially established by looking through the Action physics literature and identifying the concepts that occurred most regularly. My supervisor and I had many discussion to accurately define the concepts and establish which are most important. ANU Action experts, academics who teach Action or use it in their research, were consulted for their opinions on the draft concept list. Academics who had published papers about Action Physics or produced a resource to teach Action Physics were identified as international Action experts and were consulted to ensure the validity of the concept list.

The concept list was then used to create draft concept inventory questions. Questions were designed to probe student thinking about Action concepts. Students and Action experts were interviewed, following a strict 'think aloud' protocol (see section 3.5.2) to establish expert thinking and identify student misconceptions.

The questions were transferred into multiple choice format. Misconceptions identified from student interviews were used to create (incorrect) distracter options [25]. Think aloud interviews were used again to probe both student and expert thinking while attempting the multiple choice test. This ensured correct answers were chosen for sound reasons and distracters were chosen for the non-expert-like thinking that they were intended to probe.

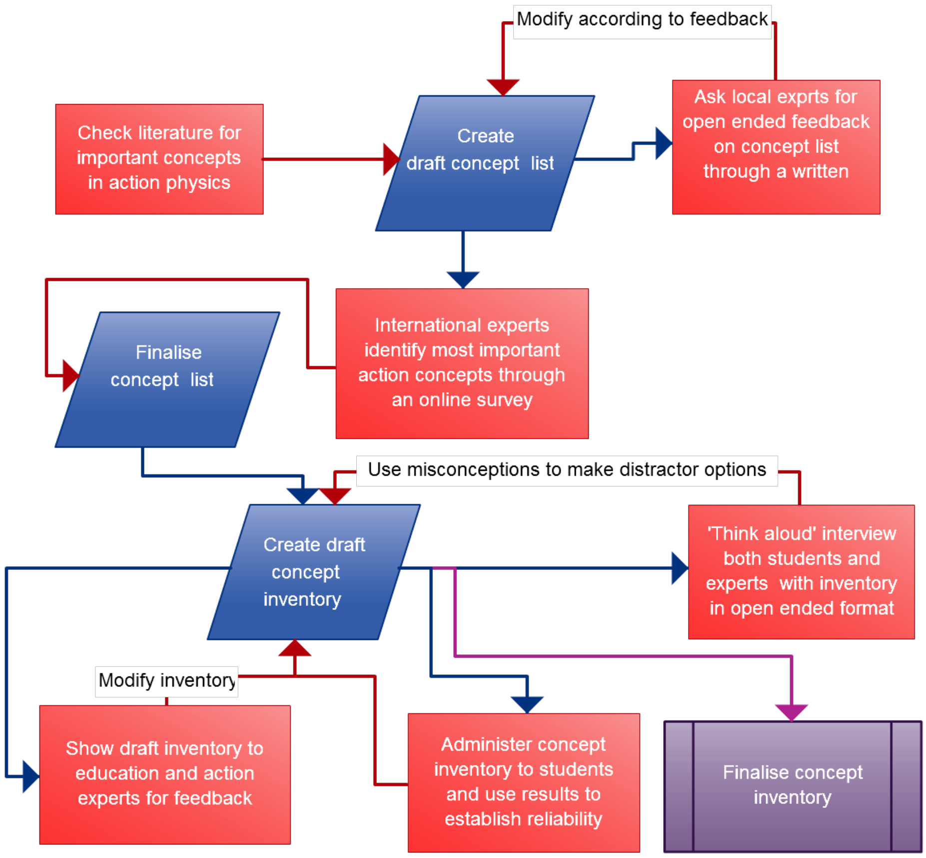


Figure 1.1: Process for developing a concept inventory. Key components of developing a concept inventory are constant consultation with experts and feedback. This ensures that all key concepts are identified and fully developed and that the inventory is valid. Blue steps correspond to creative processes while red correspond to feedback and modification.

The final inventory contained 18 multiple choice questions and took most students $20 - 30$ minutes to complete.

1.5 Statistical Analysis

Once the Action Concept Inventory was developed, its reliability and validity were established by administering the test to the PHYS1201 class, a first year ANU physics class taking a unit on Action. Although a class size of a few hundred would have been desirable [25], the PHYS1201 contained 108 students which was sufficient for the statistical methods I used to establish validity and reliability.

The test was administered to the students both before and after they studied an 8 lecture unit on Action Physics (called pre-test and post-test). When administering the test to students it was essential to obtain as high a response rate as possible and have the students take the test seriously. Pre-test and post-test data was used to carry out reliability and validity checks and item analysis.

Reliability

A reliable ruler would measure the length of a piece of string to be exactly the same every time. Unfortunately unlike measuring a piece of string, it is impractical to measure student understanding over and over again to check the reliability of an inventory.

Usually the reliability of an inventory is established by administering the test to two equivalent populations (usually two consecutive years of the same course) and obtaining a test-retest stability coefficient. Due to time constraints this was not done for the Action concept inventory, instead internal consistency coefficients were calculated.

Validity

A ruler is designed to measure length, however due to thermal expansion most rulers are affected by temperature variations. A valid ruler would measure length alone. A valid Action Concept Inventory should measure students' understanding of Action, not their general intelligence or their ability to guess answers (called test-wiseness see [25] and [27]). Validity is primarily established by consultation with experts and by observing (interviewing) students while they complete the inventory. However validity can be confirmed by checking the correlation between questions testing the same concepts and the correlation between student inventory scores, homework and course exam marks.

Discrimination and Difficulty

There is no point using a micrometer to measure the length of a table. Testing the item difficulty and item discrimination of the inventory ensures that it is the appropriate scale for what it is intended to measure.

1.6 What next?

With the Action Concept Inventory validated and published, instructors can start using it to measure effectiveness of and improve Action instruction. Physics education researchers will be able to answer the question:

What is the most effective way to teach physics through Action?

Once effective methods of teaching Action have been established physics education researchers will be able to make a proper comparison between teaching physics through Action and through Newton's laws. Action Physics may prove to be an effective way to introduce students to physics and link many topics for continuing students.

Thesis Outline

A brief introduction to each of the chapters in this thesis is given below. Rather than having a separate literature review chapter, the relevant literature was integrated into chapters 2, 3 and 4.

- Chapter 2 - *Physics through Action*, is an outline of Action pedagogy. It contains Action resources, teaching methods, applications and derivations that should prove useful to any instructor wishing to teach Action. Chapter 2 also shows how Action can be linked back to other areas of physics.
- Chapter 3 - *Developing The Action Concept Inventory*, describes the procedures used to create the Action Concept Inventory. Chapter 3 also contains an in depth discussion of the concepts tested on the Action Concept inventory and common misconceptions.
- Chapter 4 - *Statistical Analysis*, explores the statistical methods used to evaluate the Action Concept Inventory.
- Chapter 5 - *Results and Discussion*, displays the results of the Action Concept Inventory pre-test and post-test, student homework and the student survey. I discuss the validity and reliability of the inventory, and lessons learned for future Action teaching.
- Chapter 6 - *A Novel Idea: A second Dataset*, briefly describes the results from giving the Action Concept Inventory to higher level students and academics.
- Chapter 7 - *Conclusions*, summarises the main findings from the research.

Chapter 2

Physics Through Action

In 'A Call to Action' Edwin Taylor outlined a new way of teaching introductory physics based on The Principle of Stationary Action [5]. In this paper he speaks about completely revolutionising introductory physics, introducing classical mechanics without focussing on Newton's laws, introducing quantum mechanics without the Schrodinger Equation and relativity without using tensors. Table 2.1 compares teaching physics through Action to teaching physics through the traditional method using Newton's laws:

Table 2.1: Comparison of teaching introductory physics through Newton's laws and teaching introductory physics through Action

	Physics Through Newton's Laws	Physics Through Action
Central Concepts:	Force (vector)	Energy (Scalar) and Action (Scalar)
Mathematics Required:	Algebra Vectors	Algebra Complex Numbers Differential Equations
Main Advantages:	• Links easily to everyday experience • Was discovered first historically • Teaching is already well established (many resources available)	• Relates to contemporary physics research • Connects different areas of physics together • Energy is a contemporary issue and is central in chemistry and biology
Main Disadvantages:	• Not central to modern physics research • Students must completely change their intuition when finally taught Quantum Mechanics or Relativity	• Not many teaching resources available • Only allows fundamental interactions (Deals awkwardly with friction forces)

In this chapter I outline how to teach physics through Action and address the problems posed as disadvantages in Table 2.1. My hope is that this section will be a useful resource for any instructor looking to start teaching Action at an introductory or advanced level. In this section I summarise the literature on introductory Action pedagogy. In addition I show the links that will allow an instructor to link almost any physics topic back to the Action principle.

2.1 Action Resources

To date there are no textbooks devoted entirely to Action or Action principles. However there are some other great resources: books, videos, interactive software and webpages that can be used by an instructor when teaching Action at an introductory level.

To gain an understanding of the principles that underly the many path formulation of quantum mechanics, *Feynman's QED: The strange Theory of Light and Matter* is a brilliant starting point [28]. Two videos [29] and [30] have been created which summarise the early chapters of QED and these can be viewed online at `http://www.youtube.com/watch?v=AfsZiHEcoxk#aid=P-xeitdAXgE` and `http://www.allthingsscience.com/video/752/Quantum-Electrodynamics` respectively. For more advanced students looking to gain a good understanding of how many paths integrals are performed one of the best resources is Feynman and Hibb's *Quantum Mechanics and Path Integrals* edited by Dan F. Styer [31]. Jon Ogborn and Edwin Taylor also wrote notes for any instructors wishing to teach students the link between classical mechanics and quantum mechanics in their paper *'Quantum physics explains Newton's laws of motion'.* Chris Gray wrote a scholarpedia page which provides an elementary yet comprehensive overview of the Principle of Stationary Action [10].

In 1998 Edwin Taylor, Statmis Vokos, John M. O'Meara and Nora S. Thornber wrote a paper 'Teaching Feynman's Sum-Over-Paths Quantum Theory' [32]. This paper outlines a way that computer programs can be used to teach students the Feynman many paths approach to quantum mechanics. There are computer programs that go with the paper which can be downloaded for free and used as computer based tutorial or laboratory exercises, they are available at `www.eftaylor.com/leastaction.html` [6]. Although these programs are written in an old language that is no longer supported (cT), they the most comprehensive set of Action teaching software available.

Edwin Taylor and Slavomir Tuleja also created an online tutorial for introductory Action students [33]. Although less comprehensive than the computer programs mentioned above, it is very easy for instructors and students to access online. In addition to having some simulations of Euler's Method as built in Java applets it also contains text explaining concepts to students and guiding them through the simulation. This simulation can be accessed at `http://www.eftaylor.com/software/ActionApplets/LeastAction.html` [33].

A more modern Action simulation [34] has been created by Todd Timberlake and published including a program which allows students to explore paths of least Action using Euler's method (see Section 2.2.4). It is available online at `http://www.compadre.org/portal/items/detail.cfm?ID=12400` [34].

Thomas Moore explores implications for the higher levels curriculum if Action physics was taught to first year curriculum. In 'Getting the most out of least Action: A proposal' [8] he highlights new opportunities that arise when teaching later year quantum mechanics, classical mechanics and electromagnetism.

Usually whenever a new physics concept is taught, the instructor introduces the history and context which lead to the development of that concept. A number of authors have highlighted the story of how Action has developed [10, 9, 35]. The next Section, *The History of Action* summarises their work, outlining the history and context in which Action was developed so that instructors may use this to introduce Action to their students.

2.2 The History of Action

The sections below outline the history of Action principles. It starts with Fermat's principle (proposed by Fermat in 1658) [36], then looks at the historical Maupertuis Principle (1744) [10] before looking at the Hamilton's Action Principle (1834) [10] (popularised by Feynman as the Action Principle) and finally looking at some of the work done to develop Action principles over the last few decades.

2.2.1 Fermat's Principle

Figure 2.1: Pierre de Fermat. Image from http://en.wikipedia.org/wiki/Pierre_de_Fermat

Pierre de Fermat introduced the principle of least time for light rays (Fermat's principle) in 1658. It states that light takes the path of least time between fixed initial and end points. [9].
During Fermat's lifetime he was unable to answer the question: "How does light know in advance which path will be the quickest?" [9]. This problem was resolved by Christian Huygen's idea that each point on a wavefront can be treated as a source of 'wavelets' that superimpose constructively and destructively. Along the path of stationary time these 'wavelets' interfere constructively [9].

Fermat's principle can be used as a great introduction to variational principles and can be used to explain a broad range of optical phenomenon to students including refraction and reflection. One particularly interesting example is Einstein's mirror, Fermat's principle can be used to show that for a mirror that is moving the angle of incidence is not equal to angle of reflection.

Fermat's principle is still used to design lenses [37] and it can also be used to calculate the motion of light rays in general relativity [38].

2.2.2 Christiaan Huygens and the wavelet principle

In 1690 Christiaan Huygen's developed 'Huygen's Principle' and was the first person to account for the laws of ray optics using a wave theory [39]. Huygen's principle states that the propagation of light can be stated as [40]:

- Any point reached by a wave front may be considered from that instant as a point source of light.
- If a wave front with its forward and rear region is given at one instant $t_{initial}$, then the wave front at the next instant, a time Δt later, is found at the envelope of the elementary wave surfaces after time Δt around all the points of the original wave front which lies on the forward side.

This is illustrated in Fig. 2.3:

Figure 2.2: Christiaan Huygens. Image from `http://en.wikipedia.org/wiki/Christiaan_Huygens`

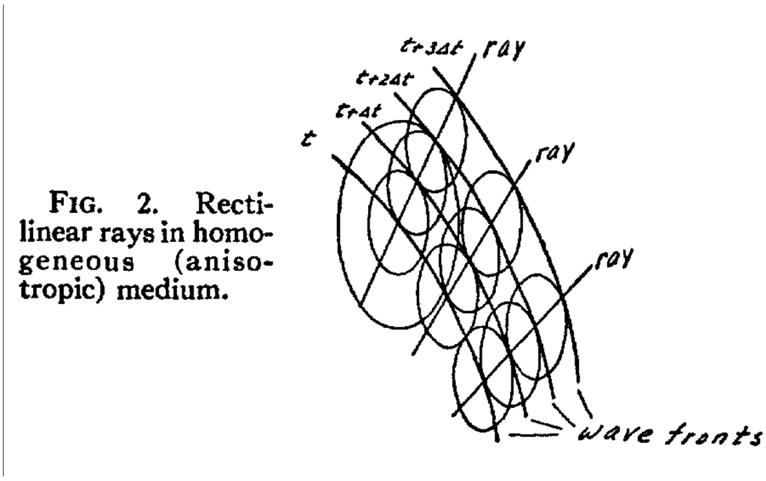


Figure 2.3: Illustration by A. Witte [40] of how Huygen's principle is used to predict an arbitrary wavefront after times of Δt, $2\Delta t$ and $3\Delta t$. Huygens showed that the propogation of a wave could be modelled by treating every point on the wavefront as a point source of 'wavelets'. Richard Feynman used a similar method to describe the motion of small particles [9].

Fermat's principle arises from Huygen's principle in the limit of small wavelength [40].

2.2.3 Maupertuis Action: The first Action Principle

Figure 2.4: Pierre Maupertuis. Image from `http://en.wikipedia.org/wiki/Pierre_Louis_Maupertuis`

The principle of least Action was first proposed by Pierre Louis Maupertuis (1698-1759) in 1744 [10, 35]. He speculated that in all events in nature the quantity called Action (which he defined as $S = mvs = mass \times velocity \times time$) is minimised [9]. For Maupertuis the least Action principle had religious significance. He believed "The laws of motion and rest deduced from the attributes of God" and stated "Here then is this principle, so wise, so worthy of the Supreme Being: Whenever any change takes place in Nature, the amount of Action expended in this change is always the smallest possible".

Leonhard Euler(1707-1783) realised Maupertuis principle is meaningless unless you assumed conservation of energy (conservation of energy allows the formulation to take time and potentials into account). He also redefined the Action as [9]:

$$W = \int_{\text{initial position}}^{\text{final position}} mvds \tag{2.1}$$

Where W is the Maupertuis Action, m is the mass, v is the velocity and s is the position.

Euler also realised that if the Action integral is minimised along the entire path, it must also be minimal for every subsection of the path [9]. This allowed him to create a graphical method of calculating the path of minimum Action. Joseph Louis Lagrange worked with Euler to create the Euler-Lagrange equation, a much more sophisticated (but less intuitive) method of calculating the path of stationary Action.

Lagrange's method of the calculus of variations very easily produces differential equations and analytical solutions for the path of stationary Action. For this reason the Euler-Lagrange equations dominate the way Action is taught in higher level classical mechanics [9]. However Euler had a graphical method of understanding Action that can now be easily simulated on computers and could prove to be a more intuitive way for students to find the paths of stationary Action [9].

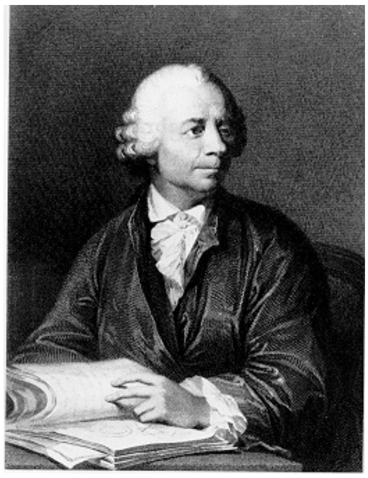


Figure 2.5: Leonhard Euler. Image from `http://www.usna.edu/Users/math/meh/euler.html`

2.2.4 Euler's Graphical Method of finding paths of Least Action

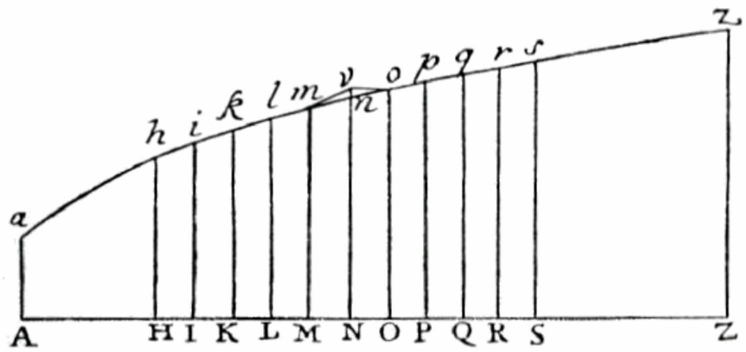


Figure 2.6: Euler's original figure used in his derivation of finding paths of stationary Action [41, 9]. This figure illustrates Euler's method of finding paths of stationary Action. Euler realised that if Action is stationary for an entire path then it must be a stationary for every small segment along that path. An equation is generated using the stationary condition on the middle point of each triplet along the path for example point.

Euler's method of finding paths of stationary Action is described thoroughly by Jozef Hanc in his paper *'The original Euler's calculus-of-variations method: Key to Lagrangian mechanics for beginners'* [41]. Euler's method involves dividing a path into a series of small timesteps, approximating the stationary path as a piece-wise function, see Figure 2.7.

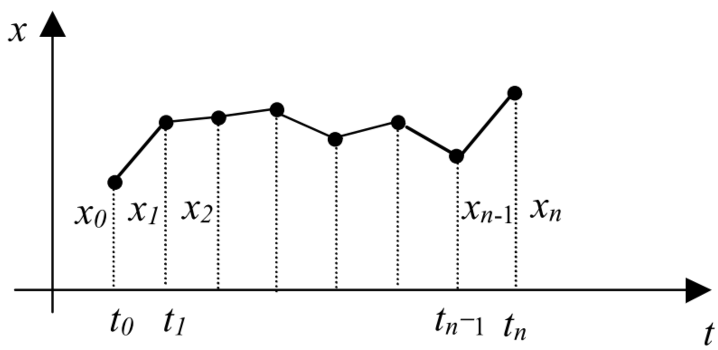


Figure 2.7: Diagram from [41] showing how Euler divided a path into a series of n segments. The two boundary positions x_0 and x_n are fixed. The stationary condition is used to create an algebraic equation for each position x_1 through to x_{n-1}. Solving these equations approximates the stationary path.

In Euler's time this would have been an extremely tedious process involving solving many simultaneous equations by hand. Lagrange's method (the Euler-Lagrange equation) would have been a much more convenient way to find paths of stationary Action. However with the technology available today and modern computers Euler's method is a great way to introduce students to the Action Principle and allow them to intuitively find paths of stationary Action [41]).

In modern times there are more effective methods than Euler's method for accurate numerical predictions of physical processes, most recently the Finite Element Method developed in 1943 [41].

2.2.5 Hamilton's Action

Figure 2.8: William Rowan Hamilton. Image from `http://hu.wikipedia.org/wiki/William_Rowan_Hamilton`

A hundred years after Euler in 1834 William Rowan Hamilton (1805-1865) developed the definition of Action that Richard Feynman came to call the Action Principle [9, 10]. Hamilton defined Action as:

$$S = \int_{\text{initial event}}^{\text{final event}} L dt = \int_{\text{initial event}}^{\text{final event}} (K - V) dt \tag{2.2}$$

Where S is Hamilton's Action, L is the Lagrangian, t is time, K is kinetic energy and V is potential energy.
Hamilton also realised that a physical path did not necessarily have minimum Action but was always stationary compared with Action along adjacent alternative paths between the same initial and final events (event referring here to a position and times a the concept of events was developed by Albert Einstein nearly 100 years later) [9].

2.2.6 Einstein and the principle of Maximal Ageing

In the 20th century Einstein developed the principle of maximal aging. Einstein hypothesised that the path (worldline) for a free particle makes the total proper time maximal. For a free particle the relativistically correct expression of Action is [9]:

$$S = -mc^2 \int_{\text{initial event}}^{\text{final event}} d\tau \tag{2.3}$$

Where S is the Action, m is the particle's mass and τ is the proper time. When this definition of Action is used the principle of maximal aging is entirely equivalent to the Principle of Least Action if a relativistic Lagrangian is used 2.5. The minus allows the principle of *maximal* aging to be reconciled with the Principle of *Least* Action.

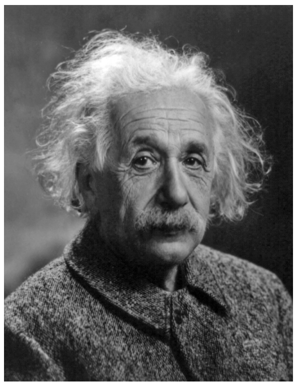


Figure 2.9: Albert Einstein. Image from `http://en.wikipedia.org/wiki/Albert_Einstein`

2.2.7 Feynman and the Many Paths Formulation of Quantum Mechanics

Figure 2.10: Richard Feynman. Image from http://www.atomicarchive.com/Bios/FeynmanPhoto.shtml

As a high school student Richard Feynman was shown the Principle of Stationary Action by his physics teacher Mr. Bader [42]. Ever since Feynman found the Principle of Stationary Action fascinating [42]. Richard Feynman worked on the Principle of Stationary Action as a PhD student, which lead to his development of the many paths formulation of quantum mechanics. The many paths formulation of quantum mechanics is equivalent to the Schrodinger and Heisenberg versions, but it provides a direct explanation for the Principle of Stationary Action. This helped develop 'Feynman diagrams' which are used for calculations in quantum electrodynamics for which he won a Nobel prize with Schwinger and Tomonaga [9].

In Feynman's Action model of quantum mechanics, an electron (or any particle) takes all possible paths. Each path has an 'arrow' or phasor associated with it, where the direction of the arrow (phase) is determined by the Action of the path. To find the overall probability of a transition occurring the arrows (phasors) of every possible path must be added together 2.3 (like waves constructively and destructively interfering).

Only the phasors around a path of stationary Action will add up constructively. All other paths will add up destructively and there will be almost no probability of these transitions occurring. That is why the path of stationary Action is the only path observed in classical situations.

It so happens that Feynman's treatment of particles in his many path formulation of quantum mechanics is entirely equivalent to Huygen's principle for light [9] [43]. Feynman simply generalised Huygen's principle to all particles, effectively treating all particles as waves.

2.2.8 Recent Developments: Generalised Hamilton and Maupertuis Action

Recently, Chris Gray and Gabriel Karl and V. A. Novikov wrote a paper 'Progress in classical and quantum variational principles' [11] in 2004. In this paper they review the development and use of the Maupertuis and Hamilton Action principles. In this work they discuss the generalised Hamilton and Maupertuis Action principles, using the actions given in Equation 2.4 and Equation 2.5.

Both Actions are formulated in terms of generalised co-ordinates q and their conjugate momenta p (Note q and p stand for a complete set of generalised co-ordinates $q_1, q_2, ..., q_f$ and their conjugate momenta $p_1, p_2, ..., p_f$, where f is the number of degrees of freedom and $p_i = \partial L/\partial q_i$)[11]. Gray, Karl and Novikov define the Hamilton Action as [11]:

$$S = \int_{t_A}^{t_B} L(q(t), \dot{q}(t), t)dt \tag{2.4}$$

Where S is the generalised Hamilton Action, t_A is the time of the initial event, t_B is the time of the final event, L is the Lagrangian and t is the time. Note that for the Generalised Hamilton

Action the starting and ending *events* must be fixed. However for the Generalised Maupertuis Action the starting and ending positions q_A and q_B must be fixed, but the times are allowed to vary. Gray, Karl and Novikov define the Maupertuis Action as[11]:

$$W = \int_{q_A}^{q_B} p\, dq = \int_{t_A}^{t_B} 2K\, dt \tag{2.5}$$

Where W is the Maupertuis Action, q_A is the initial position, q_B is the final position, K is the kinetic energy. Note pdq stands for $p_1dq_1 + ... + p_fdq_f$. The Generalised Maupertuis Action principle requires that the time average energy of the path $\bar{E} = \frac{1}{T}\int_{t_A}^{t_B} H(q(t), p(t))dt$ is fixed, a weaker assumption than conservation of energy (note $\bar{E}$ is the time average energy, T is the total time and H is the Hamiltonian) [11].

Gray, Karl and Novikov introduce a shorthand to summarise the two generalised Action principles. The shorthand for the Generalised Hamilton Principle is [11]:

$$(\delta S)_T = 0 \tag{2.6}$$

This is shorthand for: 'For a fixed time interval (T) a small perturbation to the classical path makes no first order change to the Hamilton Action (S)'. Another way of saying this is 'for the classical path of fixed time (T) the Hamilton Action (S) is stationary'. This notation writes the constraint of fixed travel time T explicitly to emphasise this constraint on the path [10]. Note that we are still assuming the path has fixed starting and ending positions, however this constraint will be applied to all Action principles that we will consider so it is left implicit [10].

The Generalised Maupertuis Action Principle is described in this notation in Equation 2.7:

$$(\delta W)_{\bar{E}} = 0 \tag{2.7}$$

This translates to 'for a classical path of fixed average energy ($\bar{E}$) the Maupertuis Action (W) is stationary'. Note that the assumption that ($\bar{E}$) is fixed is much weaker than assuming energy is conserved and it is in fact possible to derive conservation of energy from this [11]. Notice that this principle very closely represents Fermat's Principle of Stationary time for a photon (if you assume that photons experience no potential).

An explanation and justification for each of these principles comes from Richard Feynman's many path formulation of quantum mechanics. Taylor suggests that an easy way to help students understand the link between classical mechanics and quantum mechanics is to provide them with a conceptual understanding of the many paths formulation. Therefore this is discussed in detail the next section (Section 2.3).

To understand the many paths formulation of quantum mechanics students will already need to know what energy is and understand the different forms of energy. If teaching Action to a first year university physics class this should be pre-requisite knowledge and it is then possible to begin teaching physics through Action with the many paths formulation.

However for a highschool class it would be necessary to introduce the concepts of kinetic and potential energy first. Taylor and other Action experts [5, 44] recommend that the instructor teaches conservation of energy and allows students to solve one dimensional equations of motion with it so students can see application of energy to everyday life phenomena and do some classical mechanics calculations straight away. Once students are familiar with the concept of energy then they can be shown the many paths formulation of quantum mechanics.

2.3 Quantum Mechanical Basis of Action: Many Path Formulation of Quantum Mechanics

In 'QED: A Strange Theory of Light and Matter' Feynman explores the conceptual basis of the many paths formulation of quantum mechanics [28]. In 'Quantum Mechanics and path Integrals' Feynman and Hibbs discuss the mathematics that underpins the many paths formulation of quantum mechanics [31]. Feynman's 'QED' was written for an audience with no knowledge of physics so it makes a great text for students taking introductory physics. Edwin Taylor and other authors have iterated and refined the approach Feynman uses in QED [32]. I will refer to this as the 'QED approach' to teaching quantum mechanics. A brief outline of the QED approach is given below.

Richard Feynman based his formulation of quantum mechanics on two postulates, they are [45]:

1. *If an ideal measurement is performed to determine whether a particle has a path lying in a region of spacetime, then the probability that the result will be affirmative is the absolute square of a sum of complex contributions, one from each path in the region.*

2. *The paths contribute equally in magnitude but the phase of their contribution is the classical Action (in units of $\hbar$).*

These postulates can be expressed mathematically by:

1. $P \propto |\sum_{\text{All Paths}} \Psi_{\text{path}}|^2$

2. $\Psi_{\text{path}} = e^{\frac{iS}{\hbar}}$

Where P is the transition probability, S is the Action, Ψ_{path} is the complex amplitude of a given path, $\hbar$ is Planck's constant.The normalisation factor for postulate 1 can be found in Chapter 2 of Feynman and Hibb's 'Quantum Mechanics and Path Integrals' [31].

Feynman makes his many paths formulation of quantum mehcanics accessible to introductory students and the general public by the constant use of 'arrows' and diagrams [28]. One of the diagrams that Feynman uses in QED [28] is shown below to introduce the reader to this approach. Feynman represents the complex amplitude of each path as an arrow (phasor) and explains quantum mechanics through this strong visual representation [28]. Rather than providing a complete description of Feynman's explanation here, I will simply direct the reader to 'QED: A Strange Theory of Light and Matter' [28].

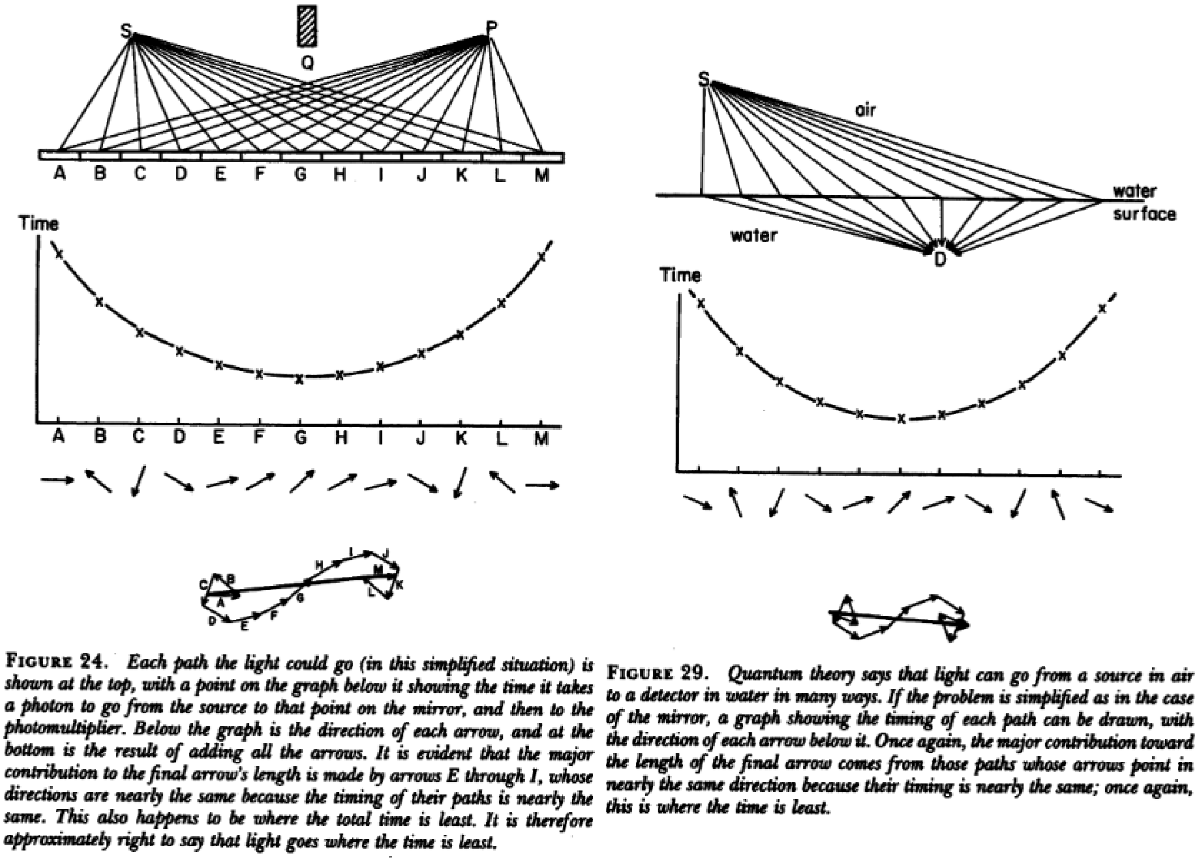


Figure 2.11: Diagram from Feynman's 'QED: A Strange Theory of Light and Matter' [28]. The figure is included here to give the reader an idea of how Feynman introduces quantum mechanics to non-physicists. This method is recommended by Taylor to teach physics to introductory students [32, 8]

2.4 Classical Mechanics from Action

Feynman's many paths formulation is completely counter-intuitive. In our everyday experience of classical mechanics we do not see a football taking all possible paths when it is passed from one player to another. How can Feynman's two postulates be reconciled to our everyday experience of classical mechanics?

The answer is through the Principle of Stationary Action: For paths that do not have stationary Action the phase can vary rapidly for adjacent paths and the sum of the contributions is negligible. We assume that only paths where Action is stationary allow constructive interference. Therefore only the paths near the stationary path will make a significant contribution to the total amplitude (and probability) of a transition.

If we take a specific look at Action, we can see for a free particle or a particle in a gravitational potential field that the mass is directly proportional to the Action (and therefore the phase). As mass increases the phase difference between adjacent paths also increases, narrowing the 'pencil' of paths that make a significant contribution to the overall phase.

In the limit where mass is very large (classical situations) the pencil of path narrows completely until only a tiny region around the stationary path that is making a significant contribution to the overall probability. The question is how can we predict what these stationary paths will be? How can we find stationary paths?

Finding Physical Paths - The Euler-Lagrange Equation

Now we have established that paths of stationary Action are classical paths, the question arises, "how do we find these stationary paths?" To do this, we could use Euler's method, the Direct Variation Method (also known as the "Rayleigh Ritz method", see [10]) or we could use the Principle of Stationary Action to derive Euler-Lagrange equation. There are a few standard ways of doing this. Pedagogically, I believe the best way is using a method first shown to be by Joe Hope when I took the third year theoretical physics course last year (lectures for this class are available online at `https://edge.edx.org/courses/ANU/Phys3001/2014Semester1/info`). The reason I prefer this method of deriving the Euler-Lagrange equation is that it makes explicit what is meant by a small perturbation in a path by using the $d/d\epsilon$ notation. Other derivations refer to quantities that may seem vague or unclear to students such as ΔS and δx (as done in [46]). Another derivation specifically designed for highschool students is described in [47]. First consider two events a and b. For simplicity we will just consider one spatial dimension at the moment however this derivation is easily generalised to higher dimensions and generalised co-ordinates. Consider any arbitrary path $x(t)$ between a and b.

Let's assume there is some path $x'(t)$ that has stationary Action. Then any arbitrary path can be written as $x(t) = x'(t) = \epsilon n(t)$, where $n(t)$ is an arbitrary function that takes the value of 0 at the endpoints (otherwise the path would no longer join a to b). Therefore $n(a) = n(b) = 0$.

Since Action is given by:

$$S = \int_a^b L(x(t), \dot{x}(t), t)dt \tag{2.8}$$

If a given path $x'(t)$ is stationary then:

$$\frac{dS}{d\epsilon} = 0 \tag{2.9}$$

Substituting S gives:

$$\frac{dS}{d\epsilon} = \int_a^b \frac{dL}{d\epsilon}dt = \int_a^b \left(\frac{\partial L}{\partial x}\frac{dx}{d\epsilon} + \frac{\partial L}{\partial \dot{x}}\frac{d\dot{x}}{d\epsilon} + \frac{\partial L}{\partial t}\frac{dt}{d\epsilon}\right) dt$$

Note $\frac{dt}{d\epsilon} = 0$ because ϵ does not depend on time. This can then be separated into two integrals:

$$\frac{dS}{d\epsilon} = \int_a^b \frac{\partial L}{\partial x}\frac{\partial x}{\partial \epsilon}dt + \int_a^b \frac{\partial L}{\partial \dot{x}}\frac{\partial \dot{x}}{\partial \epsilon}dt$$

Note that $\dot{x}(t) = \frac{dx}{dt}$ and that $\frac{dx}{d\epsilon} = n(t)$. Then:

$$\frac{dS}{d\epsilon} = \int_a^b \frac{\partial L}{\partial x} n(t) dt + \int_a^b \frac{\partial L}{\partial \dot{x}} \frac{d^2 x}{d\epsilon dt} dt$$
$$\frac{dS}{d\epsilon} = \int_a^b \frac{\partial L}{\partial x} n(t) dt + \int_a^b \frac{\partial L}{\partial \dot{x}} \frac{dn(t)}{dt} dt$$

Now integrating the second term by parts gives:

$$\frac{dS}{d\epsilon} = \int_a^b \frac{\partial L}{\partial x} n(t) dt + \frac{\partial L}{\partial \dot{x}} n(t) \Big|_a^b - \int_a^b \frac{d}{dt} \left(\frac{\partial L}{\partial \dot{x}} \right) n(t) dt$$

Note that since $n(a) = n(b) = 0$ (the end points are fixed), the second term on the right hand side of this equation is 0. Factorising $n(t)$ out of the remaining terms gives:

$$\frac{dS}{d\epsilon} = \int_a^b \left(\frac{\partial L}{\partial x} - \frac{d}{dt} \frac{\partial L}{\partial \dot{x}} \right) n(t) dt$$

Since $n(t)$ is an arbitrary perturbation along the path the term in the brackets must be 0 for all values of t. Therefore in order for a path to be stationary the following equation (the Euler-Lagrange equation) must hold at every point along the path:

$$\frac{\partial L}{\partial x} - \frac{d}{dt} \left(\frac{\partial L}{\partial \dot{x}} \right) = 0 \tag{2.10}$$

Once the Euler-Lagrange equations have been derived students can use them to solve classical mechanics problems that are very difficult to solve using Newton's laws. One classic example of this is the double pendulum, obtaining the equations of motion using Newton's laws is very difficult because the constraining forces constantly change directions. However it is relatively easy to formulate the equations of motion in terms of Lagrangian mechanics.

2.5 Action and Relativity

In relativity, the Principle of Stationary Action is expressed as, the Principle of Maximal Aging:

"A free particle will take a path between two events that maximises its proper time"

The link between the classical mechanical Principle of Stationary Action and relativistic principle of maximal aging is simple to show by assuming the relativistic Lagrangian takes the form $-mc^2\sqrt{1-v^2/c^2}$ [48].

$$S = \int_{\text{initial event}}^{\text{final event}} -mc^2\sqrt{1-v^2/c^2}dt$$

$$S = \int_{\text{initial event}}^{\text{final event}} -mc^2 d\tau$$

$$S = -mc^2 \int_{\text{initial event}}^{\text{final event}} d\tau$$

Where τ is the proper time. Note that in this free particle Lagrangian does not take the form $L = K$ as relativistic kinetic energy is given by $K = \frac{mc2}{\sqrt{1-\frac{v^2}{c^2}}} - mc^2$ [48]. Once the principle of maximal aging has been established, it can be used to solve some conceptual problems in relativity immediately, for example the twin paradox.

Teaching introductory physics students according to the procedure outlined above gives them an exposure to quantum mechanics and relativity and allows students to see how these ideas are consistent with the classical mechanics they experience in everyday life. This also exposes students to the basis of fundamental physics which is formulated in terms of Lagrangians.

2.6 Extensions of the Action Principle

It is possible to go one step further and use the Principle of Stationary Action to introduce a few other very powerful concepts in physics including generalised co-ordinates and symmetry. Generalised coordinates allow students to solve even more difficult classical mechanics problems using Lagrangian mechanics. Noether's Theorem allows students to see the basis for the conservation laws, some of the most fundamental laws that they have been told to assume are true.

2.6.1 Generalised Coordinates

In physics it is often more convenient to use coordinates that represent the symmetry of the situation. For example if we are considering a particle that is moving around on the surface of a sphere, rather than using three Cartesian co-ordinates (x, y and z), we could instead use spherical polar co-ordinates and describe the position of the particle with just two angular co-ordinates (ϕ and θ). This coordinate system not only decreases the number of variables, but also naturally constrains the particle to rest on the surface of the sphere.

There are three main coordinate systems that are simple to understand and can be easily introduced to the students, they are Cartesian coordinates (x, y, z which the students should already be familiar with), spherical polar coordinates (r, θ, ϕ) and cylindrical coordinates (ρ, ϕ, z).

Once students are familiar with the co-ordinate systems, they can be shown the generalised Euler-Lagrange equations for any coordinate q:

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}}\right) - \frac{\partial L}{\partial q} = 0 \tag{2.11}$$

Once students have been given the Euler-Lagrange equations in generalised coordinates they are able to solve many problems that would have been extremely difficult using Newton's laws.

2.6.2 Noether's Theorem

Once students are familiar with the Euler-Lagrange equation and generalised co-ordinates. It is a small extension to tell students about Noether's theorem and the origin of the conservation laws that they are familiar with. Although Noether's theorem in its full generalised form would require unfamiliar mathematical notation and be extremely complicated for the students, there is a simple version that can be shown simply using the Euler-Lagrange equation.

Conservation of Momentum

First assume the Lagrangian takes the form $L = \frac{1}{2}mv^2 - V(y,z)$, where $v = \sqrt{\dot{x}^2 + \dot{y}^2 + \dot{z}^2}$ and note the potential V does not depend on x. Now consider the Euler-Lagrange equation for the x coordinate:

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{x}}\right) = \frac{\partial L}{\partial x}$$

$$\frac{d}{dt}\left(\frac{\partial(\frac{1}{2}m(\dot{x}^2 + \dot{y}^2 + \dot{z}^2) - V(y,z))}{\partial \dot{x}}\right) = \frac{\partial \frac{1}{2}mv^2 - V(y,z)}{\partial x}$$

$$\frac{d}{dt}(m\dot{x}) = 0$$

Therefore the quantity in brackets on the left hand side of the equation does not change with time (therefore it is a conserved quantity). Students should recognise $m\dot{x}$ as classical momentum, showing that conservation of momentum arises from translation symmetry.

Conservation of Angular Momentum

If we assume that the Lagrangian does not explicitly depend on an angular co-ordinate, $\partial L/\partial\theta = 0$ then the Euler-Lagrange equation for that co-ordinate becomes (note I have used 3 dimensional polar-coordinates to generalise to 3D without making technical inaccuracies):

$$\frac{d}{dt}\left(\frac{\partial L}{\partial\dot{\theta}}\right) = \frac{\partial L}{\partial\theta}$$
$$\frac{d}{dt}\left(\frac{\partial(\frac{1}{2}mv^2 - V(r,\theta,\phi))}{\partial\dot{\theta}}\right) = 0$$

And note that v has three components, $v^2 = \dot{r}^2 + r^2\dot{\theta}^2 + r^2\dot{\phi}^2$, therefore:

$$\frac{d}{dt}\left(mr^2\dot{\theta}\right) = 0$$

Therefore the quantity in the brackets is conserved. Students may recognise this as angular momentum if it is expressed as $mr^2\omega$ or mr^2v_p where ω is the angular velocity and v_p is velocity perpendicular to r. This allows students to see that conservation of angular momentum arises from angular symmetry.

Beltrami Identity and Conservation of Energy

In cases where the Lagrangian does not depend explicitly on time there is a simplified version of the Euler-Lagrange equation that can be used, known as the Beltrami Identity. The Beltrami Identity is:

$$L - \dot{q}\frac{\partial L}{\partial\dot{q}} = C \tag{2.12}$$

Where L is the Lagrangian, q is a generalised co-ordinate and C is a constant. It is straightforward to derive the Beltrami Identity from the Euler-Lagrange equation if you assume $\partial L/\partial t = 0$, an introductory physics class should be able to follow (or complete) the derivation (shown below) and doing so may help them familiarise themselves with the difference between total differentiation and partial differentiation.

$$\frac{d}{dt}\left(\frac{\partial L}{\partial\dot{q}}\right) = \frac{\partial L}{\partial q}$$
$$\dot{q}\frac{d}{dt}\left(\frac{\partial L}{\partial\dot{q}}\right) = \dot{q}\frac{\partial L}{\partial q}$$

The chain rule gives:

$$\frac{dL}{dt} = \frac{\partial L}{\partial q}\dot{q} + \frac{\partial L}{\partial\dot{q}}\ddot{q} + \frac{\partial L}{\partial t}$$

$$\dot{q}\frac{\partial L}{\partial q} = \frac{dL}{dt} - \frac{\partial L}{\partial \dot{q}}\ddot{q}$$

Since $\partial L/\partial t = 0$. Substituting this back into the Euler-Lagrange equation gives:

$$\dot{q}\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}}\right) = \frac{dL}{dt} - \frac{\partial L}{\partial \dot{q}}\ddot{q}$$

Then by the product rule we have:

$$\dot{q}\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}}\right) = \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}}\dot{q}\right) - \frac{\partial L}{\partial \dot{q}}\ddot{q}$$

Therefore:

$$\begin{aligned}\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}}\dot{q}\right) - \frac{\partial L}{\partial \dot{q}}\ddot{q} &= \frac{dL}{dt} - \frac{\partial L}{\partial \dot{q}}\ddot{q} \\ \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}}\dot{q}\right) &= \frac{dL}{dt} \\ \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}}\dot{q} - L\right) &= 0\end{aligned}$$

Integrating with respect to time gives:

$$\frac{\partial L}{\partial \dot{q}}\dot{q} - L = C$$

Where C is a constant of the motion. Note that if we assume that L depends on multiple coordinates $q_1, q_2, ..., q_n$ then the application of the chain rule yields more terms. Each of the terms can be treated with the product rule as for one coordinate. The end result is that there is a $\frac{\partial L}{\partial \dot{q}_i}\dot{q}_i$ for each coordinate q_i in the final equation.

Once the Beltrami Identity has been derived it is easy to show that if we assume that the Lagrangian takes the form $L = K - V = \frac{1}{2}m\dot{q}^2 - V(q)$, this is equivalent to conservation of energy that the students are already familiar with:

$$\begin{aligned}\frac{\partial \frac{1}{2}m\dot{q}^2 - V(q)}{\partial \dot{q}}\dot{q} - \frac{1}{2}m\dot{q}^2 + V(q) &= C \\ m\dot{q}^2 - \frac{1}{2}m\dot{q}^2 + V(q) &= C \\ \frac{1}{2}m\dot{q}^2 + V(q) &= C\end{aligned}$$

Or for multiple coordinates:

$$\frac{1}{2}m\dot{q_1}^2 + \frac{1}{2}m\dot{q_2}^2 + ... + \frac{1}{2}m\dot{q_n}^2 + V(q) = C$$

Students should recognise this as conservation of energy that they are already familiar with $K + V = 0$. This shows students the result from Noether's theorem that time symmetry leads to conservation of energy.

2.7 Overcoming the problems of the Teaching Action in Introductory physics

In Table 2.1 three disadvantages of teaching introductory physics through Action physics are highlighted. These were: lack of available teaching resources, difficulty when dealing with frictional forces and the difficulty of the mathematics involved. In this section I highlight how each of these problems can be overcome.

Problem 1: Lack of Available Teaching Resources

Although there is no one complete textbook that contains all the explanations about Action, complete with exercises, problems and solutions, there are resources out there for Action instructors to use. Hopefully Section 2.1 will allow Action instructors to know what resources are currently available and make the most of those.

In addition there are many classical mechanics textbooks that contain not only explanations of Lagrangian mechanics but also include many examples and practice problems from other sources. Textbooks that I would recommend include:

- *'Classical Mechanics'* by Herbert Goldstein, Charles Poole and John Safko - Chapter 2 [49]
- *Classical Mechanics* by David Morin - Chapter 6 [50]
- *Problems and Solutions of Mechanics* edited by Lim Yung-kuo - Page 461-521 [51]
- *Classical Mechanics* by R. Douglas Gregory - Chapter 12 [52]

Problem 2: Dealing with Frictional Forces

Lagrangian mechanics does not naturally incorporate the frictional forces which commonly play a key role in classical mechanics problems (especially statics problems). This is a fair criticism of the Lagrangian mechanics, it only works well with fundamental forces that can be expressed in terms of Lagrangians.

There are many cases where friction or drag is a dominant force, but we do not wish to track the energy or momentum being imparted from a system to its surroundings. In these cases we consider only the energy and momentum of the objects we are modelling, and we find that energy and momentum are not conserved in these cases.

The Action Principle (Lagrangian mechanics), does 'work' in these cases. It could be used in cases where friction and drag apply and it would give correct results, but it would involve modelling the fundamental force (most likely electromagnetic interaction) from every molecule of air that exerts a drag force on the object. These calculations are completely unnecessary at a first year level (and impossible to do without a computer) and simply over-complicate the situation. In these cases it is most appropriate to make the approximation that energy is simply being lost from the system and forget about the surroundings.

In these cases energy dissipation does not fit 'naturally' into the Lagrangian. You can force it, but it is mathematically awkward. Chris Gray discusses when friction can and cannot be included into the Stationary action principle [10]. In 'A Call to Action', Edwin Taylor addresses the difficulty of dealing with friction forces [5]. Taylor states:

> *'The comment that the Principle of Least Action deals awkwardly with dissipative and frictional forces is certainly valid. Under our proposal force is not eliminated but becomes a secondary concept, available to analyse friction and dissipative forces. Making force secondary clears the center stage for the Principle of Least Action to predict motions of a huge variety of simple and complex systems.'*

In many introductory physics courses, instructors often ask students to ignore friction and drag to simplify problems. Assuming vacuums and frictionless surfaces is not a novel idea in introductory classical mechanics and in these cases the Euler-Lagrange equations can be applied.

Just as there are tricks and simplifications when teaching students introductory physics through Newton's laws, there are also tricks that can be used to simplify mathematics when teaching students physics through Action. Some of these tricks are outlined in the the next Section: *Difficulty of Mathematics*.

Problem 3: Difficulty of Mathematics

Many authors have published papers where they outline methods of teaching Action physics which only require low levels of mathematics (algebraic manipulation and in some cases basic calculus). I have listed some of the key papers below:

- *From Conservation of Energy to the Principle of Least Action: A story line* by Jozef Hanc and Edwin Taylor [44].
- *'The original Euler's calculus-of-variations method: Key to Lagrangian mechanics for beginners'* by Jozef Hanc [41].
- *'Symmetries and Conservation Laws: Consequences of Noether's Theorem'* by Jozef Hanc, Slavomir Tuleja and Martina Hancova [53].

To use the Euler-Lagrange equations, it is essential that students can know how to take partial derivatives and total derivatives. The mechanics of differentiation is usually taught in later secondary school, so hopefully students will be partially familiar with differentiation already. The instructor will need to emphasise the difference between partial derivatives and total derivatives.

Students don't need to be able to solve the differential equations they generate from the Euler-Lagrange equations. It is quite common even in higher level courses to simply ask students to express their answer as a differential equation. In reality solving differential equations usually comes down to numerical simulations performed by a computer.

Instructors can use this as an opportunity to introduce students to mathematics software. The ability to use this software is an essential skill for many physics researchers. The students can use numerical solutions to check their solutions, practicing basic checks such as solutions having approximately expected behaviour (for example cyclic, oscillator or decaying) and appropriate behaviour in given limits.

This is not saying that students no longer need to learn the theory behind solving differential equations at some point in their education, but rather that this no longer holds us back from teaching Action at an introductory level.

As part of the project I observed students in tutorials. Although most students struggled setting up the Lagrangian and performing the differentiation initially, with the help from tutors they were able to get through and a problem using the Euler-Lagrange equation in a 50 minute tutorial. Analysing student surveys (see Section 5.7.2) showed that they did not find mathematics the most difficult part of learning Action.

So even though some of the mathematics may be new for the students, and is complicated, there are strategies that can be used to make it manageable for the students. The observations and survey data indicates that difficulty of mathematics is not a barrier that prevents introductory physics students from learning Action principles.

Conclusion

In the introductory physics through Action approach students are given a brief introduction to quantum mechanics through the Richard Feynman formulation, then they are shown the link between quantum mechanics and classical mechanics. Finally students are shown how the principle of Stationary (least) Action is equivalent to the principle of maximal aging, linking classical mechanics and relativity. Through this treatment students are able to view three of the major branches of physics as one consistent picture and don't have to continually 'unlearn' or 'relearn' physics every time they come to a new topic.

In an introductory physics class students will have been exposed to three of the biggest areas of physics and the key ideas that underpin each of them. This would not replace the QM or relativity classes (or even classical mechanics classes), but expose students to these important ideas early and give them a chance to sink in before further instruction. For students who don't continue on in physics this would have opened their minds to see more depth of physics than basic classical mechanics and Newton's laws, while still learning the important concept of energy.

Although multiple resources have been created by Action instructors [32, 5, 7, 53, 44, 41, 8, 34, 54, 9, 55, 10] (see also [6] and references therein), so far there has been no formal way of measuring which of these teaching methods are most effective. In order to evaluate and improve these methods we need a way of measuring student understanding. This provides the motivation to develop an Action Concept Inventory as outlined in the next chapter.

Chapter 3

Developing The Action Concept Inventory

A concept inventory is a conceptual test that measures students' understanding of a particular topic. Concept inventories and concept evaluations now exist for most branches of physics [26, 56, 27, 57, 58, 59, 25], however no one has yet developed am Action Concept Inventory.

3.1 The purpose of Concept Inventories

It is important that we define what the purpose of this Action Concept Inventory is early on. Although there are papers that have great recommendations for how to design valid and reliable concept inventories [25], [60], these recommendations should only be followed so far as they support the purpose of the inventory.

The primary use of a concept inventory is to measure the effectiveness of instruction [25] [61]. The standard way of doing this is to measure the change in student understanding of concepts before and after instruction and therefore the purpose of a concept inventory is to measure how well students understand a certain set of concepts [60] [62] [61] [25] [63].

The purpose of the Action Concept Inventory was to measure how well students understood the primary concepts that underpin the Principle of Stationary Action. All the processes and procedures outlined in this thesis, from inventory development, to international collaboration, to student interviews, to statistical analysis were all performed to ensure that the inventory fulfilled this purpose.

The purpose of the inventory will be frequently referred to throughout this thesis as it is the motivation for this project, so it is worth clearly setting it out so the reader may refer back to it. See below:

> *The purpose of the Action Concept Inventory is to measure how well students understand the concepts that underpin the Principle of Stationary Action.*

There are many other qualities of the concept inventory that may be desirable, such as convenience to administer and grade, reliability (student scores are consistent) and ability to discriminate between students, but these properties are all secondary to the purpose of the inventory. Indeed as you will see later in the thesis a compromise needs to be made between many of these qualities. They should ultimately support the purpose of the inventory, which is to measure how well students understand the concepts that underpin the principle of stationary Action concepts.

3.2 Developing Concept Inventories

The first concept inventory was developed by Hestenes, Wells and Swackhamer in 1992 [26] and since then a plethora of inventories have been created for a wide range of subjects [27]. There is now a standard method of developing reliable and valid concept inventories [25].

In 2010 Physics Nobel Laureate Carl Wieman, and Wendy Adams wrote a paper summarising the key techniques used to develop valid and reliable concept inventories [25]. I followed the procedures they outlined as closely as possible in the time frame available. I also developed a few techniques to improve the process of performing think aloud interviews (see Section 3.5.2 *Post Test Questioning* and Section 3.5.2 *Summarising Interviews*).

The key stages in developing the concept list and the concept inventory in this project are summarised in Figure 3.1.

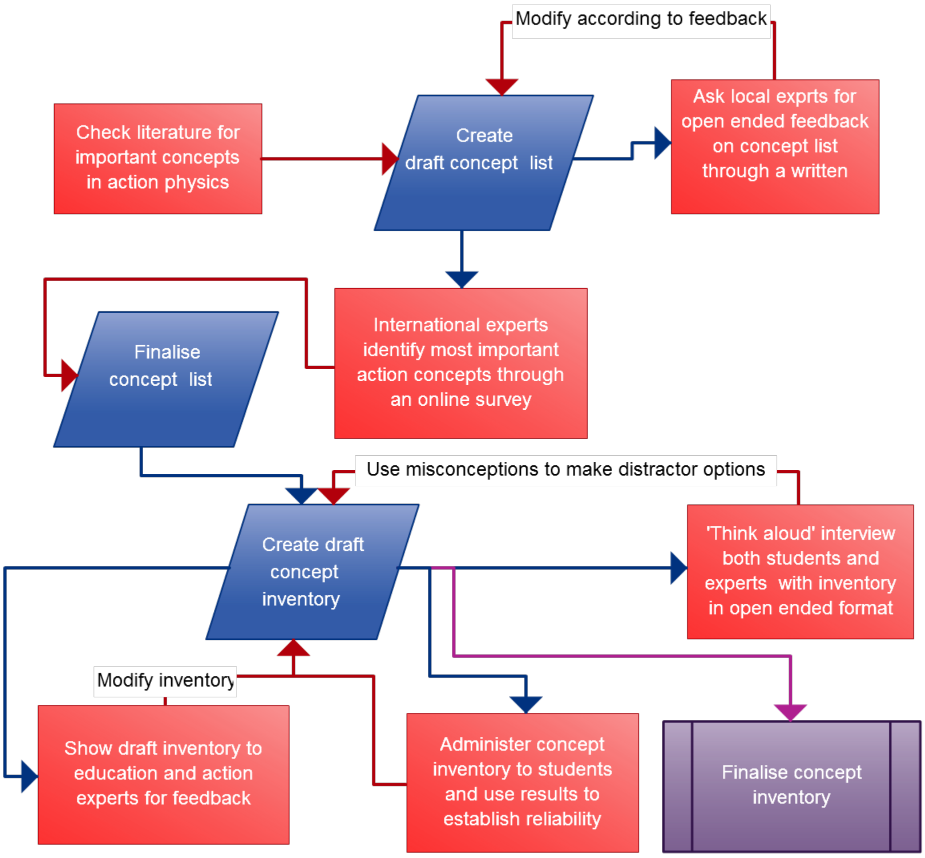


Figure 3.1: Process for developing a concept inventory. Key components of developing a concept inventory are constant consultation with experts and feedback. This ensures that all key concepts are identified and fully developed and that the inventory is valid. Blue rhomboids correspond to creative processes while red rectangles correspond to feedback and modification.

First the literature was reviewed to identify key concepts and create concept list was created. Action experts were then consulted to determine which concepts should be included in the inventory. A draft Action Concept Inventory was written and 30 formative think aloud interviews were performed with a wide range of students (and one academic), allowing the common misinterpretations to be identified and corrected. The Action Concept Inventory was delivered to the first year class as pre-test and post-test. Physicists also completed the Action Concept

Inventory. The results from both groups were statistically analysed. Validation interviews where then performed with the first year students and experts consulted to confirm the validity of the inventory.

3.2.1 Literature Review

Unlike Newtonian mechanics, electromagnetism or special relativity, no one has published the key concepts in Action physics or key misconceptions. Therefore a key component of this project was to establish and publish the important concepts of Action physics for future introductory Action physics research. A concept list is also vital for developing a concept inventory.

Initially I carried out a literature review of Action teaching materials to identify key concepts in Action physics. When developing a concept inventory, identifying key concepts from textbooks is one of the best places to start. Unfortunately, no Action physics textbooks have been published at the time of this project. There are several classical mechanics textbooks that include chapters on Lagrangian methods and the Principle of Stationary Action, however these mainly emphasize the use of Lagrangian Mechanics as a computational tool, not the conceptual basis behind the Stationary Action Principle. *'QED: A Strange Theory of Light and Matter'* by Richard Feynman [28] was one of the key sources for identifying the quantum mechanical Action concepts.

The majority of the key papers on teaching Action as introductory physics can be found on Edwin Taylor's website [6]. 'A Call to Action' [5] is Taylor's original paper where he highlights the advantages of teaching physics through Action. In 'From Conservation of energy to the Principle of Least Action' [44], Edwin Taylor and Jozef Hanc outline a specific method for teaching introductory physics through Action. 'Quantum Physics Explains Newtons Laws of Motion' [55] outlines methods for teaching quantum mechanical principles through the many path formulation of quantum mechanics. These explanations are very similar to Feynman's in QED. In addition there is a review paper of the Hamilton and Maupertius Action principle, 'Progress in classical and quantum variational principles' [11] which contains more advanced Action principles.

After identifying Action concepts from the literature, my supervisor Craig Savage and I discussed which of these were important and which ones we thought should be included in the Action concept inventory. This allowed us to create a preliminary list of 13 concepts. This preliminary list of concepts was reviewed by five ANU Action experts. They made many suggestions about which concepts should be included, which concepts were missing and how to improve the wording of the concepts.

3.2.2 Surveying Action Experts

International Action experts were then identified. Researchers who had published papers in Action physics education literature were included, as were developers of Action physics teaching resources. The Action concept list was sent to 21 international Action experts in the form of an online survey. Most Action experts contacted were from the United States of America, however two were from Russia and two from Slovakia. The ANU Action experts were also asked to complete the survey. A list of the Action experts who responded to the survey and the papers they published on Action physics literature is included.

In total 15 Action experts responded to the survey indicating which concepts they thought should be included and which concepts should not be included. Experts were also asked for their comments on each of the Action concepts and to mention any concepts they though were missing from the inventory.

In analysing the results it appeared that one expert accidentally reversed the strongly agree to strongly disagree scale. Not only did his (or her) responses stand out as an extreme outliers among the Action expert responses, but his comments contradicted his responses. The expert did not leave his (or her) name or contact details so he could not be contacted to clarify his response. For these reasons we decided to remove this response from the data set.

3.3 Concept List

Only concepts for which at least 70% of Action experts agreed or strongly agreed were tested in the inventory. They are shown in Table 3.1 below:

Table 3.1: Table summarising the concepts included in the inventory, the number of experts who agreed that the concept should be included in the inventory (as a %). The questions in the inventory which address each concept are also listed.

Action Concept	% Agreement	ACI Questions
Principle of Stationary Action	93%	Q7, Q9, Q16
Path	79%	Q10, Q11
Stationary	100%	Q4, Q12, Q13
Fermat's Principle	71%	Q6
Explore all Paths	93%	Q8
Compex Amplitude	93%	Q15, Q18
Superposition Principle	100%	Q15, Q18
Probability	100%	Q14, Q17
Conservation of Energy	79%	Q1, Q2
Principle of Stationary Potential Energy	79%	Q3, Q5

Each of the Action concepts is explored in detail in the sections below:

Principle of Stationary Action

The classical (observed) path is the one for which the Action is stationary

This concept is the centre of Action physics and connects classical mechanics to quantum mechanics. All ANU experts agreed that this was an important concept to include in the inventory. This principle raises many questions, for example "what is meant by a path?", "what do you mean by stationary?" and these are addressed in the later concepts.

Two Action experts made the comment that there are actually two Stationary Action Principles, Hamilton's Principle and Maupertuis' Principle. In this inventory we regard only Hamilton's

Action Principle as the Principle of Stationary Action. Hamilton's principle defines Action as:

$$S = \int_A^B L \, dt$$

The common misconceptions about the Principle of Stationary Action are:

- For a path to be physical it must have the minimum Action, rather than stationary Action.
- It is nonsensical to have multiple paths for which the Action is stationary.

There is no reason why the path of minimum Action should be more probable than other paths with stationary Action. For example consider the experimental set up shown in Question 6 of the concept inventory in Figure 3.2:

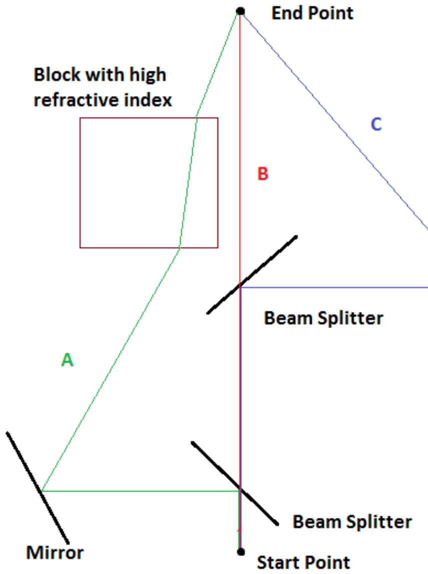


Figure 3.2: Photons are released from a point and are observed at an end point. The Action of a path taken by a photon is proportional to the time the photon spends travelling on the path. Each beam splitter transmits 50% of the light and reflects 50% of the light. Therefore in this experimental set up, path A, which has a greater time (and therefore greater Action) is more probable than either path B or C which both have a lower value of Action.

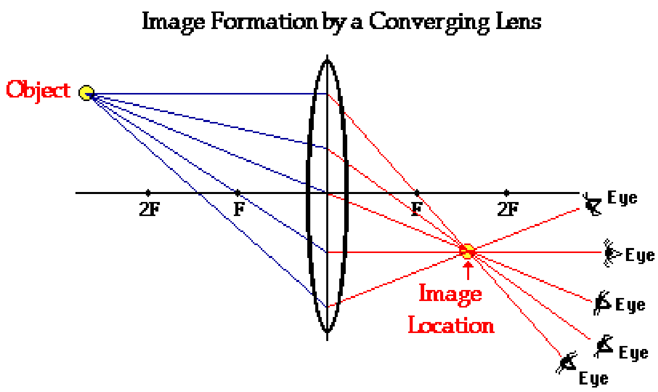


Figure 3.3: For a converging lens there are many paths that light can take for the object to the image location. Each of these paths have the same time and the same Action, therefore they are not minima with respect to all neighbouring paths. However they are first order stationary with respect to variations in the path. Image from The Physics Classroom: `http://www.physicsclassroom.com/class/refrn/Lesson-5/Converging-Lenses-Ray-Diagrams`

First order stationarity is the only requirement for neighbouring paths to add up constructively and therefore produce a significant probability of being observed. More information is required to determine which path (and its neighbours) make the greatest contribution to overall probability.

Many students, even one of the ANU Action experts struggled with the idea of multiple stationary paths and believed there is only one path for which the Action could possibly be stationary. However this is not true, when designing a lens that focuses light from one point to another point, Fermat's principle can be used. In this case the intention is to design the lens so there are many stationary paths between the two points as illustrated in Figure 3.3.
Think aloud interviews and student homework showed that the most common misconception is the 'Least Action' Misconception:

- **'Least Action'** misconception - The least Action misconception is: 'The classical (observed) path is the one for which the Action is *minimum*' rather than *stationary.*

ACI distractor options Q6(b), Q7(a)(b)(c), Q9(a)(b) and Q16(b) correspond to the 'Minimum Action' misconception and model analysis was used to provide information about the number of students who suffered from this misconception (see Sections 4.6 and 5.4.1).

Path

A path is a curve in space parametrised by time

One of the key points about paths in calculating Hamilton's Action for a path is that it needs to be a path parametrised by time, not a simple spatial trajectory. One common student misconception is that paths refer to a geometric or spacial path. In order to calculate Hamilton's Action for a path it must be a space-time trajectory or worldline. This distinction is illustrated in Figure 3.4.

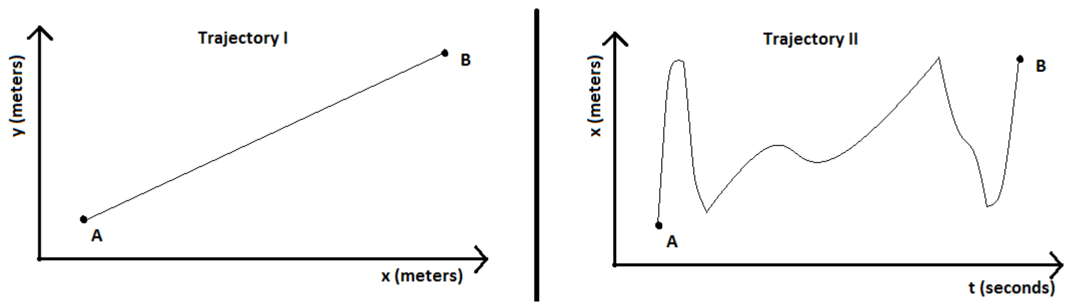


Figure 3.4: Trajectory II (shown right) is parametrised by time and hence only one value for Hamilton's Action can be calculated. By contrast trajectory I (shown left) is not parametrised by time and the description of the motion is incomplete. For example the particle could move at a constant speed along the path or in could move slowly along the beginning of the path and then accelerate. Therefore Hamilton's Action cannot be calculated for trajectory I.

We specifically worded this concept to avoid the relativistic concepts of space-time, 4-space and worldlines. These concepts are not necessary for students to understand what is meant by a path. The final wording was chosen to emphasise the importance of time in the path without venturing into relativistic concepts.

It was also noted that it was important that students understand that 'path' does not exclusively refer to the classical path, but can refer to any curve in 4-space from the start to the end point.

Three of the ANU experts queried what do we mean by the word 'possible'. The standard formulation of the many paths formulation is presented in 'Quantum Mechanics and Path Integrals' by Feynmann and Hibbs [31]. In this 1-Dimensional formulation a particle is allowed to move to any position at discrete timesteps. Then the limit where the time between steps approach as 0 is considered. Under this formulation relativistically forbidden (particles travelling faster than the speed of light) paths are allowed, but paths that travel back in time are not.

To calculate the Action of a path it needs to be parametrised by time and it needs to have fixed starting and ending *events*, not fixed starting and ending positions. This misconception is specifically tested in Question 10 of the Action Concept Inventory.

One Action expert pointed out that a path does not have to be parametrised by its own proper time:

> *The geodesic equation for the propagation of a photon in a curved spacetime can be derived from an Action, but the proper time cannot be defined for a photon. In this case another parametrisation must be chosen.*

This is true, however when calculating the Action for a photon the trajectory must still be parametrised by co-ordinate time even if it is not the photon's proper time.

Stationary

A first order variation of the path makes no first order difference to the Action

The most common misconceptions about the stationary concept are:

- **'Second Order Greater than First Order'** misconception - At a stationary point very small first order changes are not necessarily larger than very small second order changes.
- **'No Change'** misconception - Around a stationary point of $y(x)$, arbitrarily small changes in x make zero difference to the value of y.
- For a point to be stationary it must be either a minimum or a maximum

Many students did not realise that a first order change is much larger than a second order change for small displacements. Especially if diagrams make the slope seem extremely gentle. For example, in question 4 of the inventory originally a 45^o slope was shown (See Figure 3.5). However think aloud interviews revealed that students with the 'Second Order Greater than First Order' misconception would choose the correct answer simply because the slope looked steeper. These questions were adjusted to so very gentle slopes were shown (See Figure 3.5).

More advanced physics students and research physicists were more prone to the 'no change' misconception, that an arbitrarily small perturbation will make absolutely no change to the Action of a stationary path. This is false, there will be a change in the Action, it will be a second order change.

There is an extension of the real numbers, called the infinitesimals which have the property $x > 0 \not\Longrightarrow x^2 > 0$. However if you extend the real numbers to include these infinitesimals then you loose certain properties/axioms of the real numbers, for example ordering $(x \neq 0 \implies |x| > 0)$ [64]. When talking about macroscopic systems referring to such small displacements is not sensible and it is certainly not useful to do this when trying to perform a sum over all paths calculation.

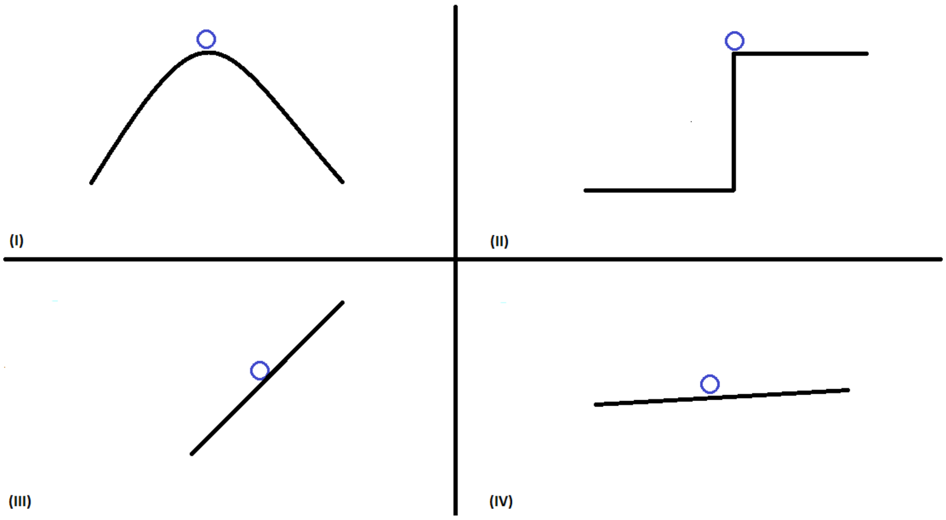


Figure 3.5: Slope shown for Question 4 of the Action Concept Inventory before and after think aloud interviews. *"Each ball is displaced a very small (arbitrarily small) amount to the left, order the change in height from greatest to least"*. Students would choose the slope III to have greater change than slope I simply because it looked the steepest. The modified slope used in the final inventory is shown in pane IV. During think aloud interviews it was observed that only students who think about the concept of stationary and first order change would chose slope IV to change by more than slope I.

Another common student misconception is that students may think that a path has to be local maximum or minimum to be stationary, however this is not generally true it may be saddle point or a flat region (for example the lens in Figure 3.3). The important concept is that a first order variation in the path will lead to no first order change in the stationary quantity. In the case of Action this means there is no first order change in phase, leading to constructive interference between neighbouring paths.

Both the 'No Change' and 'Second Order Greater than First Order' misconceptions were tracked using Model Analysis (see Sections 4.6 and 5.4.1). ACI distractor options Q3(c)(d)(e)(f), Q12(a) and Q13(a) correspond to the 'No Change' misconception and Q4(c), Q12(b)(d) and Q13(b)(e) correspond to the 'Second Order Greater than First Order' misconception.

Fermat's Principle

When travelling from point A to point B a light ray will take the path for which the time is stationary

Fermat's principle is a simple extension of the Principle of Staionary Action applied to ray optics. Out of the concepts kept on the inventory, Fermat's principle received the least agreement from Action experts (71%) and only one question was included on the inventory to test Fermat's principle. Fermat's principle is entire equivalent to the Principle of Stationary Action when we consider than the Action of a photon is directly proportional to the time it travels, $S = \int_a^b \hbar\omega \, dt$. Because of this similarity the Fermat's Principle question was grouped with the Principle of Stationary Action questions when performing model analysis, factor analysis and looking for correlation pairs (see Chapter 4 *Statistical Analysis*).

The argument for not including Fermat's principle on the concept list was that it is not central to understanding the Principle of Stationary Action, which is true. The methods of teaching the Stationary Action Principle used by Feynman and Taylor introduce Fermat's principle at an early stage as a way of introducing the idea that each path can have a quantity associated with it [28, 44]. In this case of Fermat's principle this quantity is time, which is more familiar to students than the newly introduced Action. Fermat's principle is regularly used to introduce students to path optimisation [65], [55], [44]. Since it is key to Action instruction and was a borderline case in terms of expert agreement, Fermat's principle was kept on the final concept list.

Historically Fermat's principle is stated as 'The path travelled between two points by a light ray is the path that minimises the time travelled'. However this is technically incorrect. The path must have stationary time not minimum. One example that makes this particularly clear is shown when considering an object of extremely high refractive index, see Figure 3.6.

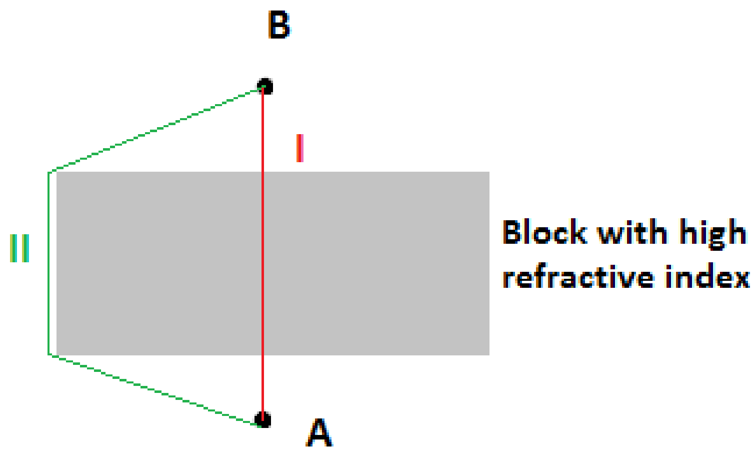


Figure 3.6: Here the shaded medium has an extremely high refractive index and the path shown in red is the path of minimum time. A principle of minimum time would predict that path II is the path taken by the light ray. This path may have minimum time, however the path is not a stationary point. A slight variation of the path into the high refractive index material would cause a discontinuous derivative of the time taken by the path (singular point, not stationary point). This means that neighbouring paths have first order variations in Action (therefore phase) and do not interfere constructively. Path I is a local minimum (stationary point) and there are no first order variations in the Action associated with a small variation in the path, hence neighbouring paths constructively interfere and it is the observed path.

Explore All Paths

In Quantum Mechanics a particle takes all possible paths when moving between two states

Expert Agreement: 93%
Inventory Questions which address this concept: Q8

This principle is one of the basic axioms that underlies Feynman's many path approach to quantum mechanics. In quantum mechanics a particle takes all possible paths, in classical physics only paths with stationary Action are observed because they are the only paths which interfere constructively with neighbouring paths.

Originally I thought there would be very few misconceptions around the 'Explore All Paths' concept. Once being exposed to the concept, students would simply understand that particles take all possible paths in quantum mechanics.

Initial data, results from the pre-test and a lecture question from the 3rd Action lecture, seemed to confirm this hypothesis. Before being introduced to the 'Explore All Paths' concept only 27% of students answered question 8 of the ACI correctly. At the end of a lecture which introduced the many paths approach to quantum mechanics, students were asked to answer a lecture question testing their understanding of the 'Explore All Paths' concept and 88% of them answered correctly.

The limitation with both of these questions (ACI Question 8 and the lecture question) was that they only tested whether or not students had a surface understanding of the 'Explore All Paths' concept. Students were asked 'which of these paths could the particle take?' and the correct answer was to simply respond 'all of them'.

The third extended response homework question probed student thinking about this topic more deeply and showed that students did not truly understand this concept. Although most students would say a quantum mechanical particle 'Explores all Paths', different students had different interpretations of what this actually meant. The common misconceptions included:

- **'Probabaility'** or **'Choose'** misconception - In quantum mechanical situations each path has a probability associated with it. This probability is higher for the stationary path. One consequence of this misconception, is that students believe that increasing the number of paths available always increases the probability of a transition (making destructive interference impossible). Another way students would phrase the misconception was 'macroscopic objects always take the stationary path, while quantum mechanical objects can *choose* from many different paths'.
- **'Least Action'** misconception - Some students believed that the lower the Action of a path the higher the probability it will be taken. This is a combination of the 'Probability' misconception and the classical 'Least Action' misconception described in 'The Principle of Stationary Action' section above.
- The number of paths available for an object to move from point A to B increases as mass decreases. Many students confused the size of region in which constructive interference

occurs (which decreases with mass), to a decrease in the number of paths available for a quantum mechanical object.

- Some students believed that quantum mechanical objects have more stationary paths than macroscopic objects and that the number of stationary paths is proportional to the mass of the object.
- **'Single Path'** misconception - Students with this misconception believe that quantum mechanical particles and only take one well defined path. Students normally have this misconception simply because they have never been exposed to quantum mechanics.

ACI distractor options Q8(a), Q14(a)(d), Q15(a)(b), Q17(b) and Q18(a)(b)(d) correspond to the 'Single Path' misconception and model analysis was used to provide information about the number of students who held this misconception (see Sections 4.6 and 5.4.1).

Complex Amplitude

Each path has a 'complex amplitude' with equal magnitude that may differ in phase, determined by the Action

Expert Agreement: 93%
Inventory Questions which address this concept: Q15, Q18

Originally this concept was simply called 'amplitude'. Multiple Action experts pointed out that this terminology could be confusing and for students who are used to amplitude having identical meaning to magnitude when referring to complex numbers or waves.

In Feynman's method of explaining the complex amplitude he asks students to imagine every path has an arrow (phasor) that represents the phase of that path. This allows students to visualise how the amplitudes of each path add together.

The most common misconceptions are:

- There is the misconception that paths around points of stationary Action are more likely or that the stationary path is more likely. This is not true, all paths have the same magnitude and make the same contribution to the overall probability. Stationary paths and their neighbouring paths simply make a larger contribution because they interfere constructively. Shankar puts this very clearly in his book 'Path integrals' *The classical path is important, not because it contributes a lot by itself, but because in its vicinity the paths contributed coherently.* [66].
- Many students don't realise that each path has a phase and can therefore add constructively or destructively. They simply believe that if some paths are removed the probability of the transition will decrease. A classic example of this is the diffraction grating, by removing some sections of a mirror it is possible to increase the overall intensity of light at a particular point.

Superposition Principle

Total amplitude of a transition from A to B is obtained by summing the amplitude of each path from A to B

Expert Agreement: 100%

Inventory Questions which address this concept: Q15, Q18

The superposition principle and complex amplitude are very closely related. As the complex amplitude (wavefunction) itself is not physical, I chose not to try and separate these two concepts in the inventory. I believe the best inventory questions ask students to make predictions about physical situations, where students who have misconceptions predict an experimentally falsifiable outcome. Therefore the superposition principle and complex amplitude concepts were paired in questions in the inventory. Therefore complex amplitude is not referred to explicitly in the inventory and the two corresponding questions have an experimental basis.

Probability

The probability of a transition is calculated by squaring the modulus of the total amplitude

Expert Agreement: 100%

Inventory Questions which address this concept: Q14, Q17

This is also known as the Born Rule after Max Born published 'Zur Quantenmechanik der Stoßvorgänge' ('On the Quantum Mechanics of Collisions')[67] in 1926.

Originally I also wanted to emphasise the experimental side of this concept and it read 'The *intensity* of a transition can be found by squaring the modulus of the amplitude.'. Intensity was chosen because it emphasised that this concept is used to make experimentally falsifiable predictions. However this wording was criticised by Action experts and taking on their advice I used the more accurate description 'probability'. However intensity was used as a way of probing student understanding of probability in the Action Concept Inventory itself.

The most common misconception was famously published in Born's paper in 1926. This misconception is that you simply take the magnitude of the complex amplitude without squaring the total complex amplitude (wavefunction). He inserted a famous footnote on the third page of his paper, correcting his misinterpretation. The English translation is: 'More careful consideration shows that the probability is proportional to the square of the quantity $\Phi_{n,m}$'. It is not too much of a stretch to say this correction won him the Nobel Prize in 1954 for his interpretation of the wavefunction [68].

Conservation of Energy

Energy is conserved

Expert Agreement: 79%

Inventory Questions which address this concept: Q1, Q2

This concept was presented by Edwin Taylor and Jozef Hanc [44] as a scaffolding concept that students could use to start practicing solving problems of particle motion in a one dimensional potential.

The original wording tried to emphasise the use of the principle for introductory classical mechanics, *'The energy of a frictionless system does not change over time'*. It received criticism from the ANU Action experts because of the use of the word 'frictionless' is technically incorrect. This could be corrected by changing the wording to '*The total energy of a non-dissipative system does not change over time*', but this could just be seen as a definition of dissipation (or a conservative system). One ANU expert who supported the inclusion of conservation of energy suggested that dissipation is an important concept in itself. It was clear that a key component of this concept was the ability to identify a non-dissipative system and question 2 of the Action Concept Inventory was devoted to this.

One ANU Action expert raised the question "If conservation of energy is included, why not conservation of momentum?". My response is, the inventory is meant to test students' understanding of Action, an energy based concept rather than a force based concept.

ANU Action experts also made the suggestion that this concept should test whether students understand how time symmetry leads to conservation of energy through Noether's theorem. This concept was considered a separate concept, 'Noether's Theorem', which was eventually rejected by the Action experts.

Principle of Stationary Potential Energy

A particle placed at rest at a point of zero slope in the potential energy curve will remain at rest

Expert Agreement: 79%
Inventory Questions which address this concept: Q3, Q5

This was suggested by Taylor and Hanc as a way of using potentials in conjunction with frictional forces and the tendency of systems to move toward increased entropy. [44]

One ANU expert challenged this statement saying that it is simply not true, small particles and even macroscopic objects move around due to thermal fluctuations. However when teaching Newtonian mechanics in an introductory physics course, complexities such as thermal fluctuations are usually ignored, there is no apparent need to include them when teaching introductory physics through Action. Although this principle may not be technically correct it can still be useful especially when considering systems which exhibit friction.

The same Action expert stated that this is technically incorrect, as someone could apply an external force to the ball (for example, hit it with a golf club) and then the ball would obviously begin to move. However if you define force as the gradient of the potential (all fundamental forces can be formulated in this way) it is not possible to apply a force to the ball without changing the potential and removing the stationary point.

This principle can be extended further, the stability of the 'equilibrium points' can be determined by the nature of the stationary points on the potential curve. For a local maximum or an inflection point the equilibrium point is unstable or arguably semi-stable in the case of the latter. For a local minimum, the equilibrium point is stable and is a final resting point for a dissipative system as it slows down and stops.

The two questions in the inventory which test this principle. Question 3 tests whether students can correctly identify stationary points. Question 5 ensures that students fully understand what properties these stationary points have.

3.4 Rejected Concepts

Concepts in Table 3.2 were not included in the final concept inventory. The reasons why each of the concepts were not included are explained in this section.

Table 3.2: Table summarising expert responses to concepts that were not included in the Action concept inventory. Concepts from the survey have the number of experts who agreed the concept should be included on the inventory (as a %). Concepts suggested by Action experts as missing, have the number of experts who suggested inclusion.

Concept	% Agreement	Number of Experts who suggested inclusion
Action	N/A	1
Equivalence	N/A	2
Generalised Co-ordinates	50%	N/A
Noether's Theorem	57%	N/A
Principle of Least Kinetic Energy	50%	N/A
Maupertuis Action Principle	N/A	2
Momentum	N/A	1
Principle of Maximal Aging	N/A	1

Hamiltonian Mechanics, Euler-Lagrange equations, Variation Calculus, Legendre transformations and Lagrange multipliers were all suggested as concepts to be included by at least one Action expert. These are all mathematical tools that can be used to solve problems. The purpose of a concept inventory is to test student's conceptual understanding of the material not their mathematical ability. Therefore none of these concepts were considered for inclusion on the Action Concept Inventory.

Each of the concepts that was included on the orignal draft concept list (sent to the experts for review) and rejected is examined in detail below. In addition any concepts that the experts suggested were missing from the list have been explored.

Action

The Action associated with a path is the product of the path's elapsed time and the time average of the Lagrangian along the path.

Originally 'Action' was suggested as one of the core concepts. However it became clear that this is an equation or a mathematical definition rather than a concept. It is much better captured in the equation for Action than in words.

$$S = \int_A^B (T - V)dt$$

For this reason it was not included in the final concept list.

Principle of Least Kinetic Energy

In the absence of a potential a particle will take the path for which the time average of its kinetic energy is minimized. [44]

This is a scaffolding principle suggested by Jozef Hanc and Edwin Taylor [44] that can be used to allow students to explore worldlines and make calculations before introducing them the Principle of Stationary Action. It gives students a chance to become familiar with possible paths and space-time diagrams before being confronted with Action.

Although this may be a useful pedagogical tool, Action experts suggested that there was nothing contained within this principle that was not already in the Principle of Least Action, it was simply a special case. Usually the most difficult concepts are the ones chosen to be tested in a concept inventory, this allows the inventory to be used as a measure of student mastery [25]. Because of its simplicity and low expert agreement, the principle of kinetic energy was removed from the final concept list.

Global and Local

If a path has stationary Action, each small section on that path must also have stationary Action

Students who do not understand that stationary Action is a local principle as well as a global principle may have misconceptions about causality [69]. For example they may think that a particle somehow magically sniffs out the entire path from A to B in advance, violating causality. One Action expert suggested that this is was a common misconception of academics who regularly use Lagrangian mechanics as a computational tool, but have not thought deeply about Action principles.

The resolution of this misconception is fairly simple. Causality is not violated because the potential must be specified at each point along the path in order to calculate Action. There is a local potential that governs the motion of the ball. Then to fully specify the path boundary conditions are required. In the case of Action principles it is more natural to specify a starting and ending event rather than an initial position and velocity (as you would with Newton's laws).

Two Action experts suggested that this was equivalent to realising that Lagrangian Mechanics is equivalent to Newton's laws, if you define force as the negative gradient of a potential. However this is not a good reason to include the concept in the inventory. If a student thought that the Principle of Stationary Action produced different results to Lagrangian mechanics or Newton's Laws, this would undoubtedly be a grave misconception. However this is not an Action specific concept, instead it probes student's understanding of science in general. Both theories must be consistent with experiment, a basic principle of science, therefore the theories must be consistent with each other. The equivalence to Newton's laws is implied in the 'The principle of Least Action' concept.

Generalized Coordinates

A system may be described with co-ordinates that embody the limits on its position due to physical constraints

Generalized coordinates are one of the greatest advantages of Action based approaches to mechanics, as they allow you to reduce the dimensions of a problem. Although generalised co-ordinates definitely aid student's ability to apply the Principle of Stationary Action, Action experts pointed out that it is not a necessary concept to understand the Principle of Least Action. For this reason this concept was removed from the final inventory.

A second concept originally accompanied generalized coordinates, '*d'Alemberts principle*', constraining forces do no work. This concept arises as the natural benefit of generalized coordinates for someone used to force based approaches to mechanics, "Why use Lagrangian mechanics? Because it allows you to ignore constraining forces". However this concept is irrelevant for students who have not learned Newtonian mechanics and who have not heard of forces. It is not fundamental to Action instruction, rather plays an important role extending classical mechanics from a Newtonian approach. Since the Action approach teaches principles of Action and energy before forces 'd'Alembert's principle' was not included.

Noether's Theorem

A continuous symmetry leads to a corresponding conservation law

Noether's Theorem uses Action based principles to derive the conservation laws, some of the most fundamental laws of physics. The ability to emphasise symmetry and teach Noether's Theorem is one of the greatest advantages of teaching physics through Action. All ANU experts agreed that Noether's Theorem should be included in the Action concept inventory.

One expert asked if 'continuous symmetry' should be a separate concept. This can be simply thought of as a mathematical definition. However it was noted that this could be a misconception of Noether's theorem, that any symmetry implies a conservation law.

Like generalised co-ordinates, many Action experts pointed out that Noether's Theorem is a useful extension of the Principle of Least Action but not necessary to understand the Principle of Least Action. For this reason, Noether's Theorem was removed from the final concept list.

Conservation of Momentum and Canonical Momentum

One Action expert suggested that if conservation of energy is to be included then so should conservation of momentum. However Action Physics emphasises energy which plays a central role, while force and momentum are secondary.

Momentum appears in both Noether's Theorem and Lagrangian mechanics in the form of conjugate (canonical) momentum. Since neither Noether's theorem or Hamiltonian mechanics featured on the concept list, canonical momentum was not considered important enough to include.

Principle of Maximal Aging

A free particle will take a path between two events that maximises its proper time

One advantage to teaching physics through Action is the simple generalisation to both special and general relativity through the principle of maximal aging. However to include this on the inventory as a standard Action concept created all kinds of complications, because then the concept of 'proper time' would have to be introduced.

All the relativistic concepts were rejected on the grounds that they were an extension to the Action principle but not fundamental to it. Most instructors who wish to include an Action unit in their introductory physics courses will not have time to introduce their students to relativity as well, therefore any relativity questions would not be useful for measuring the effectiveness of their instruction.

Maupertuis' Principle of Stationary Action

Two international experts suggested including Maupertius principle as a concept to be included on the concept list. Although the generalised Maupertuis principle is equivalent to Hamilton's principle, it was finally rejected for the following reasons:

- Hamilton's Action is the central principle that was emphasised by Richard Feynman and Taylor in their introductory physics explanations. It is possible from this fundamental principle to derive Lagrangian Mechanics and the Schrodinger equation. In addition Noether's theorem is usually derived from the Lagrangian and Hamilton's Action principle.

- Hamilton's Action has been suggested as a new point to start teaching introductory physics because it creates a clear link between quantum mechanics, classical mechanics and relativity. The Maupertuis principle does not do this.

- Maupertuis Action ties in very well with both canonical momentum and generalised coordinates, however a majority of international Action experts disagreed that these should be included on the inventory because they believed these were extensions to Action principles but not fundamental to them. Had these concepts been included in the inventory it would make a much stronger case for including Maupertius Action principle.

3.5 Creating the Action Concept Inventory and Development Interviews

Initially some draft questions were created to probe student thinking about Action concepts. I consulted psychology literature about common traps to avoid when creating questions; to prevent students from cheating on the Action Concept Inventory.

3.5.1 Preventing Students from 'Gaming' the Action Concept Inventory

Some students with good test-taking ability may be able to determine the correct answer without understanding the concepts. In 1964 Gibb [27], [70] identified 7 criteria that can be used to evaluate the 'guessability' of a question:

1. Phase-repeat - Correct answer contains a keyword or phrase that is contained in the question

2. Absurd Relationship - Distracters are unrelated to the question

3. Categorical Exclusive - Distractors contain words such as 'all' or 'every'

4. Precise - Correct answer is more precise, clear or qualified than the distracters

5. Length - Correct answer is longer than the distractors

6. Grammar - Distractors do not match the verb tense of the question or a mismatch between articles ("a", "an", "the")
7. Give Away - Correct answer is given away by another item in the test

When creating the Action Concept Inventory I checked each question against Gibb's criteria to ensure that it would be difficult for students to cheat and guess the correct answer unless they understood the relevant Action concept.

3.5.2 Developing the Inventory through Think Aloud Interviews

Think aloud interviews were used to identify common misconceptions about Action physics and therefore to create distractor options for the multiple choice test. The think aloud interviews were carried out with students of different levels of understanding ranging from no science background through to physics PhD students and even one Action lecturer. This allowed me to find the key differences between student thinking and expert thinking.

Performing think aloud interviews with as wide a range of students as possible also plays a key role in validating the concept inventory. Students from a range of ages, physics backgrounds, ethnicities and genders were tested to check that they interpreted the questions correctly.

A summary of the types of students who took the concept inventory is shown in the Table 3.3 and Table 3.4.

Table 3.3: Demographic Information of the Students who participated in the Development Interviews 1

Gender	Ages	Nationality	Occupation	Primary Discipline of Study
Female - 10	18 Years -1	Australian - 19	School student - 1	Physics
Male - 21	19 Year - 6	New Zealand -3	Undergraduate - 27	Maths -2
	20 Years - 8	Malasian -2	PhD Student - 2	Computing - 1
	21 Years - 6	Russian - 1	Academic - 1	Other Science - 4
	22 Years - 6	Chinese - 1		Arts - 3
	25+ Years - 4	American - 1		Engineering - 1
		Indian - 1		
		Mixed Australian and Other - 3		

After observing misinterpretation of questions, the wording of some questions was adjusted. This process of checking students were correctly interpreting the inventory helps validate the inventory. It checks that students are choosing the correct answers because they used correct reasoning and were choosing incorrect options because they held misconceptions. This ensures that the inventory is actually measuring what it is intended to measure, establishing validity.

Table 3.4: Demographic Information of the Students who participated in the Development Interviews 2

Average Grade (self reported)	Living Status	Highest Level of Physics Studied	Topics of Physics Studied	Other
HD - 16 Distinction - 8 Credit - 5 Pass - 0 NA - 2	On Campus - 13 Canberran - 12 Elsewhere - 6	PhD - 1 Masters - 1 Honours - 8 3rd Year - 2 2nd Year - 9 1st Year - 2 Year 11 or 12 - 2 Year 10 - 6	Taught Action - 3 Action Physics - 19 Studied Quantum Mechanics - 21	Dislexic - 1 English is Second Language - 2

An example of some of the changes made to a question as the result of the interviews is shown in Chapter 8 - *Appendix 1: Evaluating Question 13*.

Think Aloud Protocol

There is a strict, standard protocol to investigate student thinking called 'think aloud interviews' [25]. The student is told to think aloud while completing a task (in this case the Action Concept Inventory, See Chapter 10 - *Appendix 3: The Revised Action Concept Inventory*). When performing think aloud interviews the intention is to try and probe student thinking without altering it. The interviewer must be careful to avoid cueing students to think or respond in a certain way. The interviewer can remind the student to think aloud if they become quiet, however they must never ask the student to explain how they interpreted each question because this is likely to alter (and usually improve) the thought process [25].

Wieman and Adams [25], suggest that interviewers should be restricted to prompting students to think aloud when they become quiet. They do suggest that in open ended interviews the interviewer may need to probe certain areas of thinking more directly. For example asking a student to explain their answer or to say more about what a particular term or concept means to them. This will allows the interviewer to gain an understanding of what the student is thinking, while minimising the impact on their thinking.

Post Test Questioning

Once the test is in multiple choice format, Wieman and Adams suggest that strict think aloud protocol must be used in order to avoid tempering student thinking [25]. However there were some cases where I wanted to probe student thinking and better understand the reasons why students were choosing distractor options. In these cases I used a fairly simple idea, which I call "post test questioning".

During the interview I would make note of questions where I wanted to ask the students why they chose their answer. Then either after the student had turned the page, finished a section or completed the inventory I would ask them about the question.

One example of when I used this technique was after a student had finished answering question 8 (See Chapter 10). The question had the instructions 'Look carefully at the axes of the plots', however the student's think aloud comments made it unclear whether they had interpreted the plots correctly as position against time or incorrectly as position against position trajectories. I waited until the student had given their answers for both question 8 and 9 and turned the page before asking them "What where the axes of the plots on the previous page?"

This allowed me to probe student thinking and gain the understanding that I needed without interfering with the student's thinking in Question 9 or causing them to change their answer. In this particular case the student had misinterpreted the plot. Despite the 'Look carefully at the axes of the plots' comment, many students misinterpreted the axes on this plot. Therefore I changed the wording of the question so that **'Look carefully at the axes of the plots'** was in bold. In all interviews conducted after the bolding of this text students interpreted the plots correctly.

Summarising Interviews

Wieman and Adams suggest that immediately after an interview an interviewers should spend between half an hour to an hour summarizing the interview [25]. However given that I needed to perform more than 20 interviews in a time period of two weeks, I found a more effective way of summarising interviews which took less time.

The two purposes of the interviews were to identify common student misconceptions (and therefore create distractor options) and check that students were interpreting the questions correctly. So rather than making notes at the interviews, I would instead have a personal copy of the test on which I would make a note whenever a student had a misconception or misinterpreted a question. After a number of interviews I would have an organised summary of the common misconceptions and which questions needed to have their wording improved.

This was a more time efficient method of achieving the goals of the interviews.

3.6 Action Instruction, Observation and Homework

To gather as much qualitative data about the students as possible I attended most of the PHYS1201 Action physics lectures and tutorials to observe students. In addition to observing the tutorials, I asked each of the PHYS1201 tutors to make note of the common misconceptions and student difficulties they observed. A description of the lectures, tutorials, homework problems and lecture notes is contained in Chapter 9 *Appendix 2: Action Instruction in PHYS1201*.

Students had to answer conceptual extended response (200 word) questions as a part of the homework for the Action unit of PHYS1201. The three extended response questions are described in Section 9.5. I marked each of the extended response questions and student responses revealed a number of misconceptions, particularly concerning the 'Explore All Paths' concept (see Section 3.3).

As a part of the final homework students were asked to complete a survey about their experience of the Action unit of the course.

3.7 Validation Interviews

After the first year students had completed both the pre-test and post-test, I went to the lab classes and asked if any first year students would volunteer to participate in validation interviews, so I could investigate the reasoning behind their answers. As an incentive, I offered to answer any questions the students had about Action. The response was overwhelming, over fifty first years signed up to participate in the interviews. Due to time constraints I was only able to interview 19 students, which is still more than Wieman and Adams say is necessary in their paper summarizing best practice for developing concept inventories [25].

All 19 of the students who participated in validation interviews were students from the PHYS1201 class who had participated in both the pre-test and post-test. Table 3.5 below provides a summary of the demographic information of the students who participated in the interviews.

Table 3.5: Demographic Information of the Students who participated in the Validation Interviews

Gender	Age	Nationality	Living Status	Grade in PHYS1101
Female - 4	18 Years - 6	Australian - 11	On Campus - 8	HD - 5
Male - 15	19 Years - 6	New Zealand - 1	Canberran - 9	Distinction - 4
	20 Years - 3	Malasian - 2	Elsewhere - 2	Credit - 5
	21 Years - 1	Chinese - 3		Pass - 2
	23 Years - 1	Eurasian - 1		NA - 2
	26 Years - 1	Singaporean - 1		
	27 Years - 1			

Table 3.5 shows that all student demographics are well represented. Female students are slight under-represented. Besides this there is the appropriate diversity in student nationality, living status and student ability (HD students are only slightly over represented).

The validation interviews had two main purposes. First of all it was a chance for me to find any common misinterpretations of the Action Concept Inventory Questions. Secondly as shown in the results chapter, the statistical analysis showed the first years didn't understand the quantum mechanical concepts, even on the post test. There were two possible conclusions:

1. The quantum mechanical section of the inventory was invalid and the first year students really did understand quantum mechanics
2. The inventory was valid and the first years did not understand the quantum mechanical concepts

I suspected that the fraction of a lecture that Craig Savage spent teaching quantum mechanics simply was insufficient students to gain a proper understanding of quantum mechanical concepts, but performing the validation interviews allowed me to confirm this.

The validation interviews consisted of three sections. Firstly I asked the students to complete the Action Concept Inventory in the think aloud format described previously. After students had completed the inventory I asked them two further questions as a kind of oral exam to check how well students understood the quantum mechanical principles that underpin Action.

Think Aloud Interviews in Validation

The long response question of the final homework was based on a problem that the first year students completed in tutorials. In the tutorial problem the students were asked to calculate the Action for the two paths for both an electron and a 1kg mass.

One path was a straight line path, which was stationary for both the electron and the 1kg mass. There was also a parabolic path which had greater Action that the straight line path for both the electron and the 1kg mass. However the difference in Action between the two paths is many orders of magnitude larger for the 1kg mass than for the electron.

Students were asked which of the following statements they thought were true:

- The electron takes more paths than the 1kg mass
- The electron has more stationary paths than the 1kg mass.
- There is a region of constructive interference for both the mass and the electron. The region is larger for the small mass.
- There is a region of constructive interference for the electron. Only the stationary path is taken by the large mass.
- The electron could 'choose' to take many different paths, the mass takes only the stationary path.

Student responses to this question showed extremely inconsistent thinking. This could be explained by two competing models in the student's minds. A well understood classical model and an unfamiliar quantum mechanical model which they believed would give them the correct answer.

Students used phrases such as "I believe the electron could take more paths than the 1kg mass" or "If the masses use this path" which demonstrate they believe that electrons and masses only take one path, not a superposition of paths. When students talked about probabilities of paths, this indicates they think only one paths was taken. For example, one student said "The probability of this path is greater for the electron".

Explore all Paths True or False Interview Questions

The first question was designed to measure the student's understanding of the 'Explore all paths' concept. It specifically targeted misconceptions that arose in the final Action homework.

Opened Ended Interview Question

The purpose of the open ended question was to gauge the students' understanding of quantum mechanical concepts. The original format of the opened ended question was not very successful. I would show the students a list of the four quantum mechanical concepts that they were taught in the course. I then asked students which of the concepts they thought they understood. If they did understand a concept then I asked them to explain it to me (to see if they really did understand).

The problem was that many of the students did not recognise the concepts by their names. A few students tried to explain some concepts but were confusing which was which, they were able to tell me some of the things they had learned in quantum mechanics but were unable to define each of the terms.

Inspired by the oral exams of Matthew Sellars, (an ANU academic who teaches a third year course and an honours level course), I decided to change my questioning method. As a student I took both of Matt's courses and each course had an oral exam as part of the assessment. I interviewed Matt about how he conducted his oral exams. He told me the reason he likes oral exams is that if a student gets stuck then you can help them along and see the full scope of their understanding.

Matt grades students in oral exams according to the amount of help he gives them to complete the exam. While he asks the students questions, he notices how much help he has to give the students and distributes grades accordingly. In the open ended question I tried to follow Matt's format, I was not assigning students a grade, but I would make a mental note of all the help I gave students and made a note of all the misconceptions displayed. This process allowed me to gauge understanding of the quantum mechanical concepts. I always performed this part of the interview last as it would significantly alter student thinking on the topic (it was probably beneficial for the students and supported their learning).

3.7.1 Multiple Students in Think Aloud Interviews - A Possibility for more Reliable Measurement

Three of the think aloud interviews I performed were with pairs of students rather than interviewing students individually. This made the think aloud interview format much more natural for the students, being able to discuss their reasoning with each other, rather than simply speaking at an unresponsive interviewer. Although the students would alter one another's thinking, I would still see each individual's thinking and see the common misinterpretations as they discussed the question.

On the occasions where one student misinterpreted a question, discussing the question with the other student corrected this misinterpretation. Students would also discuss what assumptions they should make while answering the questions. This greatly reduced the number of questions the students answered incorrectly due to misinterpretation, effectively increasing the reliability and validity of the test.

Observing this in student interviews gave me a new idea for future research using concept inventories. Not only could think aloud interviews be performed with pairs of students (to make the experience more natural for the students), but inventory pre-test and post-test could be completed in pairs. One inventory would be shared by two (or maybe three) students who would discuss their thinking out loud. During these discussions students would point out where the other student misinterpreted a question and have a discussion about what assumptions to make. This would increase the reliability of the inventory and provide more accurate information about a class's level of understanding and therefore effectiveness of instruction.

3.8 Expert Review

Each of the Action experts who completed the online survey about the concept list, was contacted and asked if they could review the inventory.

Four of the Action experts reviewed the inventory, making suggestions for improvement. As Adams and Wieman predicted [25], the Action experts made suggestions to make questions more precise and technical than they needed to be for students to interpret them correctly. The results of the expert review can be found in Section 5.5.2.

In this chapter the process of developing a concept list and a valid and reliable concept inventory was described. In addition we have explored the qualitative methods for developing a concept inventory including expert review and think aloud interviews. Quantitative methods of validating concept inventories through statistics are examined in the next chapter *Statistical Analysis*.

Chapter 4

Statistical Analysis

I start this chapter with a special thanks to John Aslanides and Craig Savage who developed the Relativity concept inventory in 2012 [71] [56]. John Aslanides outlined a number of procedures that I used to establish the validity of the Action Concept Inventory. Thanks to his work I was able to quickly learn all the statistical methods he used and spend time critiquing and extending them.

In this chapter not only do I explore the standard techniques used to evaluate concept inventories in Section 4.2, but I also take an in depth look at the Monte Carlo analysis developed by John Aslanides [71], see Section 4.3. I then outline the Rasch intelligence model (Section 4.4), Factor Analysis (Section 4.5), Model Analysis (Section 4.6) and discuss best practice for using student self reported confidence (Section 4.7).

4.1 Why use statistics?

As mentioned in the previous chapter, once a concept inventory has been developed its primary purpose is to measure the effectiveness of different teaching methods, in order that physicists may optimise their teaching methods. To use the inventory to measure student learning we need a way of analysing the results. One may imagine that you could simply look at the increase in student total scores to measure effectiveness of teaching, but as discussed later, even this is not completely straight forward. Physicists may be interested in students' learning of particular concepts, in which case they would need a systematic method of breaking the test into subsections. If different physicists are trying and comparing different teaching methods it is essential that the statistics they are reporting are calculated according to a standard set of procedures for consistency.

In my project I use statistics not only to measure student learning but to establish the validity and reliability of the Action Concept Inventory. In this chapter I outline the standard statistical measures that are used to evaluate concept inventories. I also investigate the reasons why these different measures are used and discuss some more advanced and less frequently used statistical methods and their place in evaluating concept inventories.

In my analysis I am trying to use a single data set to both validate the inventory and make inferences about the students who took the inventory. Because I am doing this I need to be extremely careful about how I treat the data and the assumptions I make. All inferences I make about the student population who took the inventory are supported by other evidence including 'think aloud' interviews, student homework scores and observation of the students in lectures and laboratory classes.

4.1.1 Basic Statistics

I am going to assume the reader is familiar with some basic mathematical statistic procedures, so I will not outline them in this thesis. If the reader wishes to know more about any of these topics there are many textbooks and online resources available which explain, see [72, 73] and references therein.

- Mean, Standard Deviation and Standard Error
- Linear Regression
- Variance and Co-variance
- Pearson's correlation r
- P-values (note it was only appropriate to use one-tailed p-values in this analysis)
- Binomial Distribution

4.1.2 Validity, Reliability, Discrimination and Convenience

Validity, reliability and discrimination are three measures that are often considered when developing a concept inventory.

- Validity refers to the ability of the test to fulfill its purpose. A valid test measures what it was intended to measure.
- Reliability refers to the consistency of a test. A student would get the same score every time they sat a reliable test (assuming they did not learn in between tests).
- Discrimination refers to a test's ability to discriminate between individuals. If the test scores are well spread then a test has good discrimination, if the test scores are all clumped together then the test has poor discrimination.
- Convenience refers to how easily a test can be delivered and graded.

The purpose of the inventory is *'to measure how well students understand the concepts that underpin the Principle of Stationary Action.'*. Validity is a measure of how well the test achieves this purpose. For this reason I would say that validity is the most important of all these measures. The other properties are nice extras, but if the test does not achieve its purpose, I would debate that it is useless, regardless of how reliable it is or how easy it is to deliver.

Validity

A valid test measures what it was intended to measure. Referring back to the ruler analogy I used in the introduction, a valid ruler measures length, not mass or temperature or anything else. Due to thermal expansion a ruler will never be a completely valid length measuring device unless temperature is very well controlled. Similarly an inventory question will never be a perfect measure of a particular concept, you will always find fluctuations when students misinterpret or misread a question. However you can try to make wording and diagrams as clear as possible to avoid common misinterpretations.

The Action Concept Inventory was designed to measure 9 different concepts. Each question was designed to measure one specific concept. If a question does measure a student's understanding of that concept then it is a valid question.

In an ideal world if a student answered a question which was valid, they would always answer it correctly if they understood the concept. Ideal distractors would be powerful enough that students always answer the question incorrectly if they did not understand the concept.

Validity is normally established by checking correlations between items in the inventory which test the same concept or between inventory items and other previously established measures of the same concept (exams, homework and interviews). Since validity is so important, a large portion of the statistical analysis I have performed examines the validity of the inventory including the Monte Carlo Simulations, Factor Analysis (principle component analysis) and fitting the data with the Rasch model of intelligence (logistic regression).

Reliability

Reliability refers to an instrument's ability to give consistent results when used to measure the same object. A reliable tool is one for which the measurement error is small. As mentioned in the introduction a reliable ruler would measure a piece of string to be exactly the same length every time. Unfortunately it is impractical to measure student understanding over and over again to check the reliability of an inventory.

Therefore in psychometrics, reliability is measured in two specific ways. The first is test-retest reliability, a test should reproduce the same score for each student on retesting (assuming the students do not change/learn between tests) [74]. If we assume that students learned about the concepts on the Action Concept Inventory between the pre-test and post-test, then test-retest reliability cannot be used as a measure of reliability of the test. However there may be reasons to assume that students did not learn anything about certain concepts on the inventory (these concepts were not covered in class or there was no improvement of student scores on particular questions) in which case test-retest reliability could be used to establish the reliability of some of the questions.

The second quantitative method for calculating test reliability is internal consistency. It makes sense to use internal consistency as a measure of reliability in the case of an IQ test where each item is trying to measure the same variable. However in the Action Concept Inventory 9 different concepts were being tested and there is no reason why we should assume all the questions should correlate with each other for a truly valid inventory (which measures understanding of specific Action concepts, not general intelligence or physics ability). However correlation between items testing the same concept could be used to establish the reliability and validity of these items.

I made an argument for why it is not appropriate to use quantitative measures of internal consistency of an entire test to measure the reliability of the concept inventory. I still calculated these measures of reliability (see section 4.2.4), however I treated them with caution when using them to draw conclusions.

Discrimination and Difficulty

A ruler would not be able to distinguish between human hairs of different widths, it is of the wrong scale. A micrometer would be a much more appropriate instrument to measure such thin objects. When using a concept inventory to measure student learning it is also important to ensure that it has the appropriate scale. If all students answer all questions correctly both pre and post test then you have no useful information about the learning of the class or the effectiveness of instruction. To ensure a concept inventory has the appropriate scale, two measures are used, discrimination and difficulty.

Discrimination is the ability to distinguish between individuals and differentiate students according to a particular variable (which could range from academic ability to extroversion depending on the test). In the case of the Action Concept Inventory, discrimination is a test's ability to distinguish between individual student's understanding of Action concepts.

It is clear why this is an important statistic for tests whose purpose is to differentiate between individuals for university admissions or job applications. However the purpose of a concept inventory is not to distinguish between individuals, rather it is to measure the understanding of the entire class as a whole and therefore better improve instruction. For this reason I would argue that discrimination between individuals is not an important statistic when it comes to evaluating a concept inventory; in fact it could interfere with validity.

For example if a valid test is used to measure a first year class who do not have an understanding of the subject, the class should receive an extremely low score (ideally every student would get 0). This would mean that the test has a very low discrimination, however it would be extremely valid. In section 4.2, I explore the arguments used by different authors in more detail, exploring the conventional wisdom concerning standard values for item difficulty and discrimination in the case of the Action Concept Inventory.

Convenience

Convenience is not commonly reported or referred to in the concept inventory literature. I believe this is an oversight. Convenience is simply a measure of how easy a test is to deliver and grade. Convenience is one of the key reasons that concept inventories are delivered in multiple choice format.

Although there is no established formula for calculating the convenience of a test, some obvious indicators are the length of the test (how long it takes to distribute and complete the test) and how easy it is to mark.

Technology has advanced rapidly since the Force Concept Inventory was developed introduced in 1991. Inventories no longer need to be distributed in paper form and marked by hand, instead they can be delivered online and marked automatically, greatly increasing the convenience of using these tests. Despite this convenience, most inventories are still only available in pdf format by contacting the authors [57, 58, 59]. Concept inventories would be far more convenient for instructors to distribute and mark if they were available on websites such as WebAssign [75] or MasteringPhysics [76].

One conflict which Wieman and Adams mention briefly in 'Development and Validation of Instruments to Measure Learning of Expert-Like Thinking' [25] but is otherwise almost unmentioned in the physics concept inventory literature is the trade off between reliability and convenience of delivering an inventory. It is accepted that reliability increases with test length as any errors caused by students misreading or misinterpreting questions will have a smaller impact on overall score [77, 74]. However the convenience of a concept inventory decreases with length. There are concept inventories, for example the Electrical Circuits Concept Evaluation, which take 60 minutes to complete [59].

If an instructor wishes to use such an instrument to evaluate their teaching, they would need to sacrifice two hours (60 minutes pre-test and post-test) of teaching time or choose to deliver only a fraction of the test. If the test is longer than 50 minutes there is an additional complication that many universities only allocate 50 minutes for each lecture, making it impossible for students to complete in a single lecture.

In the first case two hours of teaching time is lost and many students may get tired or bored and this may compromise the validity of the instrument. In the second case, where only a fraction of the questions are given to the students, the shortened instrument has not been evaluated or had its reliability or validity established.

Convenience will be mentioned throughout this chapter, as there is a trade off between convenience and other variables such as reliability and discrimination. The Action Concept Inventory was designed to be delivered in 30 minutes or less in an online format, making it easy for an instructor to deliver in a single lecture.

4.2 Classical Test Analysis - The standard Statistics of Developing Concept Inventories

There are a number of standard measures used to calculate the reliability, validity and discrimination of concept inventories [27, 56]. These measures of the effectiveness of an inventory are known as test theory and are outlined in the sections below. To evaluate each item on the inventory, the gain, discrimination and difficulty index were calculated.

4.2.1 Difficulty Index

Difficulty index is simply the fraction of students who answered a certain question correctly. The symbol most commonly used for difficulty index in physics education research is 'd' however 'p' is sometimes used in psychology literature [78] (however this can cause confusion with probability p-values that are calculated to test statistical significance). Difficulty index is calculated by:

$$d = \frac{N_c}{N} \tag{4.1}$$

Where N_c is the number of students who answered the question correctly and N is the total number of students. Therefore the higher the difficulty index, the easier the question.

Some physics education researchers (such as Robert Beichner) state that it is desirable that the difficulty index for each item should fall between 0.3 and $0.7 - 0.9$ [79]. The reason that these ranges are appropriate is that they allow for good discrimination between students.

Robert M. Kaplan and Dennis P. Saccuzzo [80] explore this issue more deeply. They suggest that a test instructor needs to determine the probability than an item could be answered correctly by chance alone. A multiple choice question with four options could be answered correctly 25% of the time by pure chance. Therefore Kapland and Saccuzo suggest that to determine between understanding and random guessing we would require an item difficulty greater that 25% in this case [80].

Kaplan and Saccuzzo further suggest that if an item is answered correctly by 100% of individuals (or 0%) it offers little value because it does not discriminate among individuals [80]. To achieve maximum discrimination they suggest that item difficulty should be halfway between random guessing and the maximum possible score [80]. For example if a multiple choice test has 4 options, the difficulty index that will give maximum discrimination will be $d = \frac{1.00-0.25}{2} + 0.25 = 0.625$ [80].

They then suggest the items should have a variety of difficulty levels, because ideal tests discriminate at many levels [80]. They then go on to say that items in the range $d = 0.3 - 0.7$ maximise information about the difference among individuals[80] (agreeing with Beichner [79]). They then proceed to argue that this depends upon the purpose of the test.

Theresa Kline (note different to Paul Kline) makes similar arguments stating that difficulty indices of 0.5 are ideal for discriminating between participants if questions are uncorrelated [78]. However usually test items are correlated and therefore it is best to have items that have a range of difficulty indexes to maximize discrimination [78].

However I challenge the use of applying these ranges suggested by the test theory literature to concept inventories. Although this range in difficulty of items may be appropriate for entrance exams whose purpose is to discriminate against individuals, this is not a major concern for concept inventories.

The purpose of concept inventories is not to discriminate between individuals in a cohort, but to measure the conceptual understanding of an entire cohort of students (in order to evaluate instruction). In order for an inventory to be valid, it needs to measure student understanding of a given concept. If students do not understand the concept they should get the question wrong. If a question is well designed with good distractor options you should observe ignorant groups of students obtaining scores that are lower than random guessing.

4.2.2 Gain and Normalised Gain

One of the most simple measures of student learning is the average gain in scores of students who completed both the pre and post test. The gain is simply the average increase in score. This is expressed as a fraction of the total score. Therefore the formula for gain is :

$$G = \frac{T_{post} - T_{pre}}{T_{max}} \tag{4.2}$$

Where T_{pre} and T_{post} are the total scores in the pre and post test and T_{max} is the highest possible score. It has become the standard in physics education to use normalised gain, [25, 81]. Normalised gain can be interpreted as the fraction of concepts learned that students did not know prior to instruction. The formula for normalised gain is [82]:

$$G_N = \frac{T_{post} - T_{pre}}{T_{max} - T_{pre}} \tag{4.3}$$

Where T_{pre}, T_{post} and T_{pre} are the scores used in Equation 4.2. Note that G is the normalised gain for a single student, the averaged normalised gain for a class is calculated by averaging the gain for each student. This is different to Hake's normalised gain $< g >$ which is obtained by calculating the normalised gain using the class average scores. Hake discusses the difference between $< g >$ and class averaged G and usually they are within 5% [83, 82].

As a part of item analysis (evaluating each question on the inventory), it is possible to calculate a gain for individual questions. This simply measures the average improvement of a class of students on each question. This gain is given by:

$$G_{item} = d_{post} - d_{pre} \tag{4.4}$$

Where d_{post} and d_{pre} are the difficulty indexes for the pre-test and post-test.

The gain measured for a class depends on many factors. Some major factors that would be expected to affect the gain of a class include length of teaching period, style of teaching (interactive or traditional) and the level of pedagogical research in the particular subject.

Gains can vary enormously depending on these different variables, typically ranging between 0.1 and 0.6. Figure 4.1 from [84] shows the normalised gain on the Force Concept Inventory for a number of different courses.

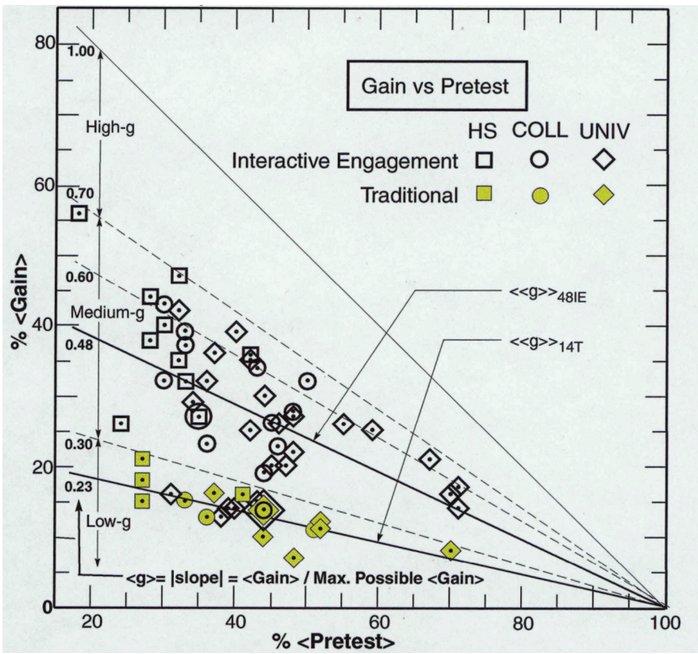


Figure 4.1: Image taken from [84]. This plot shows normalised gains of both traditional and interactive courses on the Force Concept Inventory and a similar test, the Mechanics Diagnostic.

4.2.3 Test Discrimination

There are two standard methods of measuring test discrimination, the first is the discrimination index, the second is Ferguson's Delta.

Discrimination Index

The most simple measurement of test discrimination is the discrimination index, D_i. The discrimination index for a certain question is calculated by taking the difference in number of students who answered a given question correctly, between students above the upper quartile and below the lower quartile:

$$D_i = \frac{4(N_{Hc} - H_{Lc})}{N} \tag{4.5}$$

Where N_{Hc} is the number of students in the top 25% who answered the question correctly. N_{Lc} is the number of students in the bottom 25% who answered the question correctly. N is the total number of students who answered the question.

Discrimination index ranges from -1 to 1. The higher the D_i the more the item discriminates. A positive discrimination index indicates that students in the top 25% answered this question correctly more often than students in the bottom 25%. Negative discrimination index is very unusual as it indicates that students who received a lower overall score did better on that particular question than students with higher overall score. In tests that are designed to measure overall achievement or ability, a negative D_i value indicates a poor item [78]. However in other assessment tools, negative D_i is not necessarily problematic [78] and often allows you to differentiate between different types and groups. A discrimination index of 1 would indicate that every student in the top 25% answered the question correctly while every student in the bottom 25% answered the question incorrectly.

Note that quartiles are not always used. Generally the top and bottom $25\%-33\%$ of students are chosen [78]. The most appropriate percentage to use in creating the extreme groups is 27% as this is the critical ratio that separates the tail from the mean of the normal distribution [78].

When there are uneven numbers of students in each quartile (as was the case in this experiment), it becomes important to normalise for the number of students in each quartile. The formula I used to normalise for this is [78]:

$$D_i = d_u - d_L \tag{4.6}$$

Where d_u and d_l are the difficulty indicies for the top and bottom groups respectively. This can simply be interpreted as the difference in proportion of correct answers between the top and bottom groups.

Ferguson's Delta

Ferguson's Delta [85] (δ) is a frequently used to measure the discriminating power of a test as a whole (as opposed to discrimination index which measures the discrimination power of each item). Ferguson's Delta has been quoted frequently in concept inventory literature (find multiple reference [56]), however I question it's usefulness, see Figure 4.2.

Most sources describe Ferguson's Delta with statements such as "The value of δ varies between 0 (no discrimination at all) and 1 (maximal possible discrimination)" [86] or 'it ranges from 0 (all scores the same) to 1 (each person has a unique score)' [87] [27] [88] [86]. The formula for Ferguson's Delta is shown in Equation 4.7.

$$\delta = \frac{(k+1)(n^2 - \sum_{i=0}^{k}(f_i^2))}{kn^2} \tag{4.7}$$

Where k is the number of questions on the test and n is the number of participants who take the test and f_i is the number of students who received score i.

Generally a test is considered discriminating if $\delta \geq 0.90$ [74], however this is quite easy to achieve and is not necessarily a very good measure of discrimination, see Figure 4.2. Ferguson's delta punishes any score whose frequency is somewhat comparable to the total number of items, favouring tests with a high number of items see Figure 4.2

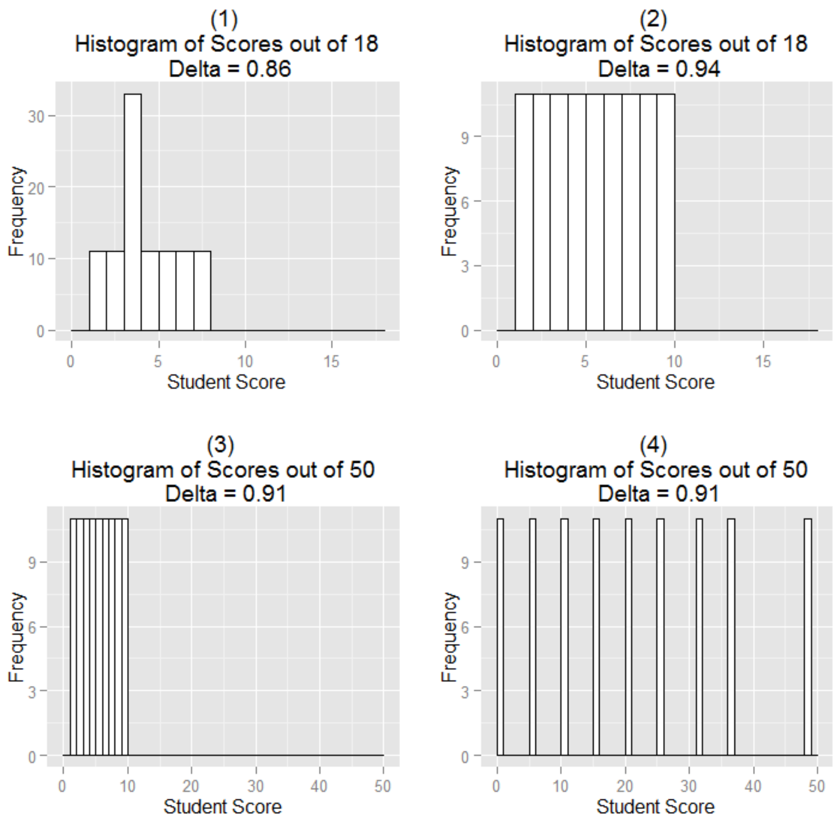


Figure 4.2: Four histograms of different hypothetical distribution of scores. All four distributions have 99 participants and (2), (3) and (4) have 11 students in each of 9 bins. Comparing (1) and (2) shows that Ferguson's delta greatly lowered if many students receive the same score. Comparing (2) and (3) shows that even if a test is more than doubled in length, the same distribution gives a very similar delta, favoring long tests. (3) and (4) show that Ferguson's delta takes no account of how well the data is spread, only the number of bins occupied.

4.2.4 Reliability

There are a number of standard measures that are used to measure the reliability of a test including calculating the point biserial correlation of each item, the Kuder-Richardson Formula 20 (KR-20), the split-half method and test-retest reliability.

Point Biserial Correlation

The point biserial correlation is a simple measure of the reliability of an item. It is the correlation between the overall score on the test and the score on the question. The formula for the point biserial correlation for a question is given in Equation 4.8 below [27]:

$$r_{pbis} = \frac{\bar{X}_c - \bar{X}_i}{\sigma_t}\sqrt{d_i(1-d_i)} \tag{4.8}$$

Where $\bar{X}_c$ is the mean total score of students who answered the question correctly, $\bar{X}_i$ is the average score of students who answered the question incorrectly, σ_t is the standard deviation of the total scores and d_i is the difficulty index for the question.

The numbers generated by a point biserial correlation analysis are simply correlations between the performance on a particular question and overall test score. Positive correlations indicate that a student who has done well on that particular question will do well overall. A correlation of zero indicates that a students score on the particular question had no correlation with their overall score. Negative correlations indicate that a student who has done well on a particular question will do poorly overall (this is unusual).

Kuder-Richardson Formula 20

The Kuder-Richardson Forumula 20 gives reliability of a test as a whole usually denoted r_{tt}. This reliability is calculated using Equation 4.9 [74].

$$r_{KR_{20}} = \frac{k}{k-1}\left(1 - \frac{\sum_i^n d_i q_i}{\sigma_t}\right) \tag{4.9}$$

Where $r_{KR_{20}}$ is the reliability of the test, k is the number of items in the test, σ_t^2 is the total score variance for the test, p_i is the portion of students who answered item i correctly (difficulty index), q_i is the portion of students who answered item i incorrectly (note $q_i = 1 - d_i$).

The split-half method

This method of reliability splits the test into two halves and measures the correlations between student scores on each of the two halves. One critique of this method is that there are many ways to divide the test into two halves, therefore it is possible to calculate many different reliability values depending on how the questions are split [27]. Just by chance, some of these values will be high. When splitting the test it is important to make the two subsections as similar as possible (same number of questions, similar difficulty level). For tests which gradually increase in difficulty it is better to use an odd-even split, rather than first half second half [80].

The test can also be split according to intelligent design, ensuring that each half of the test is similar in difficulty and concepts covered [80]. In the case of the Action Concept Inventory even-odd and first-last splits were calculated. In addition a special split was designed to ensure that question difficulty of each half was even and one question from each pair (relating to the same concept) was on each half of the test.

As mentioned previously, reliability increases with test length [77] [74]. The reliability calculated would actually be an under-estimate because each subsection is only half as long as the full test [80]. To correct for half-length the Spearman-Brown prophecy formula can be applied, see Equation 4.10.

$$r_{corrected} = \frac{2r_{half}}{1 + r_{half}} \tag{4.10}$$

Test-Retest Reliability

Test-retest reliability is a measure of reliability calculated when a test is administered to the same group at two different times. This type of analysis is only of value when measuring traits or characteristics that do not change over time [74].

A group of students from PHYS1201 did take the Action Concept Inventory twice, however for many of the questions there is reason to believe that the students' understanding of the tested concept increased between the two tests (they were being taught a unit on Action). For questions were student understanding improved between the two tests test-retest reliability cannot be used. The most simple method of calculating test-retest reliability of an individual question is to calculate Pearson's r correlation of responses to that question on pre-test and post-test. Test-retest reliability can also be determined with factor analysis (see Section 4.5).

4.2.5 Validity

Validity of an instrument refers to the extent that it measures what it is intended to measure. Validation is the process of accumulating evidence that supports this claim. In test theory there are many different forms of evidence which establish validity including content (face) validity, concurrent validity and construct validity [74, 27].

Content validity

Content validity refers to the extent that the items are representative of the knowledge base being tested and the manner in which they are constructed [77].

In the Action Concept Inventory, content validity was established by creation of the concept list as described in Section 3.3. This involved searching the literature for Action concepts, creating a draft concept list and surveying international experts to establish the validity of this list.

Concurrent Validity

Concurrent validity is assessed by correlating the test with other tests which measure the same variable [74]. However this raises a dilemma. Why create a new concept inventory if there already exists a test (say the final exam) that is good enough to act as a criterion? A concept inventory has the advantage of convenience, it is much shorter, easier to administer and quicker to mark than an exam. Also since a concept inventory is purely multiple choice the grading of answers is completely objective. A traditional exam has the advantage that it can ask open ended questions and check that students can apply the concepts they have learned.

Unfortunately, since Action is not widely taught and there are no well recognised Action exams or tests, concurrent validity becomes difficult to establish. Correlations can be calculated between final Action exam score and concept inventory score. If they correlate positively this indicates that they could both be testing student understanding of Action. Unfortunately students had not completed the final exam at the time this thesis was due, so the results of these correlations are not included in this thesis.

In the case of the Action Concept Inventory concurrent validity was established through performing interviews with students. Ordinary think aloud interviews were performed and these helped established validity and reliability by finding common student misinterpretations of the questions and then rewording them to prevent those misinterpretations. In validation interviews, think aloud interviews were followed with 'oral exam format' questions which were used to check student understanding of quantum mechanics and establish concurrent validity. See Section 3.7.

Construct Validity

Construct validity defines sub-tests in a larger instrument [27]. In the case of the Action Concept Inventory construct validity is established by looking at the correlations between questions which test the same concept.

If two items both test the same concept and if a student understands that concept, they should be more likely to answer both questions correctly. For example, Questions 1 and 2 of the inventory both test conservation of energy. If a student fully understands conservation of energy they should answer both questions correctly. If they do not understand conservation of energy they should answer both questions incorrectly. Therefore we should observe a correlation in students' scores on the two questions.

Determining the statistical significance of these correlations is essential. If you assume Gaussian statistics, you can simply use Equation 4.11 to obtain a t-value and use this to calculate the corresponding p-value.

$$t = r\sqrt{\frac{n-2}{1-r^2}} \tag{4.11}$$

Where r is the Pearson Correlation coefficient, n is the number of students (degrees of freedom) and t is the t-value.

However if you do not wish to assume that your distribution is Gaussian, you can instead use a Monte Carlo sampling method proposed by Craig Savage and John Aslanides [56]. This method can also be used to determine the significance of the absence of expected correlations.

4.3 Monte Carlo Simulations

In John Aslanide's honours thesis (Relativity Concept Inventory) [71] and in Alsanides' and Savage's paper [56] describing the Relativity Concept Inventory, they describe the use of random sampling to obtain the likelihood of seeing certain correlations values between two given questions by pure chance alone. They used these simulations as grounds to reject the null hypothesis that the responses to two questions on the Relativity Concept Inventory were uncorrelated.

Performing these simulations for every pair of questions which is to be investigated is much more difficult than simply using Equation 4.11 above. The advantage is that the assumptions are not based on Gaussian statistics, but instead are tailored to the observed data. My question would be, are these simulations actually beneficial and do they provide different results to the simple Pearson test?

In Aslanides' and Savage's paper [56], they report the statistically significant correlations and the significance values calculated using their Monte Carlo Simulations. Surprisingly they did not compare the values they obtained from their Monte Carlo Method to any other more standard procedures for calculating statistical significance. As a first evaluation of the Monte Carlo Method, I have calculated the p-values using Equation 4.11 for correlations of this strength given their data set and compared them in the Table 4.1 below.

Table 4.1: Signficance of Correlations from the Relativity Concept Inventory [56]. This table shows that Gaussian statistics seems to underestimate the frequency of rare extreme events, producing a lower p-value, effectively over-estimating the significance of correlations.

Question Pair	Pearson's r	p-value from Gaussian Statistics	p-value from Monte-Carlo Simulations [56]
1, 2	0.56	4.6×10^{-7}	5×10^{-5}
5, 6	0.56	4.6×10^{-7}	5×10^{-5}
11, 12	0.44	1.3×10^{-4}	4×10^{-4}
3, 9	0.43	2.0×10^{-4}	3×10^{-4}
15, 22	0.44	1.3×10^{-4}	5×10^{-4}
2, 7	0.39	8.5×10^{-4}	7×10^{-4}
9, 22	0.38	1.1×10^{-3}	9×10^{-4}

The only strong assumption of the Monte Carlo method is the fixing of the question difficulties. In using this method we assume that our students are representative of the overall student population and that the difficulty index for each question will be fixed (the portion of the total population who would answer each question correctly is the same as our sample). Once this assumption is made the Monte Carlo simulation provides an effective method for estimating p-values.

I reproduced Aslanides and Savage's Monte Carlo Method and applied it to check the significance of the correlations found in the Action Concept Inventory. Aslanides used Mathematica to code his simulations, whereas I used R a freeware statistics program. Now that the code is in a freeware language I am able to publish it online, making the technique more accessible to other physics education researchers who wish to use it. The next section contains an outline of the procedure I have used.

Monte Carlo Simulation Method

In order to perform these Monte Carlo Simulations I first needed to create a distribution of the student population to draw from. Since this Monte Carlo method only considers two questions at a time, (let's call them Question x (Q_x) and Question y (Q_y)), a sample of N students, has students of four different types, those who answered both questions correctly (N_{11}), those who answered only Question x correctly (N_{10}), those who answered only Question y correctly (N_{01}) and those who answered both questions incorrectly (N_{00}). Samples of N students were created and a ϕ value (equivalent to Pearson's r for dichotomous data) was calculated using Equation 4.12 [27, 78].

$$\phi = \frac{N_{11}N_{00} - N_{01}N_{10}}{\sqrt{(N_{11}+N_{10})(N_{00}+N_{10})(N_{11}+N_{01})(N_{00}+N_{01})}} \tag{4.12}$$

Following Aslanides, I assumed that the sample of students who took that ACI was a subset of a larger population that I wished to understand (for example, all students who study introductory Action physics) and that the inventory scores resemble that of the larger population [56]. I followed Aslanides' method [71] drawing N students from a multinomial probability distribution function given by [56]:

$$Pr(N_{11}, N_{10}, N_{01}, N_{00}) = \frac{N!}{N_{11}!N_{10}!N_{01}!N_{00}!}(p_{11})^{N_{11}}(p_{10})^{N_{10}}(p_{01})^{N_{01}}(p_{00})^{N_{00}} \tag{4.13}$$

Where p_{11} is the probability that both questions are answered correctly, p_{00} is the probability that both questions are answered incorrectly, p_{10} is the probability that only Q_x is answered correctly and p_{01} is the probability that only Q_x is answered correctly. In order to construct this distribution function I determined the values of p_{11}, p_{01}, p_{10} and p_{00}. To do this I needed four constraints to apply to my population.

The first constraint was a normalisation constraint (Equation 4.14). Aslanides further imposed two constraints that the portion of students who answered each of the two questions in the population is the same as the portion who answered each question correctly in the dataset (in our case the ANU first year students who took the Action Concept Inventory this semester).

$$p_{11} + p_{10} + p_{01} + p_{00} = 1 \tag{4.14}$$

$$P_x = p_{11} + p_{10} \tag{4.15}$$

$$P_y = p_{11} + p_{01} \tag{4.16}$$

Where P_x is the proportion of students who answered Q_x correctly in our sample, P_y is the proportion of students who answered Q_y correctly in our sample. In addition we also set an expected Pearson's r correlation coefficient for the two questions, r_{xy} (if we want to test the null hypothesis we set this to 0) see Equation 4.17.

$$r_{xy} = \frac{p_{11}p_{00} - p_{01}p_{10}}{\sqrt{(p_{11}+p_{10})(p_{00}+p_{10})(p_{11}+p_{01})(p_{00}+p_{01})}} \tag{4.17}$$

Where N_xy is the number of XY outcomes from a sample of N answers. 100,000 data sets of 78 students were then randomly sampled from the distribution. The p-value was taken to be the fraction which showed correlations as strong as the observed correlation.

One Subtlety - No Variance

When sampling from the multinomial distribution it is in fact possible to obtain a data set where not a single student answers one of the questions correctly. When this occurs the correlation is undefined as both N_{11} and N_{10} are 0 (see Equation 4.12). Even Pearson's R, the more generalised formula for correlation, is undefined in this case because if every student answered a question correctly then the variance is 0.

Aslanides and Savage did not publish in their paper how they dealt with this [71, 56]. I can see two logical resolutions to the problem, either discard these occurrences and sample again or treat the correlation as 0 because both the variance of one of the variables is 0. I chose to use the second resolution and in all cases where there was no variance in one of the questions the correlation was therefore 0.

4.3.1 Improving the Monte Carlo Simulation Method

The specific way that Aslanides and Savage applied their Monte Carlo method could have been improved (see [56] for their method). They calculated correlation values for all 276 possible question pairs and then use the Monte Carlo method described in Section 4.3 to estimate their significance. However there were two mistakes that they made when doing this:

1 Aslanides and Savage only looked at the highest correlations (r-) values and checked to see whether these values were significant using the Monte Carlo Method. However this automatically biases them to find correlations between questions that are of similar difficulty. See Section next section *'bias towards question of similar difficulty'*.

2 The level of signficance that Aslanides and Savage used was simply to search for p-values of less than 1/276 [56]. This means that when they calculated correlations for 276 items pairs they would expect to see one correlation by pure chance [56] and if student intelligence played a factor in overall score then you could expect to see even more. They were effectively finding correlations with a p-value of 1, rather than the more usually accepted values of 0.05 or 0.01. As a result of this choice they found many correlations which they later determined to be meaningless (after correcting with the Rasch Model of intelligence, see Section 4.4). This could have been avoided if they had been more strict in their criteria to determine statistical significance. This could also have been avoided if they only checked for correlations where they would be expected.

Bias towards questions of Similar Difficulty

When trying to correlate two questions that have very different difficulties, the strength of the possible correlation is limited. To confirm this, look at the functional form of Equation 4.12 (given below as Equation 4.18 for convenience).

$$\phi = \frac{N_{11}N_{00} - N_{01}N_{10}}{\sqrt{(N_{11} + N_{10})(N_{00} + N_{10})(N_{11} + N_{01})(N_{00} + N_{01})}} \tag{4.18}$$

In the case where the first question is much more difficult than the second question, we have the condition that N_{10} must be large and that N_{11}, N_{00} and N_{01} must be small (relative to a question where the difficulty of each question is equal). As the numerator or the correlation is given by $N_{11}N_{00} - N_{01}N_{10}$, if N_{11} and N_{00} are limited, so is the correlation between the two questions.

Performing Monte Carlo Simulations with questions of differing difficulties confirms this see Figure 4.3 below:

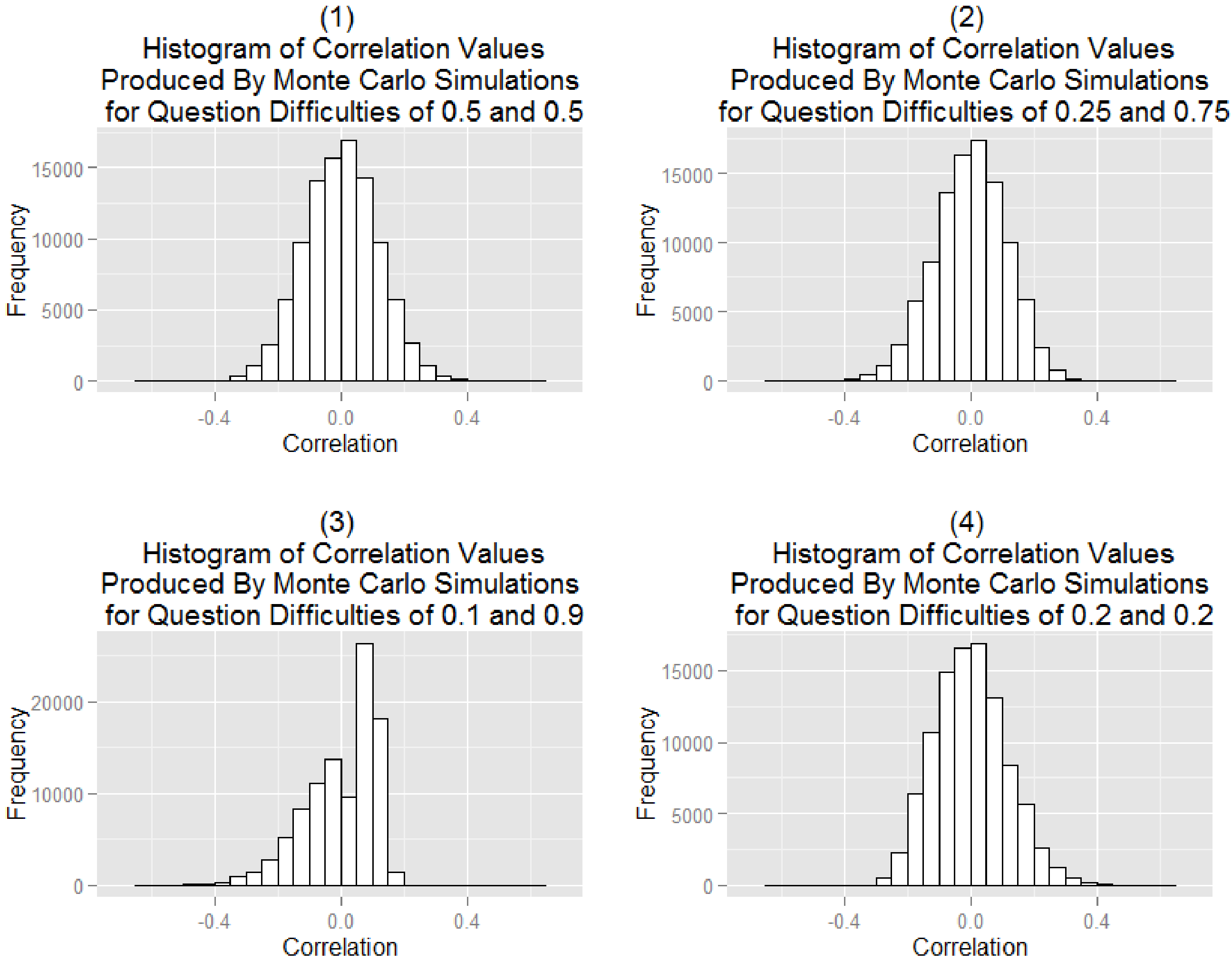


Figure 4.3: Monte Carlo Simulations are shown for 4 pairs of item difficulties. In each case 100,000 trail data sets were created each with of 78 students were created. Comparison of the four figures show that when there is a large difference between the two question difficulties, as shown in (3) the distribution is skewed, making it less likely to see correlations. (4) shows that increasing or decreasing the difficulty of both questions has no significant effect on the distribution so long as the difficulties are similar.

Therefore rather than simply assuming the largest correlation values will be the most significant, the significance of every question pair should be investigated. Alternatively only meaningful correlations could be investigated as described in the next section.

Significance Levels and Checking Specific Correlations

When making multiple comparisons it is generally accepted that you need to decrease the p-value that is being used to calculate significance. For example let's suppose you were checking a 276 matrix of values and wanted to test the null hypothesis that some of the values were significantly different from random data. If a p-value of 0.05 was used then you would expect one in twenty of the values (about 13 of them) to have this level of significance just by chance.

In this case, it is usual to follow a procedure similar to that followed by Aslanides and Savage [56], called the Bonferroni correction of family-wise error rate [89]. In this procedure you simply divide your desired significance p-value by the number of values you are testing. In the example mentioned above, if we wanted to test to a level of 0.05, we would need to look for a significance of $0.05/276 \approx 0.00018$.

The probability of finding correlations caused by random chance can be decreased by investigating less values. It is a well established procedure in psychology literature to only check for correlations that we have reason to expect to see [89]. When using the Bonferroni correction we only need to divide by the number of correlations you actually check. Following the example above, if we only checked 18 values where we expected to see correlations, then the 0.05 significance level would become $0.05/18 \approx 0.0028$.

Even in this case we, may still find that the p-value overestimates the significance of correlations; because some correlation between questions should be observed simply on the basis that some students are more intelligent than others and hence answer more items correctly. The Monte Carlo simulation doesn't take into account that strong students are more likely to get more questions correct and this would lead to a higher correlation values than the random data sets predicted by the Monte-Carlo simulation with Pearson's r set to 0.

It is actually possible to use the Monte-Carlo simulation to test for significance of hypotheses other than $r = 0$ (no correlations). If there is a correlation that we wish to prove false we can specify a different r and test its significance. John Aslanides did this in his thesis to rule out the possibility of a strong correlation between Questions 9 and 10 of the Relativity Concept Inventory [71].

Alternatively the Monte-Carlo method could be used to compensate for student intelligence by trying a different r value. To avoid choosing an arbitrary value the correlation value could be chosen (in the case of the Action Concept inventory post-data $r = 0.042$). This would be a 'one size fits all' approach to the intelligence of the class as a whole and fails to compensate for the intelligence of individual students.

A much more precise way to compensate for student intelligence used by Aslanides and Savage [71] [56] is to fit the data with an intelligence model. A process for doing this with the one parameter Rasch model is described in Section 4.4 below.

4.4 Compensating for Student Intelligence using the Rasch Model

The Rasch model is a well established model that is used to fit test data; creating parameters that describe the intelligence of students and the difficulty of test items [90, 91]. The procedure of fitting a Rasch model simply involves following a logistic regression algorithm to obtain the best possible fit to the data.

In Aslanides' and Savage's paper on the relativity concept inventory [56], they used the Rasch intelligence model to try and uncover more correlations (they attribute the idea to Paul Francis) and confirm the significance of the correlations they observed in their data.

This counters the argument that correlations occurred in their data simply because there are students of different academic ability and questions of different difficulty.

Explaining this argument further: only very intelligent students will answer very difficult questions correctly, therefore you are likely to see correlations between very difficult questions. As the strongest correlations were seen between the most difficult questions for the ACI, this argument could be used to undermine the significance of these correlations. Therefore to ensure these correlations are significant, it would be appropriate to take into account student intelligence and question difficulty.

The Rasch Model essentially creates a new set of data based on parameters student intelligence and difficulty index. My original data set is a matrix of ones and zeros where each row corresponds to a student and each column corresponds to a question. A one in position ij indicates that the ith student answered the jth question correctly, while a zero in this position indicates they answered the question incorrectly.

The Rasch model creates a matrix of the same size, but rather than being filled with ones and zeros it is instead filled numbers between 0 and 1 which indicate the probability that student i would answer question j correctly based on student ability and question difficulty [92].

How does the Rasch Model Work?

The Rasch Model uses the function shown in Equation 4.19 to model the probability of the ith student answering the jth question correctly.

$$M_{ij} = Pr_{ij} = \frac{e^{\theta_i - \beta_j}}{1 + e^{\theta_i - \beta_j}} \tag{4.19}$$

Where M_{ij} is the model probability that student i answers question j correctly, θ_i is a parameter indicating the ability of student i and β_j is a parameter indicating the difficulty of the question. Notice that if student ability is very high (or difficulty parameter is high indicating an easy question) then this model predicts that the probability of the student answering correctly approaches 1. Alternatively if student ability is very low (or a question has a low difficulty index, indicating a difficult question) then the predicted probability approaches 0.

We can use the ith student's average score to create a first guess of student ability (θ_i) and the difficulty index of each question as a first guess of question difficulty (β_j). To do this we use the Equations 4.20 and 4.21, obtained by solving Equation 4.19 for student ability and difficulty index and averaging across all questions (for student ability) and students (for question difficulty).

$$\theta_i = \log(\frac{1 - p_i}{p_i}) \tag{4.20}$$

$$\beta_j = \log(\frac{1 - d_j}{d_j}) \tag{4.21}$$

Where p_i is the fraction of answers student i answered correctly and d_j is the difficulty index of item j.

After we have this first guess we use a logistic regression to continually improve the model [93]. To do this a cost function ($J(M_{ij}, Data_{ij}; \theta_i, \beta_j)$) is required to evaluate how well it is predicting the data. Iterations are repeated until the cost function, $J(M_{ij}, Data_{ij}; \theta_i, \beta_j)$, has been minimised within a given tolerance [93]. The cost function must have certain properties to ensure that our iteration algorithm will work effectively.

Our cost function is a function that takes in both the values predicted by the Rasch Model M_{ij} (which implicitly depends on θ_i and β_j) and the Action Concept Inventory data $Data_{ij}$, and outputs a number which indicates how well the model fits the data. We could imagine that out cost function could look like either of the two functions shown in Figure 4.4.

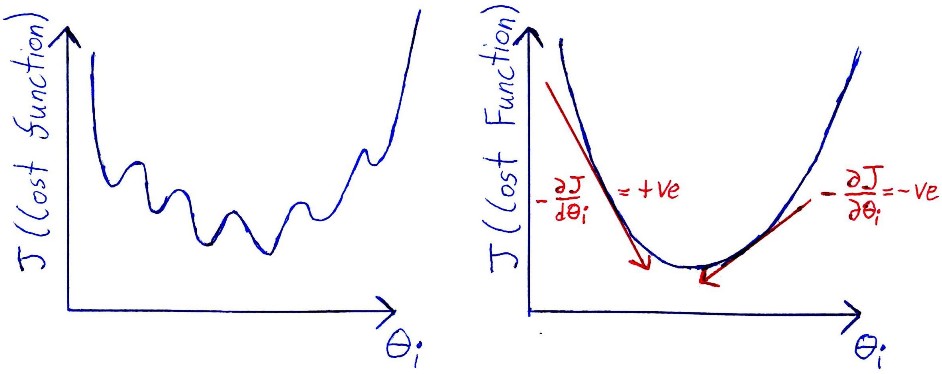


Figure 4.4: Two cost functions are shown only as a function of θ_i for simplicity. The one on the right is convex, the one of the left is not. Following Newton's method, adjusting the parameter θ_i in the direction of minus the derivative of the cost function $(-\frac{\partial J}{\partial \theta_i})$ with each iteration will lead to J moving towards the local minimum. To reach the global minimum with this algorithm it is essential that a convex cost function is chosen, otherwise the starting conditions may lead to convergence on a local minimum rather than the global minimum.

With each iteration we adjust the parameters θ_i and β_j according to Equations 4.22 and 4.23 to ensure that it moves toward the local minimum of the cost function (see Figure 4.4).

$$\theta_i \leftarrow \theta_i - \alpha_1 \frac{\partial J}{\partial \theta_i} \tag{4.22}$$

$$\beta_j \leftarrow \beta_j - \alpha_2 \frac{\partial J}{\partial \beta_j} \tag{4.23}$$

Where α_1 and α_2 are positive constants which vary on each iteration, to allow convergence to be reached more quickly. J is the cost function, θ_i is the student ability parameter and β_j is the question difficulty parameter. J and α are found using a statistical principle called the 'principle of maximum likelihood estimation' [93].

As the Rasch Model is fairly standard in item response theory there are several R packages fit the Rasch Model to data [91]. I used a package called 'eRM' to fit the Rasch Model to my data, it is available at `http://cran.r-project.org/web/packages/eRm/index.html` . A very brief tutorial on how to use this package was written by Reinhold Hatzinger and can be accessed online [90] at `http://www.stat.uni-muenchen.de/~carolin/material_psychoco/folien_fuer_homepage/reinhold.pdf` .

Once the Rasch Model (M_{ij}) was fitted to the data, the residuals (R_{ij}) were calculated by simply subtracting the Rasch Model off the original data. See Equation 4.24:

$$R_ij = Data_{ij} - M_{ij} \tag{4.24}$$

Correlation values were then calculated from the residual matrix using Pearson's R.

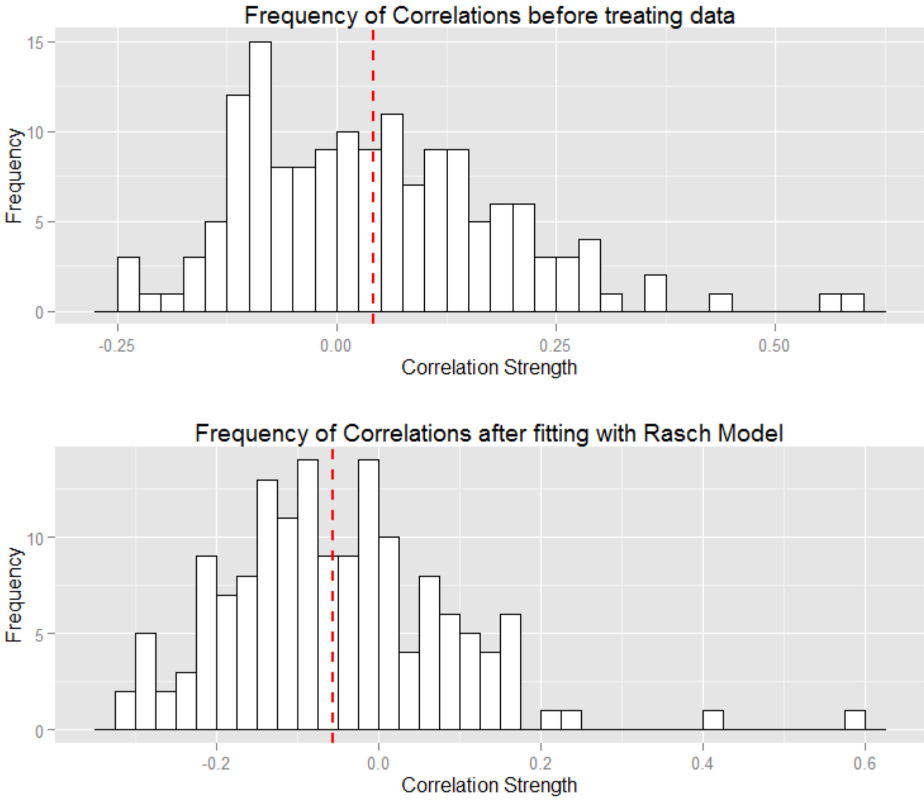


Figure 4.5: Histogram of Correlations before and after intelligence is compensated for using the Rasch Model. Mean correlation is shown by red dashed line.

4.5 Factor Analysis and Principle Component Analysis

Another method of finding correlations of students responses to different questions is to use factor analysis, a specific form of Principle Component Analysis. Factor analysis is capable of finding relationships between groups of questions and thus an advantage over the pairwise comparisons that we have discussed so far.

Factor analysis is frequently used in physics education studies [94, 63] for two primary purposes. First to reduce high dimensional data into a smaller number of dimensions to allow researches to visualise it. Secondly factor analysis is used to establish groups of correlated items, called factors, which may be used to help establish the validity of the test.

Principle component analysis is a variable reduction procedure [94]. In the Action Concept Inventory, several items are designed to test the same topic. If these items are valid they should be high correlated. Principle component analysis groups highly correlated items together into factors. In this way factor analysis provides a measure for how well different test items may be related in terms of consistencies among student responses [95].

Factor analysis is a mathematical tool and does not provide the reasoning for such relations between components that make up a factor, this must be done by the researcher.

4.5.1 Number of Factors

There are a number of measures that you can use to determine how many factors are 'significant' and actually correspond to a significant factor in he data, rather than simply describing noise. These methods include the Kaiser-Guttman rule, Cattell's Scree test, parallel analysis, optimal co-ordinates and acceleration factor [96]. For more information on these methods see [96] and references therein.

4.6 Model Analysis and Student Misconceptions

Model analysis is a good method for trying to capture a better understanding of the wider dynamics of student learning developed by Lei Bao and Edward Redish in 2006 [95]. Rather than treating each student response as simply correct or incorrect, incorrect responses are examined more closely to determine what misconceptions exist in the student population.

In model analysis we investigate all the different ways students think about a particular topic. We call each broad category of idea a model. Model analysis is primarily based on two critical facts [95]:

I Cognitive literature has shown it is possible to have multiple contradictory ways of thinking (models) at a given time. Different contexts will activate different and contradictory bits of knowledge.

II On any particular topic, the range of alternate misconceptions in a population tends to be fairly limited. Usually 2 or 3 specific ideas account for most of student thinking.

The easiest way for me to explain Model analysis is to draw a close analogy with quantum mechanics (which shares the same mathematical techniques). Model analysis treats students brains as being in a superposition state of many different models they could use to answer a physics question. When students interact with an inventory item, their brain (wavefunction) collapses to a single model which they use to answer the question.

Different questions will contain different contexts which trigger different models in each student. In this way each question on the inventory can be seen as a measurement of the state of a student's brain. By asking a student many questions we are able to obtain information about the likelihood of a student using a particular model to answer a given question.

Bao and Redish outline a particular procedure for implementing model analysis [95]. I have summarised their procedure in the steps below [95]:

i Through systematic research and interviews with students, common student models are identified.

ii This knowledge is used to design a multiple choice instrument where each of the distractors corresponds to a common student model.

iii A class of students take the test and student responses are characterised in a linear model space representing the (square roots of the) probabilities that a student will engage different common models.

iv The individual student model states are used to create a density matrix which is then averaged over the whole class. The off diagonal elements of the matrix retain information about the inconsistency (confusion) of individuals students.

v The eigenvalues and eigenvectors of the density matrix give information about the level of confusion in the state of the class's knowledge.

For the Action Concept Inventory groups of 4 or 5 inventory questions which all targeted a particular concept were grouped together to use for model analysis. Each option of each of the question was assigned to one of the student models. Usually model 1 always corresponds to the correct model. In addition the final model is always referred to as a null model. Options which do not fit nicely into any of the other models are allocated to the null model. Table 4.2 summarises the questions which test the 'explore all paths' model and is shown below as an example.

Table 4.2: Associations between physical models and the choices of the five Action Concept inventory questions which test the explore all paths concept.

Questions testing 'Explore all Paths'	Model 1 A particle takes all paths and interferes with itself	Model 2 A particle takes only one path	Model 3 Other
Q8	d	a	b,c
Q14	b, c	a,d	e,f
Q15	d	a,b	c
Q17	a,c,d	b	e
Q18	c	a,b,d	e

A vector Q_i is constructed for each student, where each entry of Q_i represents the probability that the student will use a particular model [94]. In Equation 4.25 a vector is shown for a student who answered according to model 1 for three out of five questions on a particular concept. The student used model 2 and model 3 once each to answer the two remaining questions.

$$Q_i = \begin{pmatrix} 0.6 \\ 0.2 \\ 0.2 \end{pmatrix} \tag{4.25}$$

We then take the square root of each element in Q_i to create a new vector V_i, see Equation 4.26 [94].

$$V_i = \begin{pmatrix} \sqrt{0.6} \\ \sqrt{0.2} \\ \sqrt{0.2} \end{pmatrix} = \begin{pmatrix} 0.775 \\ 0.447 \\ 0.447 \end{pmatrix} \tag{4.26}$$

We then take the outer product of V with its transpose ($V \otimes V^T$) to create the "density matrix" for each individual student, D_i [94]. Next we take the average over all the students to obtain a class density matrix D. Depending on how students use the models, the class density matrix will show different patterns. Figure 4.6 is taken from Ding and Beichner's paper 'Approaches to data analysis of multiple choice questions' and shows three examples discussed by Bao, Redich, Ding and Beichner [95] [94].

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \qquad \begin{pmatrix} 0.5 & 0 & 0 \\ 0 & 0.3 & 0 \\ 0 & 0 & 0.2 \end{pmatrix} \qquad \begin{pmatrix} 0.5 & 0.2 & 0.1 \\ 0.2 & 0.3 & 0.1 \\ 0.1 & 0.1 & 0.2 \end{pmatrix}$$

(a) **(b)** **(c)**

Figure 4.6: Examples of class density matricies as appeared in [95] and [94]. (a) corresponds to the entire class using model 1 consistently, (b) the class consists of three groups of students who each use a consistent model and (c) students use multiple models inconsistently.

In Figure 4.6, class density matrix (a) has only one nonzero element along it's diagonal, corresponding to the class using one model consistently [94]. In class density matrix (b) only the diagonal elements are non-zero, meaning that there are three subsets of students in the population each using a different model consistently. Class density matrix (c) is the most realistic and corresponds to different students using a mixture of different models when they answer different questions.

Once the density matrix is found, we solve its eigenequation and find the largest eigenvalue, λ_{max}, and it's corresponding eigenvector $U = (U1, U2, U3)$. According to Bao and Redish [95] the class probability of using each model is given by multiplying λ_{max} by the corresponding U value squared. For example to find the probability that the class uses model 1: $Pr(Model_1) = \lambda_{max} \times U_1^2$. Note that the sum of all these probabilities is frequently less than 1 [94] implying that these probabilites do not form a complete set. This originates from the second and third eigenvectors not being taken into account. Bao and Redish consider these additional eigenvectors “corrections that are not presented by the primary state" [95].

Model analysis was used on the Action Concept Inventory data to gain a greater understanding of the PHYS1201 class, however there are a number of reasons why it was difficult to implement:

- Qualitative studies of student's misconceptions is a necessary precursor to model analysis. As the common misconceptions of students relating to Action physics have not been published it will be very difficult to implement model analysis effectively for the Action Concept Inventory.
- Requires clusters of items that all test the same concept. If you wish to test 9 concepts (as in the case of the Action Concept Inventory), then to have 4 or 5 items on each concept would blow the test out to 45 questions, which could take more than an hour for students to complete. This would make administering the inventory in a single class impractical.
- Model analysis is most valuable in cases where research has documented that students enter a class with a small number of strong misconceptions [95]. Most students have never experienced quantum mechanics in everyday life and will not be likely to have strong misconceptions. This makes it less likely that students will have strong misconceptions relating to quantum mechanical topics or the many paths formulation. Students will most likely enter the class without any strong misconceptions about Action physics.

4.7 Confidence

Students completing the Action Concept Inventory were asked to rate their confidence of their answer in each question. This was first used by John Aslanides and Craig Savage in developing the Relativity Concept Inventory [71] [56]. The primary purpose of confidence data was that it allowed Aslanides and Savage to establish the existence of true misconceptions as opposed to students randomly guessing. For a distractor to correspond to a strong misconception students should have a high confidence rating.

4.7.1 Review of Psychological Literature: Use of a 7 Point Scale

Aslanides and Savage originally used a 5 point scale to measure confidence. Having an odd number of options is well accepted as standard practice in psychology literature as it allows subjects to give a neutral response. Psychology literature would also suggest that it would be much better to increase the number of points on the scale to at least 7 [74, 78]. It is generally accepted by psychologists that if a scale has only 5 options, it should not be treated as a continuous data but instead categorical data [74, 78, 97, 98]. There is debate in psychology literature about whether Likert scale type data can ever be treated as continuous (for anyone interested in learning about this debate this webpage: `http://www.theanalysisfactor.com/can-likert-scale-data-ever-be-continuous/` by Karen Grace-Martin [99] is a good starting point. Although it is not peer-reviewed, it is a good summary of the debate and references papers on both sides of the debate).

Data type is incredibly important as this greatly restricts the type of statistics you can use on the data. Mean, standard deviation and correlations such as Pearson's product moment cannot be used. Kline critiqued the use of these procedures on 5 point scales:

> *"Pearson product-moment correlations with five-point scales or below the product-moment correlation is dubious."* [74]

Savage and Aslanides were probably unaware of this psychology literature and arbitrarily chose to use a 5 point scale and then proceeded to calculate mean confidences and correlation of confidence with other variables. To make these calculations more valid, I increased the number of options on the confidence scale to 7 for the Action Concept Inventory. Although it is still controversial to treat 7 point scale data as continuous [100, 97], the integrity of this treatment is much more widely accepted [74, 78]. If confidence data is used to establish the existence of misconceptions in future, researchers should make it common practice to use 7 or more points on their confidence scales.

4.7.2 Normalising Confidence Data

In interviews, I noticed that particular students would verbally report that they did not understand what they were doing, that they were almost guessing, but would then proceed put down confidence ratings as high as 6 (or in some cases 7) out of 7. Therefore confidence data was only be treated as relative confidence for an individual student and each confidence score was normalised.

I normalised student confidence by subtracting each student's mean confidence to all questions from each of their confidence scores and then dividing by the standard deviation. This is a standard procedure producing a z-score for each confidence value. This normalised confidence is only useful for comparing different questions, not for comparing different students. I only use the confidence data to compare questions so this technique is appropriate.

4.7.3 How Will Confidence Data be Used?

The primary purpose of the confidence data is to establish misconceptions. The mean normalised confidence was calculated for each of the multiple choice options. Incorrect answers with a high confidence indicate that a student has a misconception associated with than answer.

Although not a well established technique, confidence also has the capacity to be useful in establishing the validity and reliability. For example, questions on the inventory which have low normalised confidence give an instructor good reason to assume that students are randomly guessing answers. If negative gains are observed, low confidence provides a possible explanation. In this way, confidence can be used to make conclusions about the values reported by the statistical techniques mentioned earlier on in this chapter.

In this Chapter and in Chapter 3 *Developing The Action Concept Inventory*, all the analysis techniques, both qualitative and quantitative have been outlined. In the next Chapter *Results*, each of these techniques are used to evaluate the inventory.

Chapter 5

Results and Discussion

As described in Chapter 3, there were many sources of data collected from which to draw conclusions about the Action Concept Inventory. These sources included a pre-test and post-test of the Action Concept Inventory, validation interviews, expert review, observations of students in tutorials, student homework and an attitude survey. Of these the most important is the Action Concept Inventory pre-test and post-test data.

Craig Savage, the course convener, specifically allocated half an hour of lecture time and half an hour of laboratory time for students to complete the Action Concept Inventory pre-test and post test. It was made clear to the students that participation was completely voluntary and that they were not obliged to participate in the research.

5.1 Summary statistics

Of the 108 students in the course 106 chose to participate in either the pre-test or post-test. 63% of the students who took both the pre-test and post-test where male, which roughly corresponds to the gender ratio of the class as a whole (64% male). The Venn diagram below (Figure 5.1) summarises student participation.

The fraction of students who participated in the Action Concept inventory is much higher than two years ago when similar studies were used to validate the Relativity Concept Inventory here at the ANU [71]. I believe the two main reasons for this high participation rate are:

- The convenience of the Action Concept Inventory, it was delivered online and took students approximately 20-30 minutes to complete. Because it was delivered in the online format students who were unable to complete the inventory in class due to illness or commitments were able to do so at home.
- The Action Concept Inventory post-test was given to students to complete in their lab class. Laboratory classes are compulsory. Although completing the Action Concept Inventory was not compulsory, most students treated the inventory as a learning exercise of the lab.

77 students who took the pre-test and the post test, which is very large compared to previous concept inventory studies at the ANU. This was a sufficient number of students so that I only considered students took both pre-test and post-test to obtain the statistics I needed.

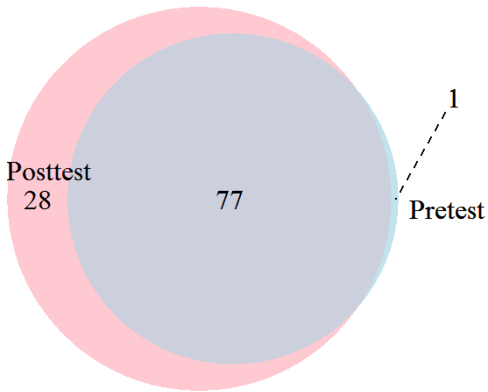


Figure 5.1: Number of Students who completed the pre-test and post-test of the Action Concept Inventory. 78 student took the pre-test and 105 students took the post test. Only the 77 students who took both pre-test and post-test were used in the data analysis.

This allowed me to make some assumptions about my sample straight away. Only students who were motivated enough to attend lectures were included in the sample. It is reasonable to believe that motivated students are more likely to take the test seriously, decreasing the probability that students randomly guessed answers. This was confirmed when the student data compared to data generated from random guessing, see the next Section.

5.1.1 Is the data random?

A group of students randomly guessing answers to a single question can be modelled by a multinomial distribution, where $p = 1/m$ is the probability of a student answering the question correctly. Therefore the mean student score is simply the mean total score of the distribution for a group N students divided by N. Since the mean of the distribution is given by $np = n/m$ the discrimination index of an inventory item if students randomly guessed would be:

$$d_{\text{random}} = \frac{1}{m} \tag{5.1}$$

This is exactly what we would expect for students guessing on a multiple choice test with m options. Determining the standard deviation uses a similar procedure, the variance of a multinomial distribution is given by $np(1-p)$. The standard deviation of difficulty indices found in of a class of n students randomly guessing is:

$$\sigma_{\text{random}} = \frac{1}{n}\sqrt{n\frac{1}{m}\left(\frac{m-1}{m}\right)} = \sqrt{\frac{m-1}{m^2 n}} \tag{5.2}$$

Random guessing has been included on the plots of item difficulties and student gain, see Figures 5.3 5.4 5.13. Student responses for most questions fall out of the range of two standard deviations of the mean that would be obtained from random guessing. This supports the hypothesis that the students took the inventory seriously and attempted to answer as many questions as possible correctly. This is further supported by investigating the frequency with which each individual option was selected, see Figure 5.7.

5.2 Reliability

The next statistics calculated were those relating to the reliability of the Action Concept Inventory. I used the KR_{20} formula which measures the internal consistancy of the test as a whole. I expected this to be low for the Action Concept Inventory because it tests many different concepts which are not necessarily closely related. The KR_{20} internal reliability co-efficient was calculated to be 0.53 for the post test results, which would be considered low as physics education (and other test theory) literature [101] [94] state $KR_{20} > 0.6$ or $KR_{20} > 0.7$s. Given the inventory tests multiple different concepts 0.53 is not alarmingly low and it does not give me reason to believe the inventory is unreliable.

The point biserial correlation coefficient for each question is shown in Figure 5.2. If the Action Concept Inventory was designed to measure and test one specific variable (such as an intelligence test), the values shown in Figure 5.2 would be low enough to indicate that there is something wrong with the test or that some items should be removed. Kline states that all test items should have point biserial correlations of 0.2 [74] or higher while Kaplan suggests that a desirable range for point biserial coefficients is between 0.3 and 0.7 [80] [101]. However the Action Concept Inventory tests multiple concepts and therefore should not have the same level of internal consistancy as a test which measures only one variable. Most of the inventory items have positive biserial coefficients and there are no strong anti-correlations. Therefore the biserial coefficients do not give reason to suggest that any of the questions should be removed.

Test-Retest Reliability

It was appropriate to calculate the test-retest reliability for some questions as there was evidence to suggest that students had not studied or learned about these particular concepts. Using both pairwise correlations and factor analysis it was shown that Questions 1, 2, 3, 5 and 14 possessed test-retest reliability. See Section 5.5.5.

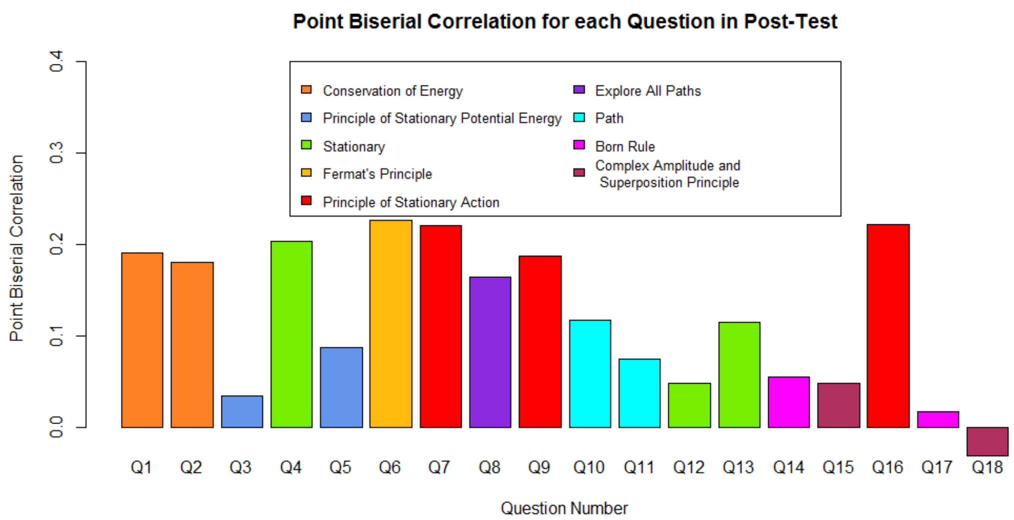


Figure 5.2: Point biserial correlation coefficient for each of the Action Concept Inventory Items in post-test. Point biserial correlation is a measure of how well performance on each item correlates with overall test score. The values for the ACI are low, as would be expected because the ACI tests a broad range of concepts which do not necessarily correlate with one another.

5.3 Discrimination and Difficulty

Next the difficulty and discrimination of the entire test and each individual item were calculated. Ferguson's Delta, a measure of discrimination power of the entire test, was calculated to be 0.93 for the post-test, indicating that there was a reasonably good spread of student scores and that the test was able to discriminate between students.

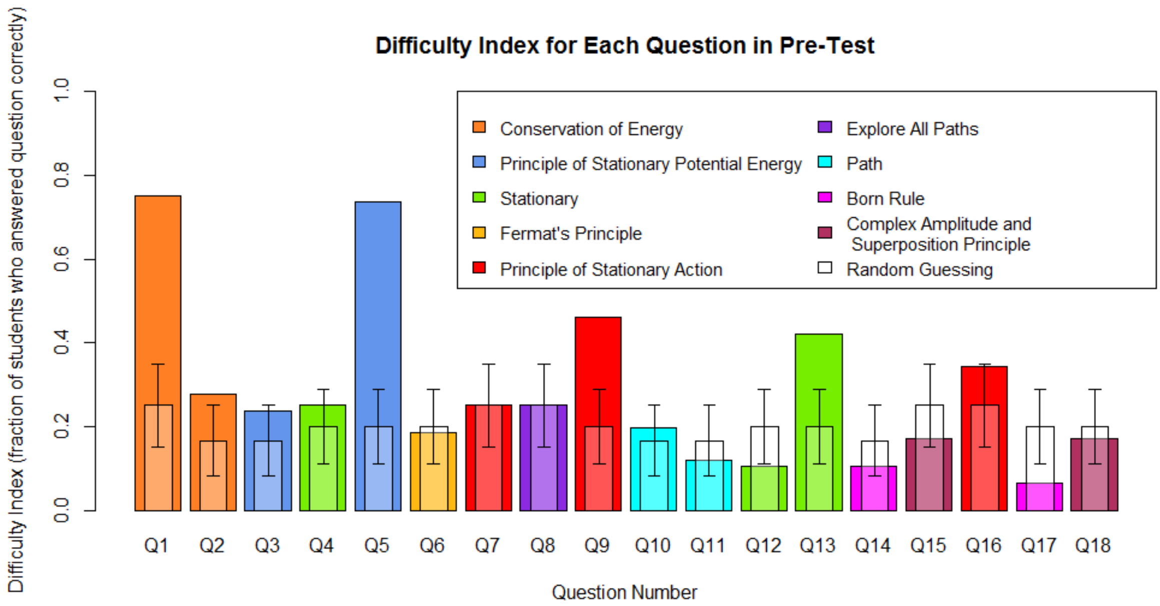


Figure 5.3: Difficulty index of each item of the Action Concept Inventory pre-test. Difficulty index is the proportion of students who answered the question correctly. Colour indicates the concept each questions corresponds to. Random guessing is shown in white with 95% confidence intervals (2 standard deviations). For Q12 and Q17 students obtained results significantly (at the 0.05 level) lower than if they had random guessed the answers, which indicates that distractor options were very effective.

The difficulty index is shown for each question both pre and post test in Figure 5.3 and Figure 5.4. Figure 5.4 and 5.3 show that the student data is different from random guessing and further supports the hypothesis that students took the test seriously and answered to the best of their ability.

The extremely low difficulty index of Q12, Q14, Q15 and Q17 demonstrates the effectiveness of the distractors and indicating that some of these options may correspond to student misconceptions. Question 13 had an unusually high difficulty index on the pre-test. Given that students had not learned any Action at this point it is unusal that so many would answer this question correctly. This is the first piece of evidence that led me to question the validity of question 13.

The discrimination index of each individual item was calculated by considering the top and bottom quartiles of students in the post-test. A histogram of student scores in the post-test highlighting the quartiles is shown in Figure 5.5. Figure 5.5 shows that the students broke nicely into four roughly even quartiles.

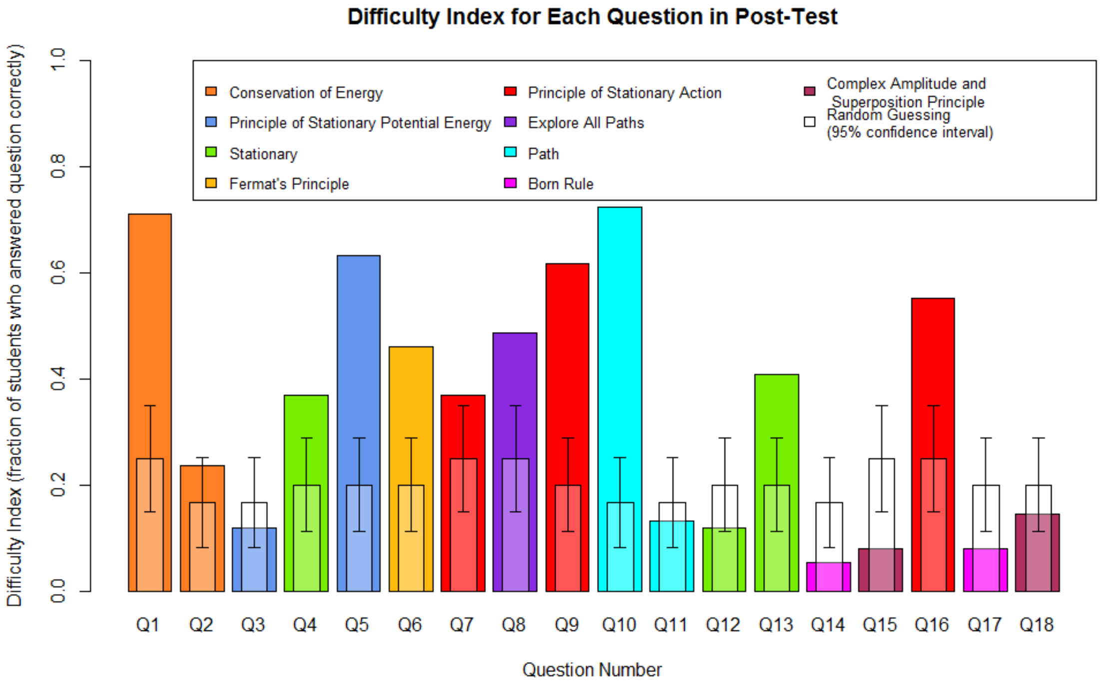


Figure 5.4: Difficulty index of each item of the Action Concept Inventory post-test. Difficulty index is the proportion of students who answered the question correctly. Colour indicates the concept each questions corresponds to. Random guessing is shown in white with 95% confidence intervals (2 standard deviations). For Q14, Q15 and Q17 students obtained results significantly (at the 0.05 level) lower than if they had random guessed the answers, which indicates that distractor options were very effective.

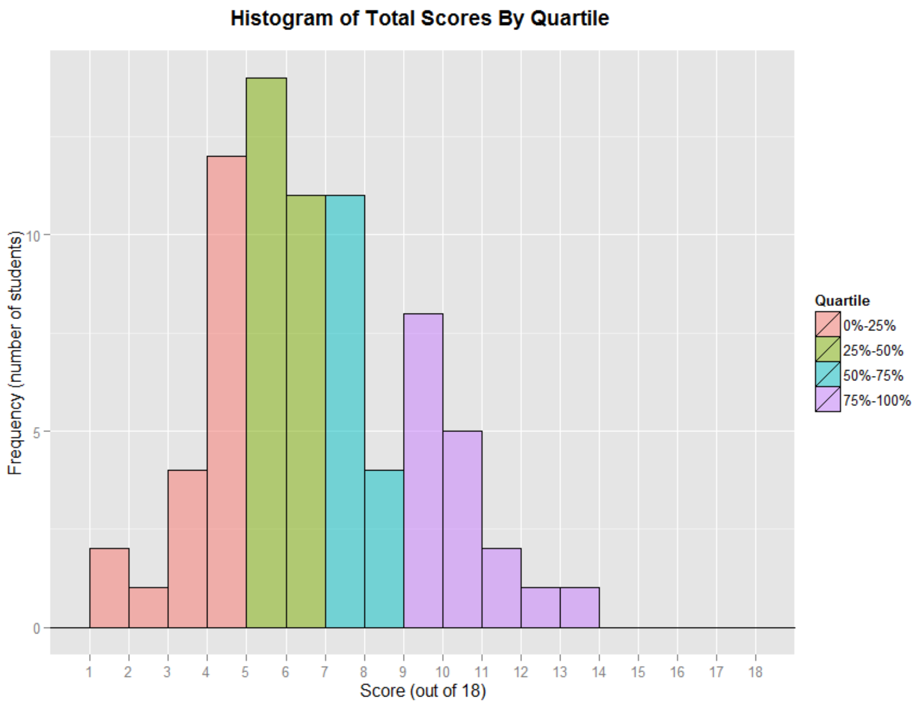


Figure 5.5: Histogram of student scores in the post test data. The four quartiles are shown in different colours. The difference in score between the 0% – 25% quartile and the 75% – 100% quartile gives the discrimination index for each question.

The top and bottom quartiles were used to calculate the discrimination index of each item. The discrimination index for each question is shown in Figure 5.6.

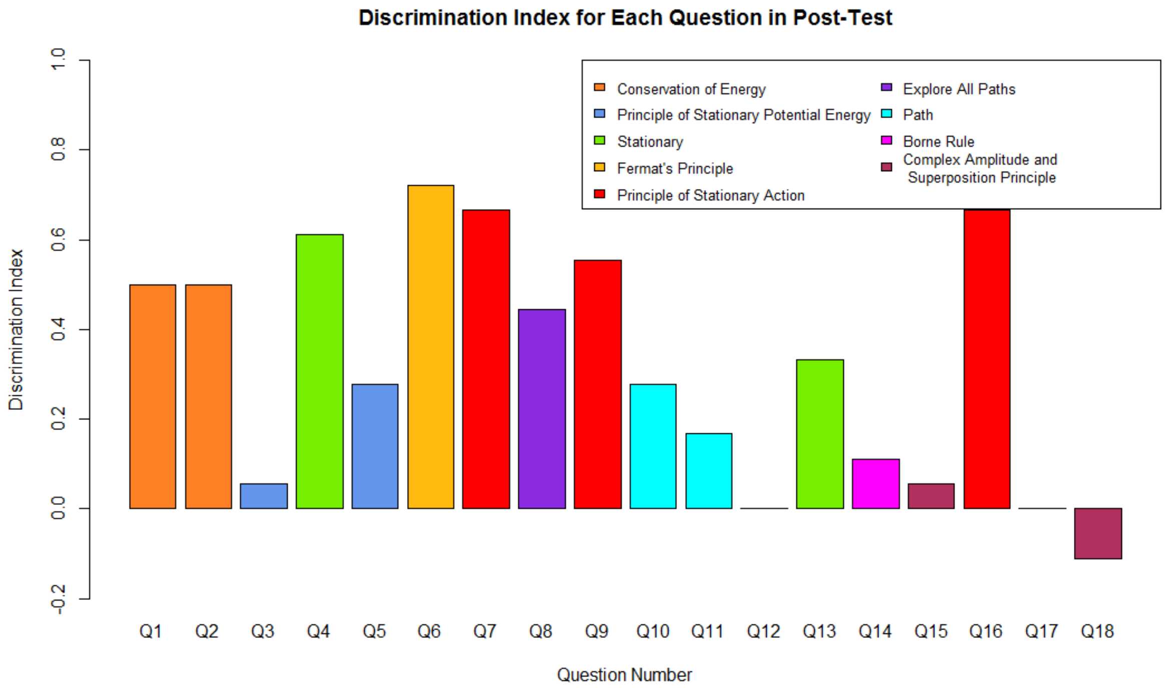


Figure 5.6: Discrimination Index for each of the Action Concept Inventory Items in post test.

The low values of discrimination index for Q3, Q11, Q12, Q14, Q15, Q17 and Q18 indicate strong students did no better on these questions than weaker students. This could indicate that either the questions are invalid or that the questions don't correlate with overall ability (the latter is likely as the ACI tests many concepts which are not closely related). In the case of the quantum mechanics questions, Q14, Q15, Q17, Q18 it was most likely that a few students had previously been taught quantum mechanics, but this did not correlate with overall ability. In validation interviews the only student who correctly answered questions 14, 15, 17 and 18 was able to do so because she had read Feynman's QED.

One question that needs to be explored: "Was the Action Concept Inventory too difficult?" The data shown here will be used to make this conclusion in Chapter 7.

Although scores on the quantum mechanical questions were low, this was to be expected. Less than one lecture was spent teaching quantum mechanical concepts. Interviews with students and their homework responses indicate that they do not understand the quantum mechanical concepts that underpin Action. Therefore the inventory was doing it's job correctly by showing that students do not understand quantum mechanics. Therefore I would conclude not that the inventory needs to be changed, but instead either more time or a different method needs to be used to teach students the quantum mechanical principles.

Question 3 was also very low scoring on the post test as were questions 11 and 12. Although Question 3 may have been valid, it may have gone into unnecessary mathematical depth for an Action Concept Inventory designed for first year students, especially since it is meant to be testing one of the scaffolding (easier) concepts.

Question 11 was completely changed as the result of expert review and validation interviews which showed that students with correct reasoning were choosing the wrong answer. In Question 12, the wording 'arbitrarily', may have caused confusion, interviews revealed that this wording was causing problems for students.

5.4 Misconceptions

To establish misconceptions, the frequency of selection and student confidence was investigated for each option. Figure 5.7 shows how regularly each option was chosen in the pre-test and post-test. To be a good distractor an option must be selected by a large number of students. To be a strong misconception, there is an additional criterion that the student must be confident as well. Figure 5.8 shows the mean confidence for each option selected. This can be used to evaluate whether each option is a good distractor.

Note: 7b, 8a and 16a all have higher confidence than any other option for the same question. This could indicate that they correspond to actual misconceptions rather than just being good distractors. The error bars shown in Figure 5.8 correspond to one standard error in the mean of the confidence. The error bars are large enough that it is not possible to say that the differences in confidence levels between options are statistically significant.

Option 7b and 16a correspond the the 'minimum Action' misconception. Students with this misconception believe that the physical path is the one that has minimum Action not stationary Action. Both of these options had a high confidence and were selected frequently, indicating that 'minimum Action' is a real misconception that is prevalent in the first year class.

Validation interviews and expert review indicated that question 8 was not valid, therefore the high confidence and frequency of selection for this question is less meaningful. It corresponds to the misconception that a quantum mechanical particle only takes the classical paths.

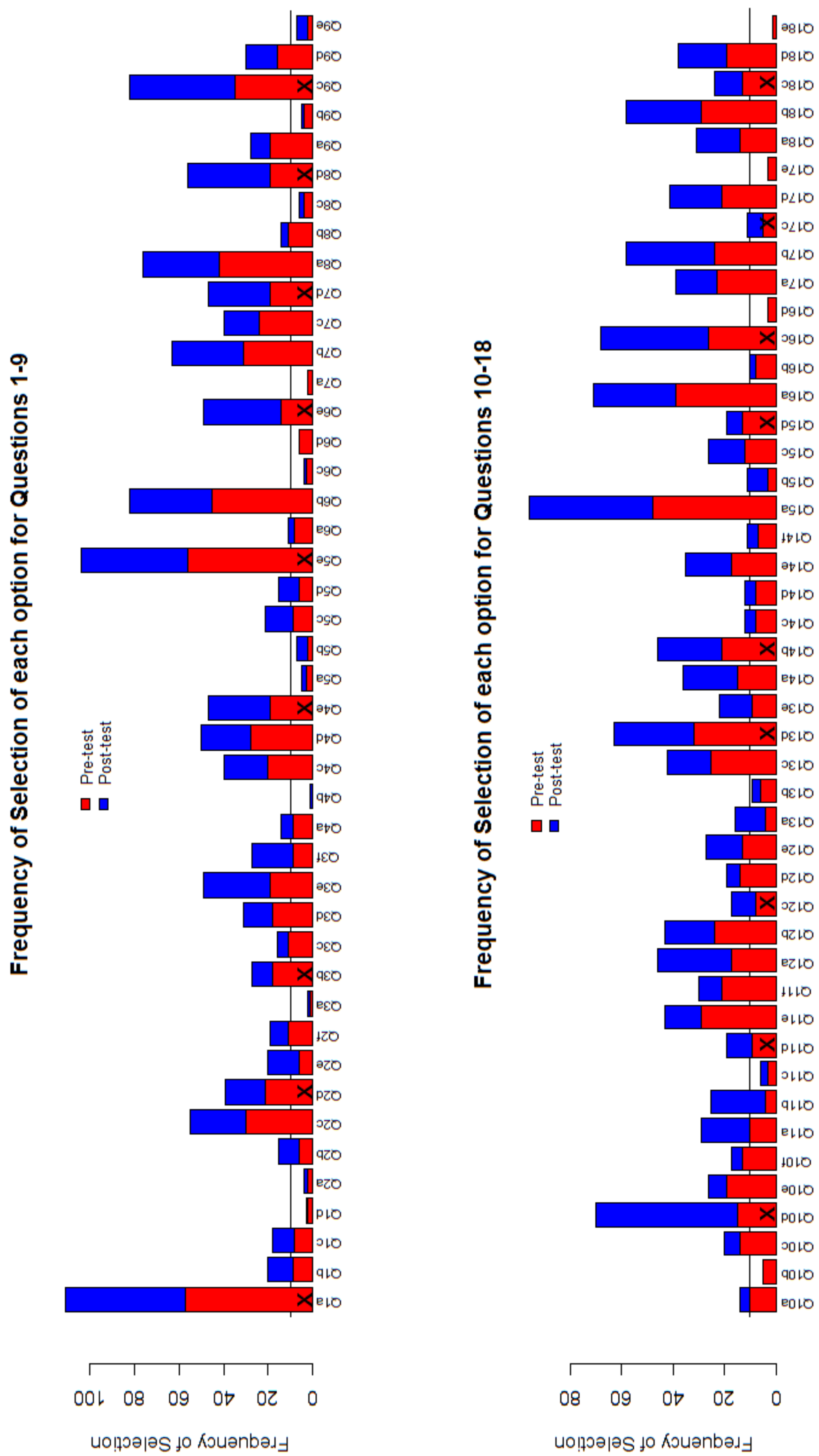


Figure 5.7: Bar plot showing how frequently each option was chosen in the Action Concept Inventory, both pretest and post test. X mark the correct answers. The most frequently chosen incorrect answers were 6b, 7b, 8a, 15a, and 16a. 6b, 7b and 16a correspond to the 'least Action' (as opposed to stationary) misconception. Question 15a corresponds to the misconception that more paths always increase probability of a transition (ignoring the possibility of destructive interference). I suspect that Question 8a was selected frequently simply because 8b and 8c are weak distractors.

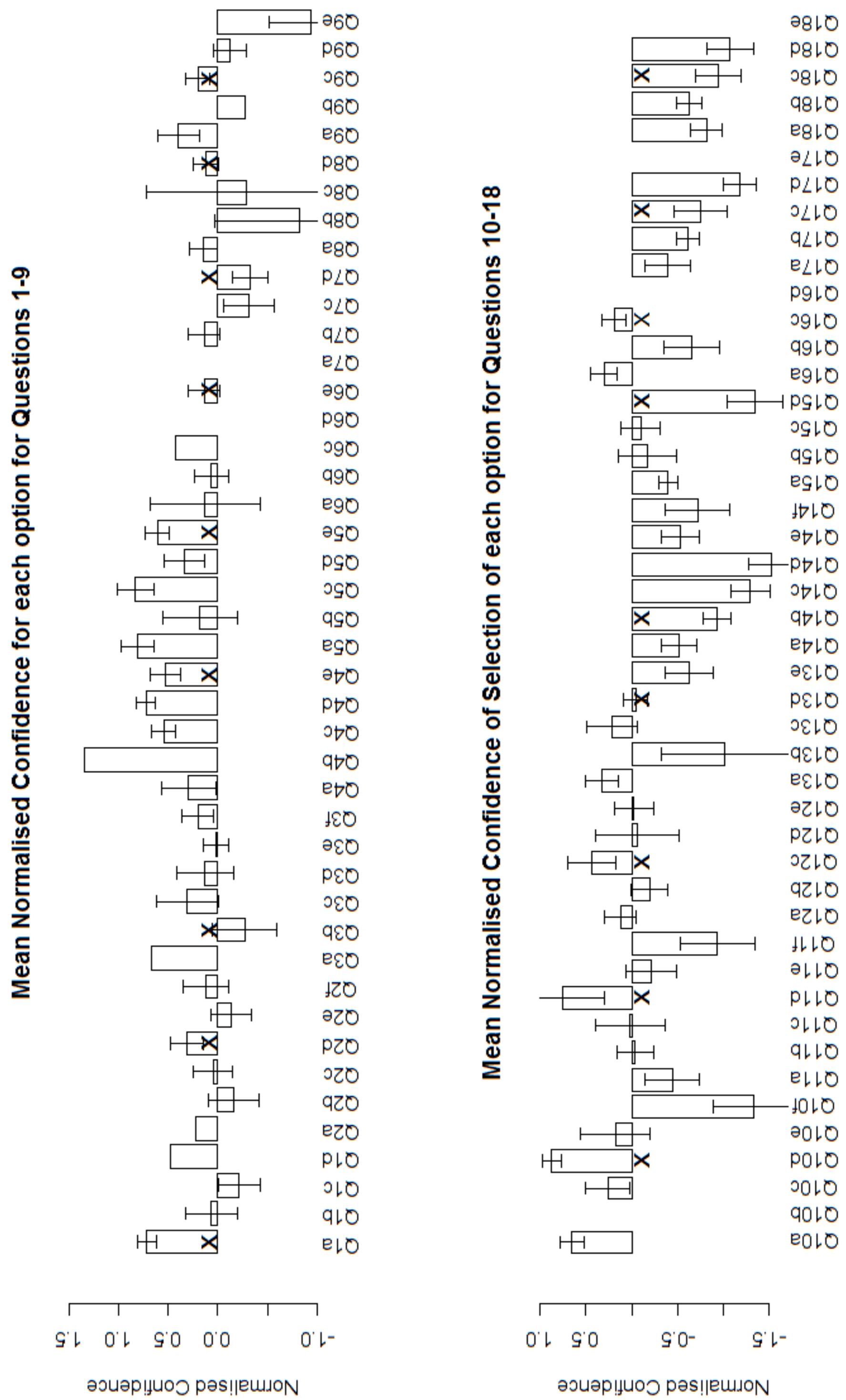


Figure 5.8: Bar plot showing the mean normalised confidence of each option selected for post test of the Action Concept Inventory. Correct answers are shown with an ‘X’. Error bars show one standard error in the confidence. Questions which do not have a column were not selected by any students and columns without error bars were options only selected by a single student. It can be seen that generally confidence is higher for correct answers than incorrect answers. As the student self reported confidence is not a well established technique, this is evidence that it provides meaningful data and is a valid technique to identify student misconceptions.

5.4.1 Model Analysis and Student Misconceptions

The results from the model analysis are shown in 'Model plots' Figures 5.9 and 5.10 below. Model plots are based on model analysis, the idea that students are in a 'superposition state' of all the possible models they could use when answering a question. Bao and Redish proposed model plots as a convenient way of showing the probability of students using two different models [95]. Model plots are a convenient way of comparing different classes or of seeing how a class's understanding changes between pre and post instruction. The closer the points are to the origin the more likely students are to engage in the random/inconsistent model.

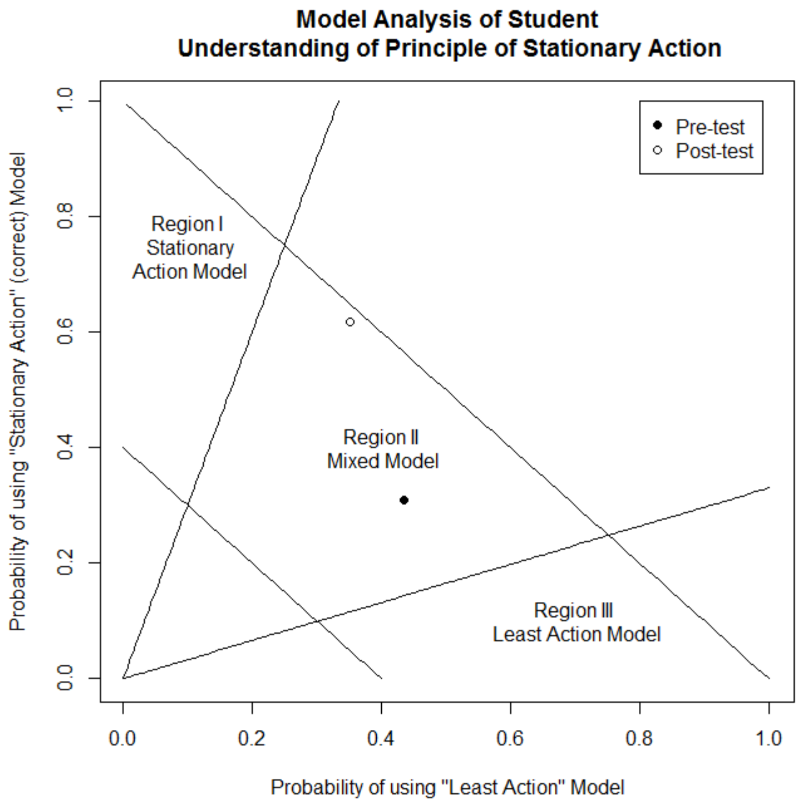


Figure 5.9: Model Analysis of student understanding of the Principle of Least Action. Here we can see that after instruction students were more likely to use the model of Principle of Stationary Action rather than the Principle of Least Action.

The off-diagonal entries of the class model matrix are large in each case (approximately equivalent in magnitude to the corresponding diagonal entries). This indicates that students are inconsistent in the models that they are using. Normally this would indicate that students are at a stage where they are moving between models, they are still learning.

There is another possibility. As the misconceptions surrounding Action physics have not been established, the Action Concept Inventory could not have been designed to investigate well understood student misconceptions. Therefore it is possible that the options do not necessarily correspond to strong student misconceptions causing large off-diagonal elements.

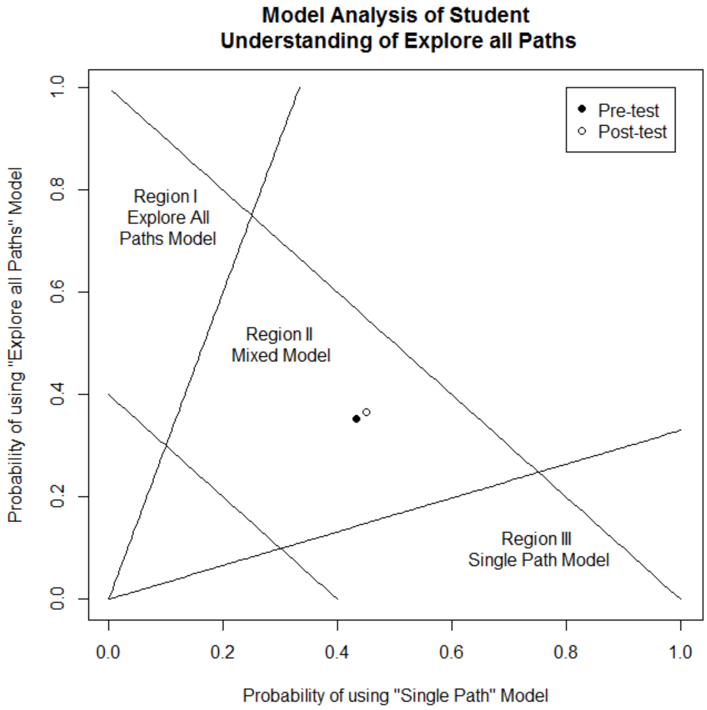


Figure 5.10: Model analysis of student understanding of Explore all paths. The quantum mechanical concepts were not given much emphasis in the course, so it is not surprising that models that students use to think about quantum mechanical concepts were not significantly modified by instruction.

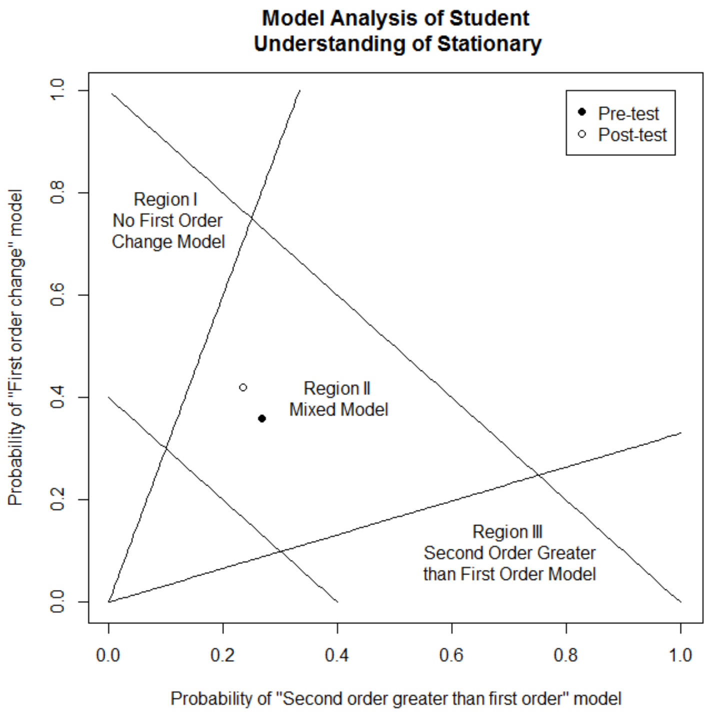


Figure 5.11: Model analysis of student understanding of the 'Stationary Concept'. This plot together with Figure 5.12 looks at student understanding of the 'Stationary' Concept. This plot shows that students moved away from the 'second order greater than first order' misconception towards the correct model.

Together Figures 5.12 and 5.11 explore the change in student's understanding of first order and second order changes. During instruction students moved away from the 'second order greater than first order', however they became more likely to use the 'no change' model. This is evidence for hypothesis that the 'no change' misconception is prevalent among experienced physicists.

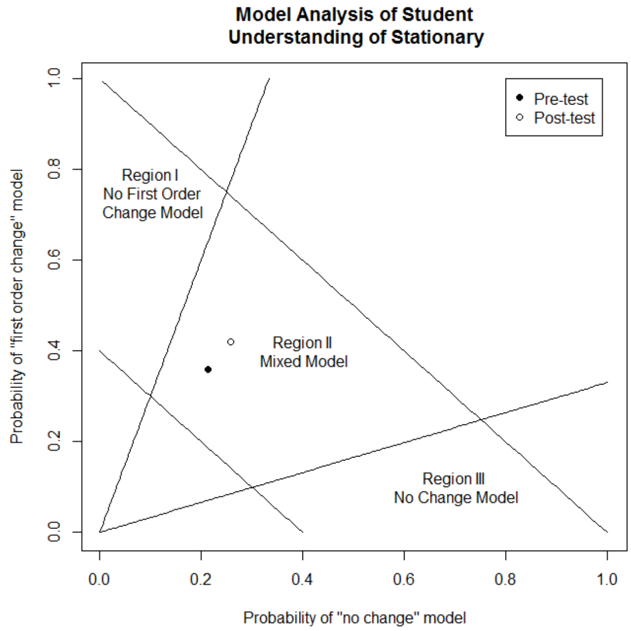


Figure 5.12: Model analysis of student understanding of the 'Stationary Concept'. This plot together with Figure 5.11 looks at student understanding of the 'Stationary' Concept.

Overall the model analysis shows that students understanding of the Principle of Stationary Action greatly improved, which is consisted with the gains observed for Questions 6, 7, 9 and 16. Students understanding of quantum mechanical concepts (Explore all Paths) was almost unaffected by instruction. Although overall the class's understanding of stationary improved, the number of students who held the 'no change' misconception increased.

5.5 Validity

5.5.1 Validation Interviews

To establish the validity of each of the questions, 20 validation interviews were performed with the first year students after they had taken the Action Concept Inventory post-test. The process used in these interviews is described in detail in Section 3.7. Most questions were found to be valid, students only picked the correct answer if they were using expert-like thinking. In response to some student misinterpretations, some revision of wording were made for Questions 2, 4, 5, 10, 12 and 15.

Question 11 was found to be invalid, students knew that Action could not be calculated for a path unless it was parametrised by time (displaying expert thinking). However students did not know if this corresponded to being able to calculate infinitely many Actions or no Actions (note expert reviewers picked up on this problem as well). As a result question 11 was revised.

Firstly the wording of the question was changed to emphasise the different axes being used for the two motions. Secondly the question was changed so that instead of asking about the number of Actions that can be calculated it instead asks: 'for which of the motions is it possible to calculate the Action?'

Question 13 was also found to be invalid see Section 8 *Appendix 1: Evaluating Question 13.*

After the think aloud component of the interview I probed the students' understanding of the quantum mechanical concepts. In general, students lacked confidence when discussing the quantum mechanical concepts. Many students displayed strong misconceptions about the 'Explore All Paths' concept. The most common misconception was that a quantum mechanical particle ***can*** take any path, but it only *chooses* one path. These students fundamentally misunderstood the idea of superposition.

Three students demonstrated that they understood the 'Explore all Paths' concept, of these only one was able to explain the remaining quantum mechanical concepts. Overall it appears that the students in the PHYS1201 class did not understand the quantum mechanical principles. This is what the concept inventory measured supporting the validity of the inventory.

5.5.2 Expert review

Four international Action experts reviewed the Action Concept Inventory. As Adams and Wieman predicted there were many small suggestions about tightening of wording of questions [25]. I followed the recommendations and did not necessarily implement all changes recommended about tightening the wording, especially when validation interviews showed that students had no trouble interpreting the question.

One example of this is question 1. The preamble of question 1 begins with "Two identical balls are originally stationary at the top of two frictionless ramps. Ramp I is twice as steep as Ramp II." Two Action expects claimed that this question was ambiguous: does twice as steep mean twice the slope or twice the angle. The correct answer is the same regardless of which interpretation is made and validation interviews showed that students had no problems interpreting this question. Therefore I did not change the wording in this case despite the expert review.

As the result of expert review there were small corrections made to most questions. The definitions and preamble on the front page of the inventory were changed substantially. In addition the preamble on question 10 was reworded.

The Action experts also pointed out the invalidity of question 11: *"No Action versus an infinity number of Actions? The point is that with no time information it is impossible to calculate a unique Action, but I don't know if this means no solution or an infinity number of solutions."*

One of the Action experts noted that the diagram for question 8 and 9 was inaccurate, as path IV needed to be parabolic for this question to be technically correct. Although validation interviews showed that students had no problem interpreting the diagram, I used R to plot create a more accurate diagram.

The wording for question 8 was also technically inaccurate, it originally read "If there are no forces acting on the electron, which of the paths *could it have taken* between the measurements at A and B?". This is encourages the misconception that the particle takes only one of many paths available to it. The wording was changed so it read "which of the paths *did it take*", emphasising that the particle actually takes all paths.

Because of this technical inaccuracy question 8 was invalid and was be revised. This was an opportunity to incorporate some of the discovered misconceptions about the 'Explore all Paths' concept into the Action Concept Inventory.

5.5.3 Gain

The average normalised gain for the inventory was calculated to be $G = 0.09$. This is quite low for an experienced instructor using interactive learning techniques, for which I would expect to see average normalised gains of the range $0.2 - 0.5$. Two years ago, John Aslanides observed a averaged normalised gain of $G = 0.38$ for the relativity concept inventory (note: this difference could be partially explained by Craig's emphasis on concepts when teaching relativity compared to his focus on computation when teaching Action). Immediately this would put the validity of the inventory into question. To investigate what was going on, the item gain for each question is shown in Figure 5.13 below:

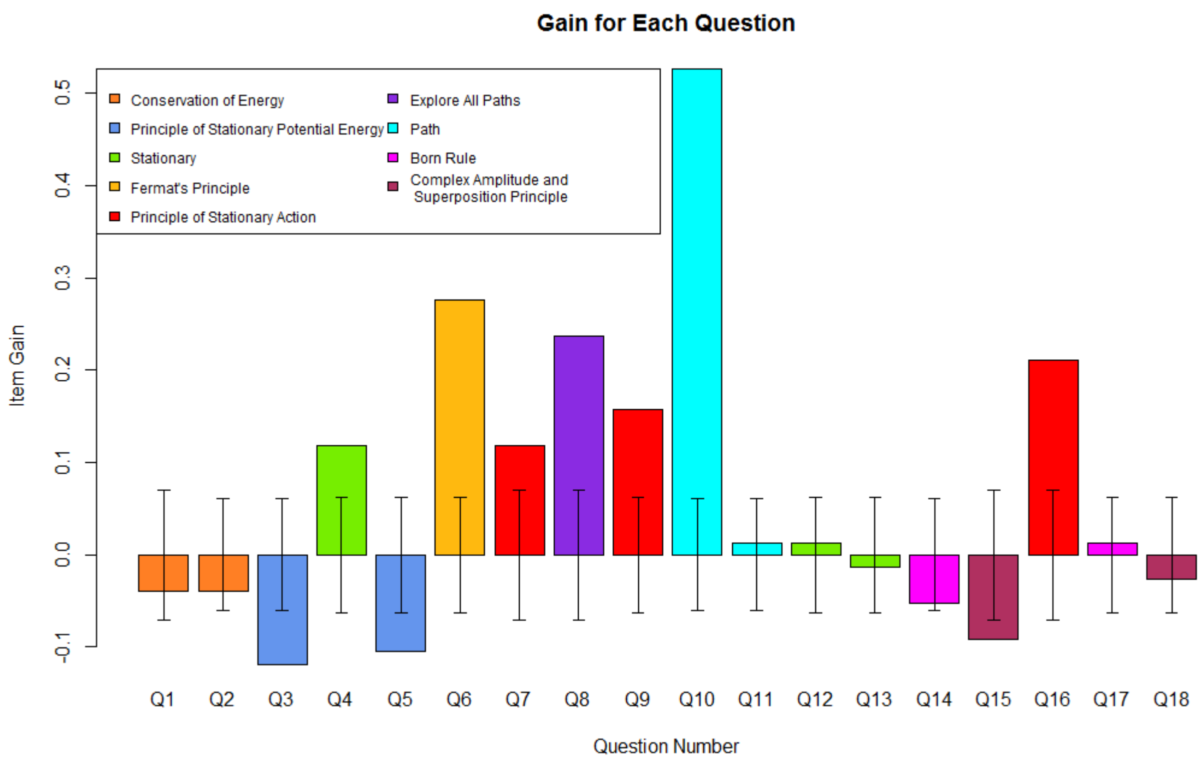


Figure 5.13: Item gain is the change in the fraction of students who answered a certain question correctly. Here the gain is shown for each of the Action Inventory Questions. If students randomly guessed answers we would anticipate that the average mean gain for question would be 0. One standard deviation of the fluctuation you would expect due to random guess is shown by the error bars.

Since the students were taught Action for two and a half weeks between the two tests, we should expect to see a positive gain for each question as negative gain indicates that students performed worse on the post-test than the pre-test. For questions where negative gains were observed, some sort of explanation is needed to justify that these questions are still in fact valid.

Conservation of energy was not taught by Craig Savage in the Action unit of PHYS1201 (unfortunately he did not cover time symmetry when explaining Noether's Theorem), so we would expect to see no significant gain in Question 1 and Question 2, which was observed. Therefore this does not contradict but indeed support the validity of Questions 1 and 2.

The lack of gain in Questions 4, 12 and 13 agrees with the factor analysis (Figure 5.12) the models students were using became more consistent, but they were no more likely to engage the correct model in the post-test than the pre-test. This showed that students did not learn the mathematical meaning of what is meant by stationary (in the context of stationary points or stationary paths).

The negative gain in question 3 and question 5 is not unexpected. The principle of stationary potential energy was not addressed directly in the course. The level of mathematical detail to correctly answer question 3 was not covered in the course.

Quantum Mechanical Concepts

The low gain on Question 8, 14, 15, 17 and 18 can be explained by the amount of time spent teaching quantum mechanics. Less than one lecture was spent teaching the four quantum mechanical principles and part of this teaching was immediately before the pretest. It is not surprising that the student scores on these questions hardly improved.

Student homework and follow up interviews confirmed that students had not grasped the quantum mechanical concepts that underpin the Principle of Least Action. For the homework task, students had to give a 200 word answer explaining the physical consequences of regions of constructive interference. Marking these responses showed that students had many misconceptions about the explore all paths concept. Many students believed that a particle 'sniffs out' all paths, but then it still only takes one. Twenty follow up interviews were performed with the students to further probe their understanding of the quantum mechanical concepts. Three students were able to explain the explore all paths concept and only one student was able to successfully explain all of the quantum mechanical concepts (this particular student had read Feynman's QED [28] before taking the course).

The data suggests that one lecture was insufficient time to teach the quantum mechanical concepts. Craig tried his own way of teaching the quantum mechanical principles that underpin Action. This method had not been tried and iterated as many times as the way of introducing the Action quantum mechanical concepts used by Feynman [28] and recommended by Edwin Taylor [32].
I would recommend that future instructors allocate more time to the quantum mechanical concepts which underpin Action. Knight [3] states that one of the biggest lessons learned from physics education research is to deal the student misconceptions explicitly. This could not be done when teaching Action to the PHYS1201 class because the common misconceptions had not been identified. Now some of these misconceptions have been established (see Section 5.4 and Section 5.6) and can be used to improve future Action instruction.

5.5.4 Correlations and establishing Validity

In order to establish the validity of the test it is important to find correlations between questions. These correlations indicate that questions are in fact testing the same concept. A very important part of this analysis is determining whether the correlation values are statistically significant or not. If you assume Gaussian statistics you can simple use Equation 4.11 to obtain a t-value and use this to calculate the corresponding p-value.

Another method of determining the significance of the correlations, without assuming Gaussian statistics, is using Monte Carlo simulations. This technique was used by John Aslanides and Craig Savage [56] to evaluate the strength of correlations in the relativity concept inventory. I have overviewed the technique in the statistical analysis chapter 4.3.

Although I recommended a more systematic method in the statistical analysis section, a good starting point to get an estimate of the significance of correlations is to look at a histogram of all the correlations as shown in Figure 5.14.

At a first glance it appears that there is an obvious break with two correlations with values $r > 0.5$. These seem to be different to the other correlations and we may expect these to be significant. According to the more systematic approach, all the expected correlations were checked in both pre-test and post-test. The significance of each correlations was found using both Gaussian and Monte Carlo Methods. The results are summarised in Table 5.1 on the next page.

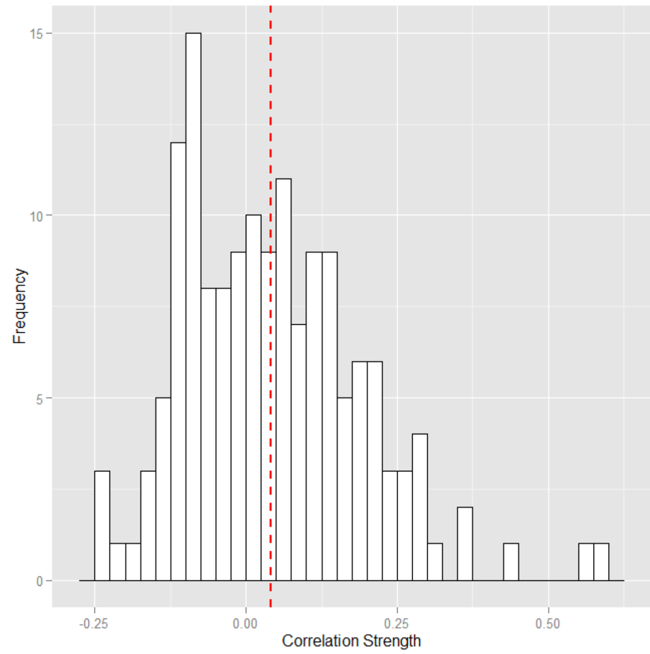


Figure 5.14: Histogram of all 153 correlation values calculated. The mean correlation (0.042) is indicated by the red dashed vertical line. This mean correlation could be used to try and compensate for student intelligence, see end of Section 4.3. A first glance shows there are 2 correlations which stand out as significantly different.

Although not all the correlations are statistically significant, every question pair is positively correlated in the original data (see 3rd column of Table 5.1). Although these correlations alone are not enough strong to establish the validity of every concept pair, it certainly supports the hypothesis that these question pairs test the same concept and that the inventory is valid.

Table 5.1: Table of Correlations between questions. The strongest correlations between each of the question pairs from both the pre-test and post-test were included in the table. Using the Bonferroni correction the p-value required for statistical significance is $p = 0.0036$. Significant correlations are shown in bold. Note the correlation between Q4 and Q12 is close, but not statistically significant by this criterion.

Concept	Question Numbers	Original Correlation	p-value (Gaussian)	p-value (Monte Carlo)	Rasch Model Correlation
Principle of Stationary Action	7, 9	0.13	0.1331	0.13	-0.12
	7, 16	0.22	0.028	0.0265	0.04
	9, 16	0.17	0.0715	0.0692	0.04
	4, 12	0.3	0.0042	0.00816	0.06
Stationary	4, 13	0.09	0.2198	0.216	-0.02
	12, 13	0.14	0.1138	0.1125	0.17
Principle of Stationary Action/Fermat	6, 7	0.25	0.0147	0.0143	-0.01
	6, 9	0.14	0.11385	0.11	-0.15
	6, 16	0.57	$\mathbf{3 \times 10^{-8}}$	$> \mathbf{1 \times 10^{-6}}$	0.4
Conservation of Energy	1, 2	0.23	0.02275	0.0217	0.14
Principle of Stationary Potential Energy	3, 5	0.14	0.1139	0.111	0.13
Complex Amplitude	15, 18	0.54	$\mathbf{2.2 \times 10^{-7}}$	$\mathbf{2.2 \times 10^{-5}}$	0.51
Born Rule	14, 17	0.59	$\mathbf{1 \times 10^{-8}}$	$\mathbf{3.95 \times 10^{-4}}$	0.59

5.5.5 Factor Analysis

R has multiple principle component and factor analysis functions and packages available for free online (see [102]). To perform the factor analysis I chose a function that was specifically designed for factor analysis 'factanal()' because it presented the results of the analysis in the most convenient format. I checked the results of the 'factanal()' function which is based on the maximum likelihood method of exploratory factor analysis and it gave similar results to the 'princomp()' and 'principal()' functions which are basic principle component analysis functions [102].

The 'nFactors' package was used to test for the number of factors that were significant. It was used to produce the scree plot shown in Figure 5.15. Of the four different methods used to calculate the number of significant factors (see [96] for more details), on average they suggest 7 factors were significant. Therefore the loadings of the first 7 factors were investigated.

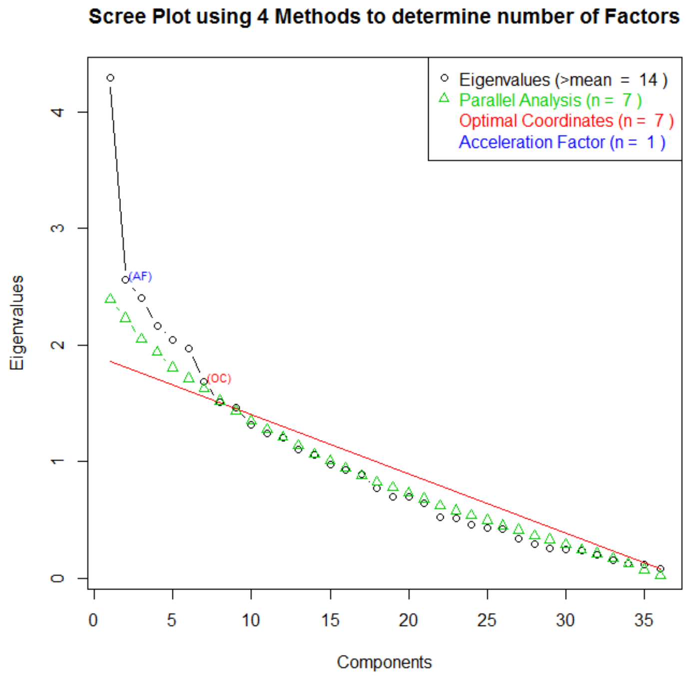


Figure 5.15: A Scree Plot Indicating how many factors should be selected as significant. This Scree plot was generated using R-code provided in the 'plotnScree' function from the 'nFactors' package [103]. This analysis suggests that only the first 7 factors are significant and should be considered.

The loadings which make up factors 2, 3, 4, 5 and 7 are shown in Table 5.2 and Table 5.3 below. Factor 1 is not included here because it contains many loadings and could simply be interpreted as a general intelligence. Factor 6 was not included because there was no easy interpretation of its meaning.

Table 5.2: Table of Factors 2, 3 and 7 and their loadings with values greater than 0.2. Loadings in italics are shown in the top section of the table and correspond to factors which can easily be interpreted as all grouped questions correspond to the same concept. These factors correspond to 'The Principle of Least Action' concept. Loadings which are bolded indicate test-retest reliabilty.

Factor 2		Factor 3		Factor 7	
Component	Strength	Component	Strength	Component	Strength
PostQ9	***0.499***	*PostQ7*	*0.319*	*PostQ7*	*0.23*
PreQ9	***0.44***	*PreQ16*	*0.419*	*PostQ9*	*0.221*
PostQ16	*0.373*	*PreQ6*	*0.297*	*PostQ16*	*0.504*
PostQ7	***0.252***				
PreQ7	***0.22***				
PostQ6	*0.223*				
PostQ1	**0.589**	**PreQ5**	**0.619**	**PreQ2**	**0.307**
PreQ1	**0.269**	**PostQ5**	**0.65**	**PostQ2**	**0.223**
PreQ3	0.381	PostQ1	0.287	PostQ17	0.232

Table 5.3: Table of Factors 4 and 5 and their loadings with values greater than 0.2. Loadings in italics are shown in the top section of the table and correspond to factors which can easily be interpreted as all grouped questions correspond to the same concept. Factor 4 corresponds to the 'Born Rule', Factor 5 corresponds to 'The Superposition Principle' and 'Complex Amplitude'. Loadings which are bolded indicate test-retest reliabilty.

Factor 4 Component	Strength	Factor 5 Component	Strength
PostQ14	*0.967*	*PreQ15*	*0.771*
PostQ17	*0.552*	*PreQ18*	*0.658*
PreQ14	*0.253*		
		PostQ10	0.344
		PostQ17	0.292

Table 5.2 shows that Q6, Q7, Q9 and Q16 are all closely correlated. Although these questions all had positive correlations when grouped together, individually none of these correlations were statistically significant. However the fact that these questions came up multiple times in the same factor in the principle component analysis further supports that they are all testing the same concept, establishing their validity. Table 5.3 further validates the questions testing the 'Born rule' and 'The superposition Principle' and 'Complex Amplitude'.

Tables 5.2 and 5.3 shows that many pre-test and post-test question pairs were grouped in the same factor (shown in bold). For topics where we can reasonably assume that students did not learn, (there was very low gain for the particular question or the material was not covered by Craig Savage in the lectures) these question correlations could be used as a measure of test-retest reliability.

Questions 1, 2, 3, 5, 11, 12, 13, 14, 15, 17 and 18 all showed either very low or negative gain. I calculated the correlations values for each of these and it was found that every pretest-posttest pair had a positive correlation. In addition Questions 1, 2, 3, 5 and 14 all had correlations of 0.3 or greater (which is statistically significant using the Bonferroni procedure in this case). Of these, questions Q1, Q2, Q5 and Q14 had post-test and pre-test loadings of greater than 0.2 in one of the significant factors. This indicates strong test re-test reliability.

5.6 Student Homework

Student's grades on the two of the extended response Action homework (Sections 9.5.2 and 9.5.3) were correlated with the inventory post-test questions. I wrote the homework questions to try and probe student thinking about paradoxes in Action physics.

The second homework question, relates to 'The Principle of Stationary Action' and 'Fermat's principle' concepts (see Section 9.5.2). Assuming that both the homework question and the inventory items both measured understanding of these concepts the two scores should correlate (concurrent validity). The correlation was calculated between the second homework question and Questions 6, 7, 9 and 16 of the inventory. The homework score correlated equally well with all of these questions, all four correlations strengths were in the range of 0.17 − 0.2. This indicates that the inventory questions and the homework question are all testing the same variable, providing further evidence towards the validity of the questions.

The third homework question relates to the 'Explore All Paths' concept and therefore should correlate with Inventory Question 8 (see Section 9.5.3). Homework score for this question was correlated with both pre-test and post-test Q8 score. The correlations were positive, but very weak (0.073 and 0.014). Expert review and validation interviews gave reasons to question the validity of question 8. The lack of correlation with the homework question provides further reason to suspect that this inventory question was not valid.

Marking the Student homework also brought to light some common student misconceptions, especially concerning the 'Explore All Paths' concept. Most misconceptions stemmed from the fact that students didn't understand/couldn't accept superposition states. These misconceptions are outlined in Section 3.3 *Explore All Paths*.

5.7 Attitude Survey

All up 93 students took that attitude survey as a voluntary component of their final Action homework. The students were asked "Do you think Action physics should be included in PHYS1201 next year?". 96% of students (89 students) responded yes.

5.7.1 Is teaching Action to First Year Physics Students Viable?

The major focus of the survey was to answer the questions 'Can first years, really learn Action?' and 'Is Action useful and interesting for students to study?'

Students were asked whether they felt studying Action had contributed to their overall understanding of physics. The results are shown in the histogram shown in Figure 5.16. This shows that students feel they gain an better understanding of physics when studying Action compared to other topics in physics.

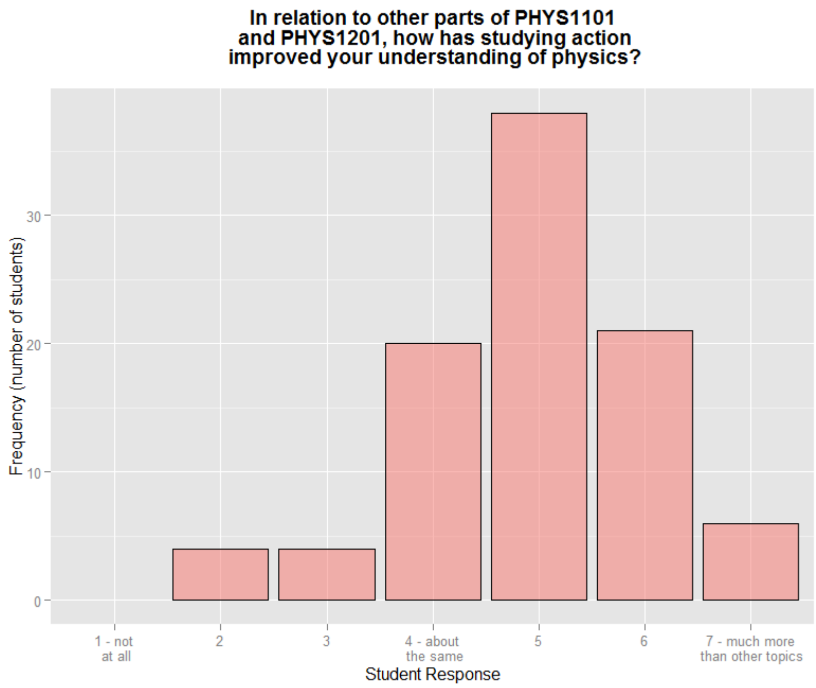


Figure 5.16: Student responses to "In relation to other parts of PHYS1101 and PHYS1201, how has studying Action improved your understanding of physics?" in the form of a histogram.

Students were asked to indicate how difficult they found each topic covered in first year physics. The question took the form “Compared to classical mechanics in PHYS1101, rate from 1 to 7 how difficult you found each of the following topics:", where 4 is about the same, 7 is much more difficult and 1 is much easier. (The students were also asked to compare their interest for each topic and how motivated they were to learn the topic (treating classical mechanics as 4, the neutral option)) The results from this question are summarised in the barplot shown in Figure 5.17.

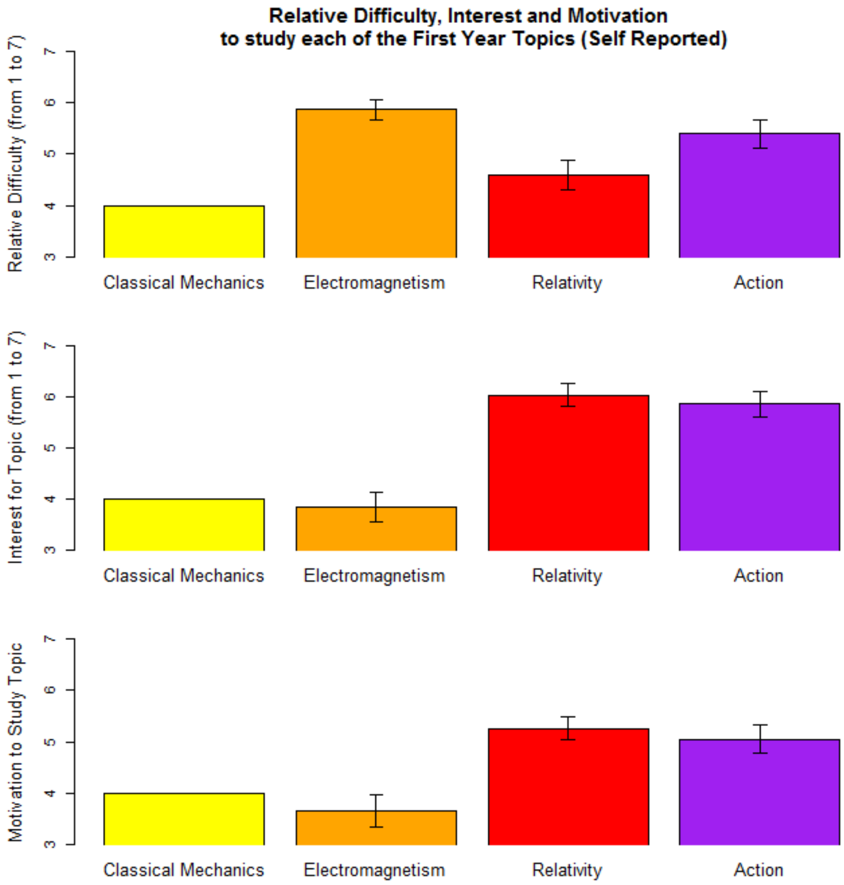


Figure 5.17: Each bar plot shows how difficult (self reported), interesting and how motivated students were to study for each of the topics covered in PHYS1101 and PHYS1201. Classical mechanics was set to 4 and students were asked “Compared to classical mechanics in PHYS1101, rate from 1 to 7 how difficult you found each of the following topics.", “Compared to classical mechanics in PHYS1101, rate from 1 to 7 how interesting you found each of the following topics:", “Compared to classical mechanics in PHYS1101, rate from 1 to 7 how motivated you are to learn about each of the following topics:"

This survey data we can conclude that students find Action more interesting than classical mechanics (Newtonian mechanics). Although students find studying Action more difficult than studying classical mechanics they find it no more difficult than electromagnetism, a topic that appears in many first year curricula. This supports the hypothesis that teaching physics through Action could be a more effective way to engage students than teaching introductory physics in the traditional way through Newtonian mechanics.

5.7.2 How do we Teach Action Most Effectively?

Students were asked what they enjoyed most about studying Action and what they found most challenging. The results are summarised in Figure 5.18 and Figure 5.19. The survey indicated that students most enjoyed seeing a new perspective on physics and seeing the link between different areas of physics. To capture students' interest I would recommend that instructors emphasise these two concepts in their instruction.

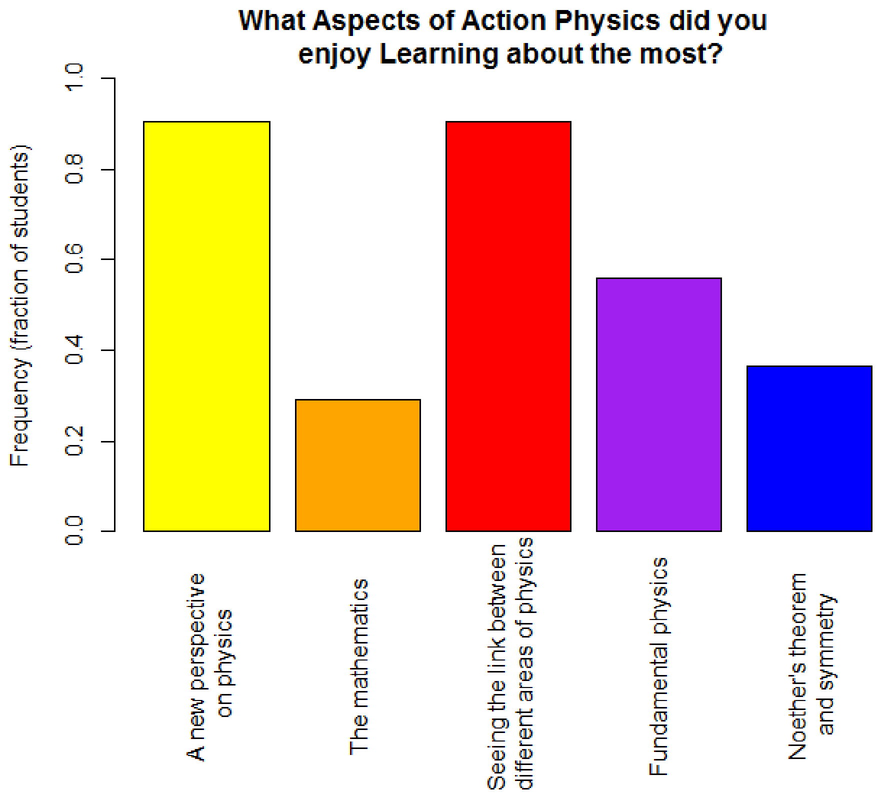


Figure 5.18: Student responses to 'What Aspects of Action Physics did you enjoy learning about the most?' This indicates that students enjoyed a new perspective on physics and seeing the link between different areas of physics.

Students could leave additional comments about what they enjoyed most about the course. Two students specifically mentioned that they enjoyed seeing a quantum mechanical explanation of classical laws and another student enjoyed seeing how Action fitted in with relativity. Two students indicated that they enjoyed learning Lagrangian mechanics because it allowed them to solve difficult problems with ease.

Figure 5.19 summarises what students found most difficult about learning Action physics. Although the mathematics behind Action physics was difficult, this data shows that it was not what students found most challenging about learning Action physics. Students found learning the concepts just as difficult as the mathematics. The most challenging part of the course was learning to use Mathematica.

In the comments section four students indicated that they found the most difficult part of Action physics was finding the Lagrangian when attempting a physics problem. Two students left additional comments emphasising that they found Mathematica particularly challenging and one student found interpreting partial derivatives challenging. Two students indicated that they found applying Action concepts difficult. One student found mathematical derivations difficult and one student commented that Action was difficult because it was new.

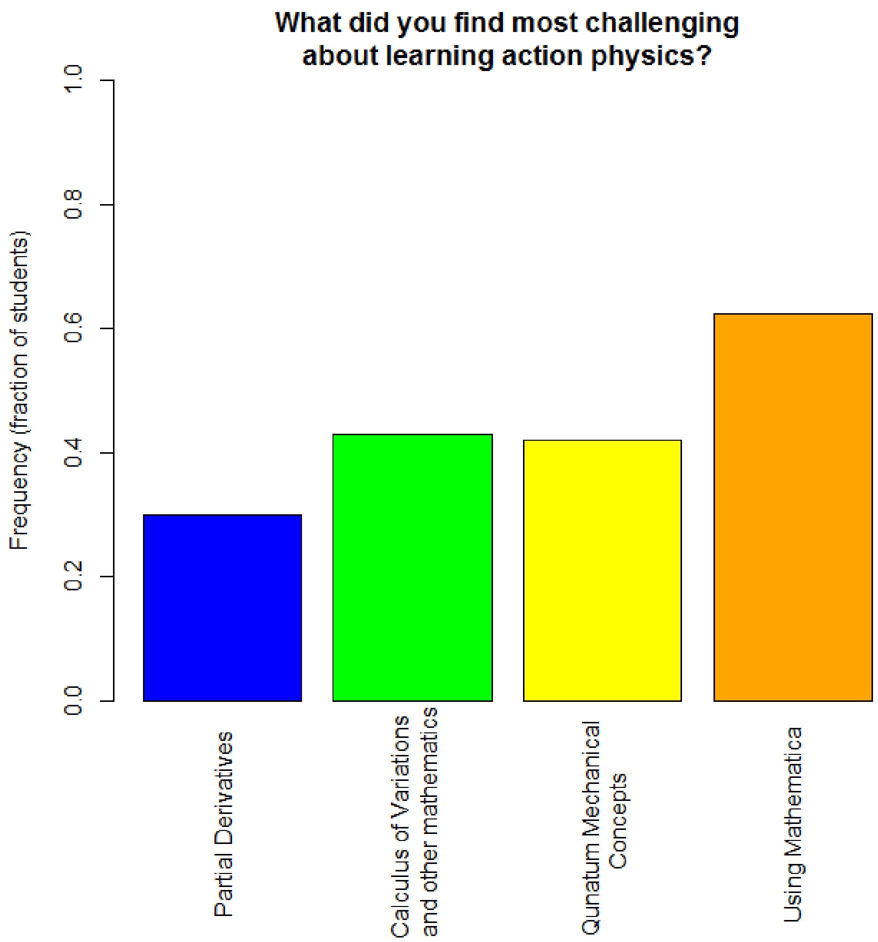


Figure 5.19: Student responses to 'What did you find most challenging about learning Action physics?' This data shows that although many students found the mathematics difficult it was not the most challenging part of the course. This undermines the hypothesis that Action cannot be taught to first years because the mathematics behind it is too difficult.

5.7.3 Additional Comments

In the survey there was space for students to leave additional comments. I have included the following quotes because I believe they give insight into the pedagogy of teaching physics through Action.

Quote 1 highlights the capacity of teaching introductory physics through Action to take advantage of the 'spacing effect' [104] [105]. By introducing Action concepts in first year, including quantum mechanics and higher level mathematics, students will have time to become familiar with the new ideas and material. Spacing effect reserach suggests that students will retain this information better in their long term memories if it is retaught some time later.

Quote 2 is additional evidence that the quantum mechanics instruction provided to the students was insufficient. In future I would recommend that instructors stick to the teaching method used by Feynman as Action instructors have focused on gradually improving this method.

> *"I really enjoyed learning about this new perspective on physics and speaking to other people apparently you would normally only start this in 3rd year? That's very cool. I am definitely supportive of keeping this in the first year curriculum. The advantage of having it introduced now is that it can take time to sink in and we can get used to it, so that when it becomes more complicated we have a more solid foundation to work with."* (1)

> *"Quantum mechanics felt like the elephant in the room in this unit. We probably didn't have time to cover it, and I realize that there are units on it later in the syllabus; but it felt like it was necessary to have background knowledge of it to understand or appreciate some parts of the Action course. It would be nice to have a lecture or some readings on relevant information, for people who have never studied quantum mechanics before."* (2)

These quotes summarise the main findings about Action pedagogy. Firstly that the students struggled to understand the quantum mechanics and the teaching methods used in this area should be revised. Secondly, although the mathematics required to learn Action physics is difficult, it did not prevent the first year students from learning Action. Finally the students really enjoy learning about Action and nearly all of them think that it should be included in the introductory physics curriculum.

In the next chapter *Conclusions* I explore these conclusions and also summarise the other key findings from the analysis performed in this chapter.

Chapter 6

A Novel Idea: A second Dataset

To further validate the concept inventory I asked academics, PhD students and higher level undergraduates to complete the Action Concept Inventory. Although concept inventories have never been statistically validated on PhD students or academics before, I hypothesised this could be useful as a second data set, particularly to draw conclusions about the quantum mechanical questions. Very few students answered the quantum mechanical questions correctly in the first year class and although the correlations between the quantum mechanical questions were very strong, the statistical conclusions would have been based on a very low number of students (only 6 students answered Q15 correct in the post-test).

It took a significant amount of time to go around to offices and explain the my project and the inventory, so once 50 participants had completed the inventory, I stopped seeking more (there was some lag between asking and the completing of the inventory, so I ended up with more than 50 participants). Participation of the physicists in the Action Concept Inventory was anonymous, though they were asked to leave demographic information.

This second data set was extremely useful. The data confirmed the conclusions I drew from the first year data. In addition some correlations that were strong but not significant in the original data were significant in the second data set, allowing me to confirm the validity of more questions.

Summary statistics

In total 61 physicists from the ANU Research School of Physics and Engineering took the Action Concept Inventory. Of those who completed the inventory 82% male. 36% were academics, 38% were PhD students and the remaining 26% were honours, masters or 3rd year students. 36% of respondents said that they were familiar with Action, a further 40% said they had studied Action but were no longer familiar and 24% said that they had never studied Action at all.

Analysis of the data

I calculated the significance of correlations using the Gaussian and Monte Carlo methods and used factor analysis to find groupings of questions in the academic data.

The correlations were much stronger in the academic data than the student data. The 'nFactors' package in R suggested that only one factor from the factor analysis was significant. Previously I have not considered the first factor because it is likely to be a factor summarising general intelligence, however since it was the only significant factor I decided to include it in the Aslanides plot for this data set. Two of the other factors were easily interpretable and were the same as the groupings found in the student data (correlations between Questions 14 and 17 and between Questions 6,7,9 and 16). The item difficulties, correlations and factor analysis are summarised in Figure 6.1 below.

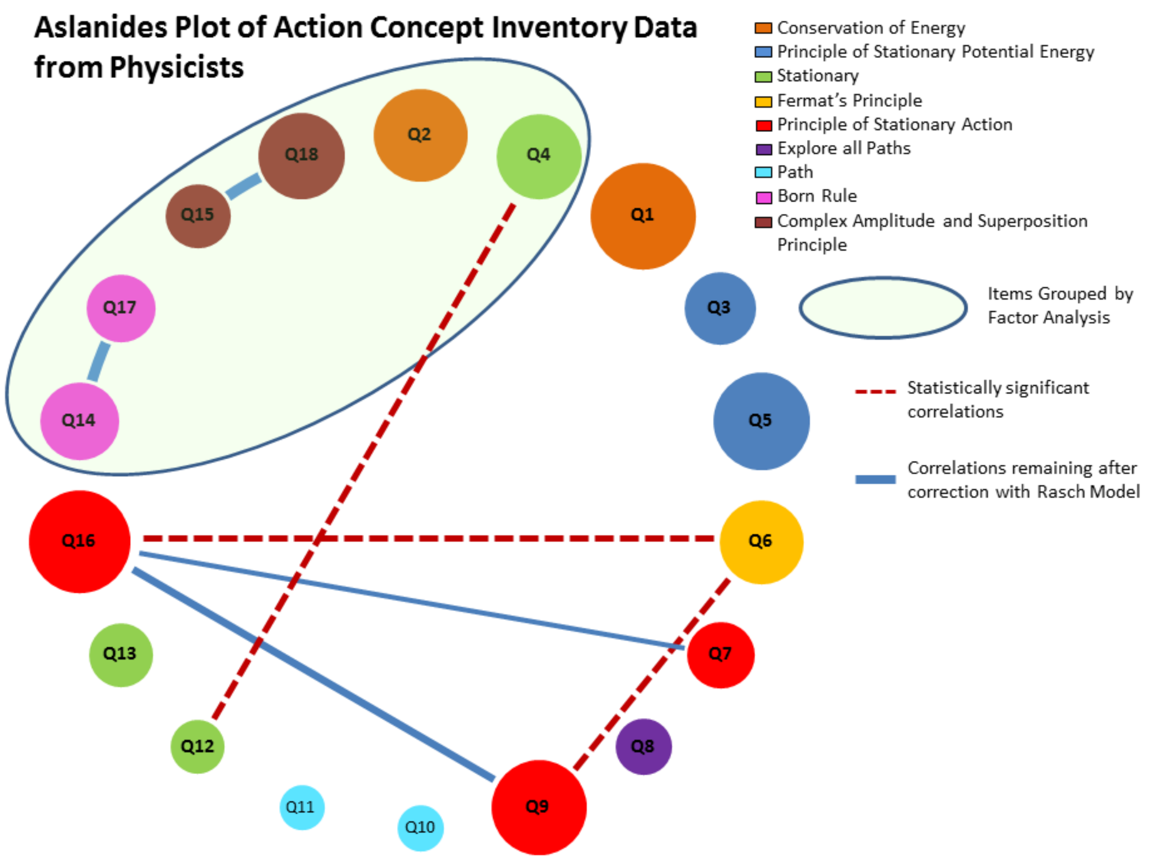


Figure 6.1: Aslanides plot of the second Action Concept Inventory data set (from physicists). The Colour of each question indicates the concept being tested. The size of each circle indicates the difficulty of the question (difficulty index is proportional to circle area). Lines between questions indicate statistically significant correlations, where width of line indicates strength of correlation. Solid blue lines indicate correlations that were still greater than 0.2 after correcting with Rasch Model for intelligence. Note that the Bonferroni Procedure was used meaning that for statistical significance $p < 0.0036$ 4.3.1, however for the correlations between Questions 7 and 16 and between Questions 6 and 9 only the Gaussian p-values satisfied this condition, Monte Carlo simulation predicted values of 0.0051 and 0.0040 respectively. To see the equivalent plot for the student data see Figure 7.1

On most questions the difficulty indicies were higher (questions were easier) for the physicists than the first year students, particularly the quantum mechanical questions. However there were some questions where the first year students actually did better than the physicists. These were questions 7, 8, 10 and 13. Both question 8 and question 13 have already been established to be invalid 5.5.1. Question 8 was a quantum mechanical question therefore we would expect the physicists (who have all studied quantum mechanics) to perform better than the first year students. Since this was not the case this confirms that question 8 is invalid.

There is no particular reason to believe that the first years would be better or worse than the academics on question 7, and the difference in difficulty index was 0.02 so I am not treating this result as significant (1 standard deviation in result you would expect to see due to random guessing is 0.05).

However the 1st year students did significantly better than the academics on question 10, the difference in difficulty index was 0.54. Question 10 requires knowledge of the boundary conditions required to calculate the Action of a path. The first years calculated the Action of a path multiple times in tutorials and homework during the two weeks leading up to the inventory while it is possible that the physicists had never calculated the Action of a path before. Therefore it makes sense that the 1st year students performed better on this question.

Besides confirming the correlations and groupings that already existed in the student data, the lack of correlations for questions 13 and 11 is consistent with the hypothesis that these questions were invalid (If questions 11 and 13 were valid I would expect them to correlate strongly with other questions which tested the same concept.)

A New Validation Technique

Testing the physicists confirmed the findings based on the student data set. In addition the second dataset had a stronger correlations than the student data, allowing previously inconclusive correlations to be established. This finding that physicists can be used as a sample to obtain statistical data is significant in itself because it opens up a new method of validating concept inventories.

Usually first year classes are ideal for validating inventories because there are a large numbers (hundreds) of students. If the instructor encourages students to participate this allows you to easily get a large sample size which allows for robust statistics.

I have explored a new possibility, using older undergraduates, PhD students and academics to validate an inventory, and it appears to have been successful. In general this may not be a useful option for physicists looking to validate inventories, because academics are generally in short supply and have very little time. However I have shown that this is a possibility for future physics education researchers looking to validate inventories.

The biggest problem I can see with using higher level students and academics to validate inventories is that these physicists may interpret questions differently to 1st year students (indeed I hypothesise that the stronger signal in the second data set is due to academics having greater ability to correctly interpret questions than 1st year students, reducing random error due to misinterpretation). Therefore it is essential that validation interviews are carried out on the population the inventory is intended for. This will ensure that the target population interprets the questions correctly and establishes the validity of the inventory.

Chapter 7

Conclusions

The conclusions drawn in this chapter fall into three categories, conclusions about the Action Concept Inventory (Section 7.1), conclusions about the methods used to evaluate and deliver concept inventories (Sections 7.2 and 7.3) and conclusions about Action pedagogy (Sections 7.4, 7.5, 7.6).

7.1 The Action Concept inventory

Overall the inventory is valid, reliable and a good measure of student understanding of Action. Many of the standard measures of an inventory such as KR-20, point bi-serial correlation and discrimination index were not appropriate in establishing the reliability and discrimination of the Action Concept Inventory because it was designed to teach many concepts which are not closely related.

Reliability of questions 1, 2, 3, 5 and 14 were established using test-retest reliability. Validity was established by correlating inventory responses with student homework and through validation interviews with the 1st year students. The inventory data was found to be consistent with other sources of information about the class.

Expert review and validation interviews established the validity of all the inventory questions except for 8, 11 and 13. Analysis of correlations provided more evidence confirming the validity of questions 1, 2, 6, 7, 9, 14, 15, 16, 17 and 18. A summary of the statistical analysis is shown in the Aslanides plot, Figure 7.1. Correlation and difficulty index data indicated that question 3 was too difficult and probes unnecessary mathematical detail of the 'stationary' concept.

Questions 3, 8, 11 and 13 need to be modified then validated through student interviews. The modified inventory will need to be given to the PHYS1201 class in 2015 to allow statistical tests to be performed that will establish its validity before it can be published. Student homework and interviews showed that 'Explore All Paths' is a concept with many misconceptions. At least one more question should be added to the Action Concept Inventory to provide instructors with information about student misconceptions about this concept.

The Action Concept Inventory tests a large number of concepts for the number of questions on the test (18 questions for 9 concepts compared to 29 questions for 6 concepts on the Force Concept Inventory [26]). There are both advantages and disadvantages to this. The primary disadvantage is that many measures that are used frequently in concept inventory literature favour tests which have fewer concepts as these tests have high internal consistency. Having only pairs of questions corresponding to each concept limits the use of factor analysis to establish the validity of the inventory. Developers of future concept inventories should consider this carefully before choosing to include a large number of concepts in an inventory.

There is an advantage to testing a large number of concepts in the Action Concept Inventory and range of difficulty of the questions. The Action Concept Inventory will be more appropriate to use across a wider variety of students. Unlike force which is primarily taught at a first year level, Action is taught at many different year levels depending on the institution. The most effective methods of teaching Action have not been found. Therefore the wide range of concepts and difficulties allows physics education researchers to evaluate instruction of different year levels.

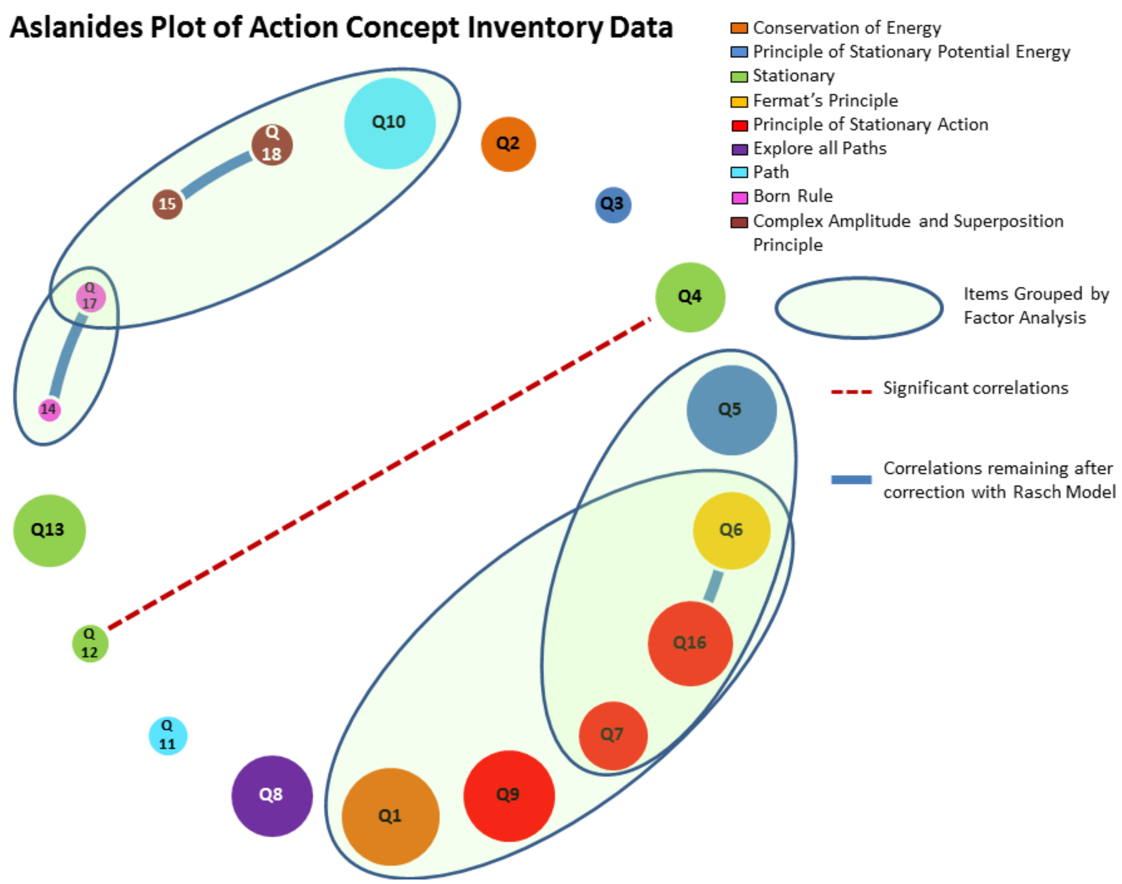


Figure 7.1: Aslanides Plot of the Action Concept Inventory post-test data. The colour of each question indicates the concept being tested. The size of each circle indicates the difficulty of the question (difficulty index is proportional to circle area). Lines between questions indicate statistically significant correlations, where width of line indicates strength of correlation. Solid blue lines indicate correlations that were still greater than 0.2 after correcting with Rasch Model for intelligence. There are statistically significant correlations between Questions 15 and 18 and Questions 14 and 17 and Questions 6 and 16 and these correlations remained after correction with the Rasch Model of intelligence. This established the validity of these questions. Factor anaylsis further confirms the validity of Questions 6, 7, 9, 14, 15, 16, 17 and 18. Note that the Bonferroni Procedure 4.3.1 was used meaning that for statistical significance $p < 0.0036$. Technically the correlation between questions 4 and 12 was not significant as $p = 0.0041$.

7.2 Analysis Methods: Lessons Learned

7.2.1 Monte Carlo Method

The Monte Carlo method is an extremely transparent and effective method of estimating the significance of correlations between pairs of questions. It gives values very similar to Gaussian statistics except in extreme cases where Gaussian statistics underestimates the frequency of rare extreme events. Because the Monte Carlo Method is computationally expensive, I believe it should only be used to calculate values for the extreme cases where Gaussian statistics fails. These cases are when there are extremely strong correlations or a large difference in question difficulty.

The Monte Carlo method could be used to correct for correlations simply due to student intelligence by setting r (used to generate the multinomial distribution) equal the mean correlation rather than $r = 0$.

7.2.2 Confidence

Using student self-reported confidence with other inventory data is not yet a well established technique. For the Action Concept Inventory self reported confidence data was used to establish the existence of misconceptions. Although this technique is yet to be well established there was evidence from this concept inventory study that shows it was a valid technique. In this study confidence data played an important role distinguishing good distractors from true student misconceptions.

To be consistent with findings from statistical and psychological literature a 7-point scale should be used when asking for self reported confidence. Confidence data was most meaningful when it was normalised for each student before taking means and averages.

7.3 Innovations for Delivering and Validating the Inventories

Distributing the Action Concept inventory during lab classes was an easy way to obtain a higher participation rate (as lab attendance is much higher than lecture attendance). Mika Kohonen suggested giving the concept inventory during labs and this proved to be much more effective for maximizing participation than other incentives and continuous promotion.

Rather than recording and then re-listening to interviews with students, it is much more convenient to take notes on a copy of the concept inventory itself during think aloud interviews. This is a quick and easy way of producing a copy of the inventory with all the appropriate notes and suggested changes in the relevant locations.

Think aloud interviews with multiple students was a more natural way of performing validation interviews. Students were able to have a discussion rather than simply talking at a non-responsive interviewer. It was noticed that when interviewed in pairs students would correct each other's misinterpretations of the questions.

This raised the question, 'Could inventories be given to groups of students to decrease the noise in the data due to student mis-interpretation?' This is a possible avenue for further research, to see if there is any merit to having students take concept inventories in pairs or small groups.

The Action concept inventory was statistically validated on higher level physics students and academics. The data generated from the physicists showed the same correlations as the student data, but the signal was even clearer. This adds more value to the inventory, it has now been validated for use across year levels not just first year students. In addition this opens up a new avenue for validating concept inventories, using academics and higher level students for statistical validation. Further research would need to be conducted before this could be considered a standard technique.

7.4 Action Misconceptions

Student homework revealed three frequent misconceptions about the 'Explore All Paths' concept:

1. The number of paths available for an object to move from point A to B increases as mass decreases.
2. Small (quantum mechanical) objects have more stationary paths than macroscopic objects.
3. Each path has a probability associated with it. Probability is higher for the stationary path. Macroscopic objects always take the stationary path, while quantum mechanical objects can 'choose' different paths.

Analysis of the frequency of each response and student confidence confirmed the existence of the 'minimum Action' misconception (that the path of minimum rather than stationary Action is taken). Student homework also revealed a strong misconception that the lower the Action of a path the more probable that path.

7.5 Recommendations for Future Action Instruction

The purpose of the Action Concept Inventory is to provide instructors with a way of measuring the effectiveness of their Action teaching and therefore improve it. Although the inventory was not fully validated, using it on the PHYS1201 class gave some valuable insights to improve the quality of Action teaching. For any instructors looking to learn from the work we have done, the main lessons were:

- Show students quantum mechanical laboratory exercises early on while learning the quantum mechanical concepts in lectures. This allows students to gain an intuitive understanding of the quantum mechanical principles that underpin the Principle of Stationary Action.
- Directly address the most common misconceptions of the quantum mechanical principles mentioned in Section 7.4. This recommendation is based on the finding that dealing with student misconceptions directly is the most effective way of correcting them [3].
- Use explicit examples of summing up the arrows or phasors as Feynman does in QED [28] and as Edwin Taylor recommends [32]. Note: It may take more than one lecture to go through an example and address the common misconceptions.
- Show students how Action links different topics in physics together: quantum mechanics, classical mechanics and relativity. The student surveys showed this was one of the things they enjoyed most about the Action unit.

7.6 Teaching Action as Introductory Physics is Viable

Observations of tutorials and student homework showed that first year students were able to learn the mathematics necessary to use Lagrangian mechanics. Student feedback shows that they didn't find the mathematics the hardest part of learning Action, it was no more difficult than learning the Action concepts.

Data from the inventory pre-test and post-test showed that students were able to understand the Principle of Stationary Action. Observation of students in tutorials showed that many of them were able to use Lagrangian mechanics to solve complex classical mechanics problems.

The survey showed that students appreciated learning Lagrangian mechanics as a new perspective on physics. Discussions with students and observation of students in tutorials showed that they were able to see situations where it will be simpler to use Lagrangian mechanics than Newtonian mechanics. The survey also showed that students found Action easier than electromagnetism, a topic that is taught routinely all around the world. They also found studying Action more interesting than studying classical mechanics or electromagnetism, and felt that studying Action increased their understanding of physics more than the other units they had studied.

In conclusion, I have taken one of the comments from the student surveys to sum up the potential of Action physics to inspire the next generation of budding physicists:

> "I want to thank you for the opportunity to learn something so profound and spectacular. When I was 14 I read the Elegant Universe by Brian Greene- everything he said was conceptually incredible, but he quite deliberately gave non-mathematical explanations for every topic. Great for a young and/or unexposed mind, but it left quite a few open-ended questions- many of which you have provided something of an answer for. Thank you for deepening my understanding of the world in which we live, even by a small amount. I'm inspired."

Bibliography

[1] Ian Chubb, Clare Findlay, Lucy Du, Bill Burmester, and Laura Kusa. MATHEMATICS , ENGINEERING & SCIENCE IN THE NATIONAL INTEREST. Technical Report May, Office of the Chief Scientist, 2012.

[2] Lorenzo J Curtis. A 21st century perspective as a primer to introductory physics. *European Journal of Physics*, 32(5):1259–1274, September 2011.

[3] Randall D. Knight. *Five Easy Lessons Strategies for Successful Physics Teaching*. Addison Wesley, San Francisco, 2004.

[4] Lillian C. McDermott. Resource Letter: PER-1: Physics Education Research. *American Journal of Physics*, 67(9):755, September 1999.

[5] Edwin F. Taylor. A call to action. *American Journal of Physics*, 71(5):423–425, 2003.

[6] Edwin F. Taylor. Principle of Least Action. `http://www.eftaylor.com/leastaction.html`, 2008.

[7] Jozef Hanc, Slavomir Tuleja, and Martina Hancova. Simple derivation of Newtonian mechanics from the principle of least action. *American Journal of Physics*, 71(4):386, 2003.

[8] Thomas a. Moore. Getting the most action out of least action: A proposal. *American Journal of Physics*, 72(4):522, 2004.

[9] Jon Ogborn, Jozef Hanc, and Edwin F. Taylor. Action on Stage: Historical Introduction. In *The Girep Conference*, pages 1–23, Amsterdam, 2006.

[10] Chris Gray. Principle of Least Action. `http://www.scholarpedia.org/article/Principle_of_least_action`, 2009.

[11] C G Gray, G Karl, and V a Novikov. Progress in classical and quantum variational principles. *Reports on Progress in Physics*, 67(2):159–208, February 2004.

[12] G Karl and Chris Gray. New variational principles in classical and semiclassical mechanics. *arXiv*, 2002.

[13] Edward F. Redish and Richard N. Steinberg. Teaching Physics: Figuring out What Works. *Physics Today*, 52(1):24, 1999.

[14] Carol A. Brewer and Diane Smith. Vision and Change in Undergraduate Education A Call to Action. Technical report, American Association for the Advancement of Science, 2009.

[15] John F. Donoghue, Eugene Golowich, and Barry R. Holstein. *Dynamics of the Standard Model*. Cambridge University Press, 2014.

[16] S J Weber, A Chantasri, J Dressel, a N Jordan, K W Murch, and I Siddiqi. Mapping the optimal route between two quantum states. *Nature*, 511(7511):570–3, July 2014.

[17] Michael Schirber. Focus: Curvey Photon Trajectories Could be Detectable. *Physics*, 7(96), 2014.

[18] A. Chantasri, J. Dressel, and a. N. Jordan. Action principle for continuous quantum measurement. *Physical Review A*, 88(4):042110, October 2013.

[19] Walter Simmons and Joel L Weiner. The principle of stationary action in biophysics: stability in protein folding. *Journal of biophysics (Hindawi Publishing Corporation : Online)*, 2013:697529, January 2013.

[20] Daniele Passerone and Michele Parrinello. Action-Derived Molecular Dynamics in the Study of Rare Events. *Physical Review Letters*, 87(10):108302, August 2001.

[21] Rahul Sawant, Joseph Samuel, Aninda Sinha, Supurna Sinha, and Urbasi Sinha. Nonclassical Paths in Quantum Interference Experiments. *Physical Review Letters*, 113(12):120406, September 2014.

[22] GL Eyink. Action principle in nonequilibrium statistical dynamics. *Physical Review E*, 54(4):3419–3435, 1996.

[23] RFW Bader. Principle of stationary action and the definition of a proper open system. *Physical Review B*, 49(19):348–356, 1994.

[24] RL Dewar. Interaction between Hydromagnetic Waves and a TimeâĂŘDependent, Inhomogeneous Medium. *Physics of Fluids (1958-1988)*, 13, 1970.

[25] Wendy K Adams and Carl Wieman. Development and Validation of Instruments to Measure Learning of Expert-Like Thinking. *International Journal of Science Education*, 33(9):1289–1312, 2011.

[26] David Hestenes, Malcolm Wells, and Gregg Swackhamer. Force concept inventory. *The Physics Teacher*, 30(3):141, 1992.

[27] Kirk Allen. *Statistical Concepts Inventory: Development and Analysis of a Cognitive Assessment Instrument in Statistics.* Phd dissertation, University of Oklahoma, 2006.

[28] Richard Feynman. *QED: The Strange Theory of Light and Matter.* Princeton University Press, 1985.

[29] Cassiopeia Project. Principle of Minimal Action and Kinetic Energy. `http://www.youtube.com/watch?v=AfsZiHEcoxk#aid=P-xeitdAXgE`, 2009.

[30] NGC1976. Quantum Electrodynamics. `http://www.allthingsscience.com/video/752/Quantum-Electrodynamics`, 2011.

[31] Daniel F. Styer, Richard Feynman, and Albert R. Hibbs. *Quantum Mechanics and Path Integrals: Emmended Edition.* Dover Publications, 2010.

[32] Edwin F. Taylor, Stamatis Vokos, John M. OâĂŹMeara, and Nora S. Thornber. Teaching FeynmanâĂŹs sum-over-paths quantum theory. *Computers in Physics*, 12(2):190, 1998.

[33] Edwin F. Taylor and Slavomir Tuleja. Principle Of Least Action Interactive. http://www.eftaylor.com/software/ActionApplets/LeastAction.html, 2003.

[34] Todd Timberlake. Least Action Gravity/Free. http://www.compadre.org/portal/items/detail.cfm?ID=12400, 2012.

[35] Clifford Swartz. EDITORIAL: Physics, Where the Least Action is? *The Physics Teacher*, 10(9):516, 1972.

[36] Encyclopedia Britannica. Fermat's Principle. http://www.britannica.com/EBchecked/topic/204700/Fermats-principle, 2014.

[37] B Nabelek, M Maly, and V Jirka. Linear Fresnel lenses, their design and use. *Renewable energy*, I(3):403–408, 1991.

[38] R Nityananda and J Samuel. Fermat's principle in general relativity. *Physical Review D*, 45(10):3862–3864, 1992.

[39] Encyclopedia Britannica. Huygen's Principle, 2014.

[40] A. J. de Witte. Equivalence of Huygens' Principle and Fermat's Principle in Ray Geometry. *American Journal of Physics*, 27(5):293, 1959.

[41] Jozef Hanc. The original Euler's calculus-of-variations method: Key to Lagrangian mechanics for beginners. *European Journal of Physics, Submitted*, pages 1–16, 2004.

[42] Richard Feynman. Feynman Lecture Principle of Least Action, 1963.

[43] Tomás Tycrská. The de Broglie hypothesis leading to path integrals. *European Journal of Physics*, 17(3):156, 1996.

[44] Jozef Hanc and Edwin F. Taylor. From conservation of energy to the principle of least action: A story line. *American Journal of Physics*, 72(4):514, 2004.

[45] R. Feynman. Space-Time Approach to Non-Relativistic Quantum Mechanics. *Reviews of Modern Physics*, 20(2):367–387, April 1948.

[46] Yuval Grossman. Derivation of the Euler-Lagrange equation. http://www.lepp.cornell.edu/~yuvalg/p3314/E-L.pdf, 2013.

[47] Jozef Hanc, Edwin F. Taylor, and Slavomir Tuleja. Deriving LagrangeâĂŹs equations using elementary calculus. *American Journal of Physics*, 72(4):510, 2004.

[48] Stephen T. Thornton and Jerry B. Marion. *Classical Dynamics of Particles and Systems*. Thomson Learning, 2004.

[49] Hebert Goldstein, Charles Poole, and John Safko. *Classical Mechanics*. Addison Wesley, San Francisco, 3rd editio edition, 2002.

[50] David Morin. *Classical Mechanics*. Cambridge University Press, Cambridge, 2008.

[51] Lim Yung-kuo, editor. *Problems and Solutions on Mechanics*. World Scientific Publishing Co. Pte. Ltd., Singapore, 1994.

[52] R. Douglas Gregory. *Classical Mechanics*. Cambridge University Press, New York, 2006.

[53] Jozef Hanc, Slavomir Tuleja, and Martina Hancova. Symmetries and conservation laws: Consequences of NoetherâĂŹs theorem. *American Journal of Physics*, 72(4):428, 2004.

[54] Dwight E. Neuenschwander, Edwin F. Taylor, and Slavomir Tuleja. Action: Forcing Energy to Predict Motion. *The Physics Teacher*, 44(3):146, 2006.

[55] Jon Ogborn and Edwin F Taylor. Quantum physics explains NewtonÂŠs laws of motion. *Physics Education*, 40(1):26–34, January 2005.

[56] J. S. Aslanides and C. M. Savage. Relativity concept inventory: Development, analysis, and results. *Physical Review Special Topics - Physics Education Research*, 9(1):010118, May 2013.

[57] Robert J Beichner and NC State Physics Department. Assessment Instrument Information Page. `http://www.ncsu.edu/per/TestInfo.html`, 2014.

[58] Coalition Foundation. Concept Inventories. `http://www.foundationcoalition.org/home/keycomponents/concept/`, 2008.

[59] Workshop Physics. Assessment Examinations Action Research Kit. `http://physics.dickinson.edu/~wp_web/wp_resources/wp_assessment.html`, 2012.

[60] Julia I Smith and Kimberly Tanner. Approaches to Biology Teaching and Learning The Problem of Revealing How Students Think : Concept Inventories and Beyond. *CBE Life Sciences Education*, 9:1–5, 2010.

[61] David Hestenes and Ibrahim Halloun. Interpretting the Force Concept Inventory. *The Physics Teacher1*, 33(NOV):502–506, 1995.

[62] James Corkins. *The Psychometric Refinement of The Materials Concept Inventory.* PhD thesis, Arizona State University, 2009.

[63] Douglas Huffman and Patricia Heller. What does the force concept inventory actually measure? *The Physics Teacher*, 33(3):138, 1995.

[64] Geoffrey Hellman. Mathematical Pluralism: The Case of Smooth Infinitesimal Analysis. *Journal of Philosophical Logic*, 35(6):621–651, 2006.

[65] Dwight E. Neuenschwander. Resource Letter NTUC-1: Noether's Theorem in the Undergraduate Curriculum. *American Journal of Physics*, 82(3):183–188, March 2014.

[66] R. Shankar. *Principles of Quantum Mechanics.* Plenum Press, New York, 1994.

[67] M. Born. Zur Quantenmechanik der Sto{\ss}vorg{\"a}nge. *Zeitschrigt fur Physik*, 37:863–867, 1926.

[68] Nobelprize.org. The Nobel Prize in Physics 1954 - Max Born. `http://www.nobelprize.org/nobel_prizes/physics/laureates/1954/born-facts.html`, 2014.

[69] Michel Gondran and Alexandre Gondran. The Principle of Least Action as Interpreted by Nature and by the Observer. *arXiv*, pages 1–8, 2014.

[70] B Gibb. *Test-Wiseness as Secondary Cue Response.* Dissertation, Stanford University, 1964.

[71] John Stewart Aslanides. *Relativity Concept Inventory.* Honours thesis, Australian National University, 2012.

[72] Andrew Conway. Princeton University Statistics One Coursera MOOC. `https://www.coursera.org/course/stats1`, 2013.

[73] Jim Morrison. *Statistics for Engineers.* John Wiley & Sons, Hoboken, NJ, USA, 2009.

[74] Paul Kline. *A Handbook of Test Psychology.* Menthuen, London, 1986.

[75] WebAssign. Welcome to WebAssign! `https://www.webassign.net/login.htmla`, 2014.

[76] Pearson. MasteringPhysics. `http://www.pearsonmylabandmastering.com/northamerica/masteringphysics/`, 2014.

[77] Jum C Nunnally. *Psychometric Theory.* Mcgraw-hill, New York, 1978.

[78] Theresa J.B. Kline. *Psychological Testing A Practical Approach to Design and Evaluation.* Sage Publications, London, 2005.

[79] RJ Beichner. Testing student interpretation of kinematics graphs. *American journal of Physics*, 62(8):750–755, 1994.

[80] Robert M. Kaplan and P. Saccuzzo, Dennis. *Psychological Testing Principles, Applications, and Issues Sixth Edition.* Thomson Wadsworth, Belmont, California USA, sixth edit edition, 2005.

[81] Vincent P. Coletta and Jeffrey a. Phillips. Interpreting FCI scores: Normalized gain, preinstruction scores, and scientific reasoning ability. *American Journal of Physics*, 73(12):1172, 2005.

[82] Vincent P. Coletta and Jeffrey a. Phillips. Interpreting force concept inventory scores: Normalized gain and SAT scores. *Physical Review Special Topics - Physics Education Research*, 3(1):010106, May 2007.

[83] RR Hake. Relationship of individual student normalized learning gains in mechanics with gender, high-school physics, and pretest scores on mathematics and spatial. *submitted to the Physics Education Research ...*, pages 1–14, 2002.

[84] Richard Hake. The Impact of Concept Inventories On Physics Education and It âĂŹ s Relevance For Engineering Education. `http://www.physics.indiana.edu/~hake/ImpactCI-D-pp.1-103.pdf`, 2011.

[85] George A Ferguson. On The Theory of Test Discrimination. *Psychometrika*, 14(1):61–68, 1949.

[86] Berend Terliun, Dirk L. Knol, Caroline B. Terwee, and Henrica CW de Vet. Understanding Ferguson's Delta: time to say good-bye? *Health and Quality of Life Outcomes*, 7(38), 2009.

[87] Matthew Hankins. Questionnaire discrimination: (re)-introducing of coefficient delta. *BMC Medical Research Methodology*, 7(19), 2007.

[88] Matthew Hankins. How discriminating are discriminative instruments? *Health and Quality of Life Outcomes*, 6(36), 2008.

[89] Jamis J. Perret, Daniel J. Mundform, and Neil J. Salkind. Bonferroni Procedure. http://knowledge.sagepub.com.virtual.anu.edu.au/view/researchdesign/n33.xml, 2010.

[90] Reinhold Hatzinger. Rasch Models and the R package eRM. http://www.stat.uni-muenchen.de/~carolin/material_psychoco/folien_fuer_homepage/reinhold.pdf, 2010.

[91] Www.rasch.org. Rasch Measurement Analysis Software Directory. http://www.rasch.org/software.htm, 2014.

[92] Mark H. Moulton. Rasch Estimation Demonstration Spreadsheet. http://www.rasch.org/moulton.htm, 2003.

[93] Andrew Ng. Machine Learning (Stanford Coursera MOOC). https://class.coursera.org/ml-005/lecture, 2014.

[94] Lin Ding and Robert Beichner. Approaches to data analysis of multiple-choice questions. *Physical Review Special Topics - Physics Education Research*, 5(2):020103, September 2009.

[95] Lei Bao and Edward Redish. Model analysis: Representing and assessing the dynamics of student learning. *Physical Review Special Topics - Physics Education Research*, 2(1):010103, February 2006.

[96] Gilles Riache, Martin Riopel, and Jean-Guy Blais. Non Graphical Solutions for the Cattell's Scree Test. http://slideplayer.us/slide/246478/, 2006.

[97] HN Boone and DA Boone. Analyzing likert data. *Journal of Extension*, 50(2), 2012.

[98] Susan Jamieson. Likert scales: how to (ab)use them. *Medical education*, 38(12):1217–8, December 2004.

[99] Karen Grace-Martin. Can Likert Scale Data ever be Continuous? http://www.theanalysisfactor.com/can-likert-scale-data-ever-be-continuous/, 2008.

[100] GH Lubke and BO Muthén. Applying multigroup confirmatory factor models for continuous outcomes to Likert scale data complicates meaningful group comparisons. *Structural Equation Modeling*, 11(4):514–534, 2004.

[101] Lorna Jarrett, Brian Ferry, and George Takacs. Development and validation of a concept inventory for introductory-level climate change science. *International Journal of Innovation in Science and Mathematics Education*, 20(2):25–41, 2012.

[102] Robert I. Kabacoff. Principle Components and Factor Analysis. http://www.statmethods.net/advstats/factor.html, 2014.

[103] R Documentation. Principle Component Analysis. http://stat.ethz.ch/R-manual/R-devel/library/stats/html/princomp.html.

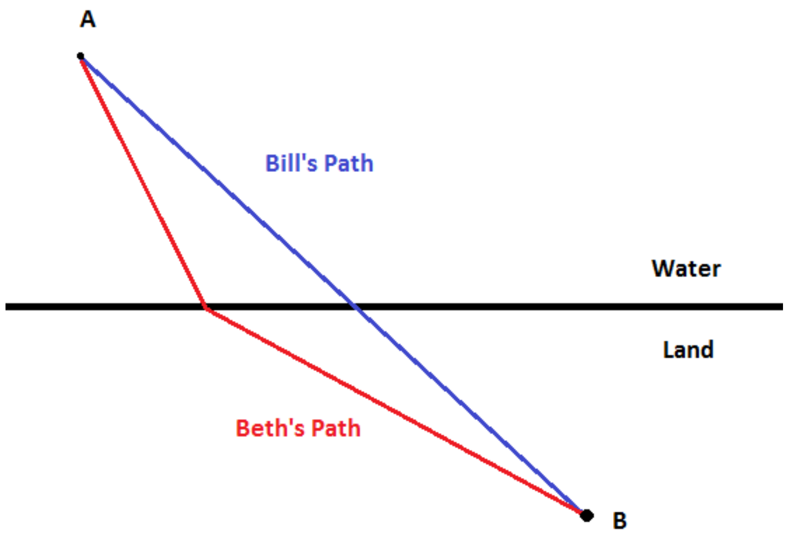
[104] HM Sisti, AL Glass, and TJ Shors. Neurogenesis and the spacing effect: learning over time enhances memory and the survival of new neurons. *Learning & memory*, 14:368–375, 2007.

[105] Will Thalheimer. Spacing Learning Events Over Time : What the Research Says. http://wp.phase-6.com/wp-content/uploads/2013/06/Spacing_Learning_Over_Time__March2009v1_.pdf, 2006.

Appendix 1: Evaluating Question 13

When writing the Action Concept Inventory questions, my favourite question was Question 13. I liked for a number of reasons. Question 13 had named characters and it was an exciting real world application of a variational principle. During their think aloud interviews a few students told me that they enjoyed Q13. In addition it tested the stationary concept in reference to a 'path', rather than a stationary point on a function. I have included the original Question 13 below:

Alice is drowning in a swimming pool at point A. Beth and Bill are both standing on the land at point B.



Both Beth and Bill rush to save Alice. Beth is a trained lifeguard and knows the exact route she should take to reach Alice as quickly as possible. Bill takes a different route.

13. On the way both Beth and Bill stumble once causing them each to arrive at the water at a slightly different location from their original paths. How does this impact on their time to reach Alice? (obviously the stumbling itself would affect their time, but in this question we are only focussing on the change in path)

    (a) It will make no difference to Beth's time, Bill could take longer or shorter.

    (b) Beth will take longer, Bill could take a longer or shorter time. Both their times will be impacted by the same amount.

    (c) Beth will take longer, Bill could take a longer or shorter time. There is not enough information to determine whose time will be affected by more.

    (d) Both times will be affected, but Bill's time will be affected more than Beth's.

    (e) Both times will be affected, but Beth's time will be affected more than Bill's.

Unfortunately multiple sources of evidence indicated that this question was invalid. The pre-test difficulty index was high (0.42), gain was not significantly different from 0 ($g = -0.013$) and it did not correlate strongly with either of the other questions which tested the same concept (question 4 and question 12).

Think aloud interviews showed that students who were unsure about question 13 would frequently guess the correct answer (d), but not be able to explain solid reasoning why. Only one student in the think aloud interviews told me that he was consciously trying to game the test and he did not attempt to game question 13.

I hypothesise that there were subconscious factors leading students to choose the correct answer (d). As mentioned when reviewing the concept inventory literature I discovered Gibb's criteria [70] for preventing students from gaming tests. Of the seven criteria, the only two that could apply to question 13 are 'length' and 'categorical exclusive'. Option (a) was not considered seriously by students in interviews and frequency of selection confirms it was a weak distractor (see Figure 5.7). Of the remaining options (d) and (e) are not only shorter, but don't force the student to commit to the statement 'Beth will take longer, Bill could take a longer or shorter time'. This statement was inserted into options (d) and (e) to address the Gibb's criteria. In addition the options were re-ordered so that 'not enough information' is the last option.

A comment from a student at the end of a think aloud interview indicated that it may have been attributes of the character Beth that was causing question 13 to be invalid. They mentioned that they saw Beth as an impervious lifesaver, a character who will be effected less by a stumble. For this reason this wording was removed. The new version of Question 13 can be found in Appendix 10

Now that Questions 8, 11 and 13 have been revised, student validation interviews will need to be performed. The revised inventory will be delivered to 2015 PHYS1201 class as a pre-test and post-test and the results will be used to validate these questions.

Appendix 2: Action Instruction in PHYS1201

Overall the PHYS1201 action unit comprised of 8 lectures, 3 tutorials, 2 sets of homework and half a lab session (which included the Action concept Inventory post-test). The action unit focussed on giving the first year students the mathematical skills required so they could use the Euler Lagrange equations. Students were also given a conceptual understanding of the principle of least action and a brief introduction to the quantum mechanical concepts which underpin it. The sections below outline each of the four areas of the course; lectures, tutorials, the computational lab and homework.

9.1 Action Lecture Notes

Craig developed his own lecture notes which he used to deliver the action unit to the first year class. He followed the notes quite closely when lecturing the students. The action lecture notes consisted of 12 sections, each is briefly outlined below:

1. **Action Principles: An Overview** - Overviews how action fits into physics as a discipline. Introduction to the concept of ‘path’ and ‘stationary’. Mentions applications of the calculus of variations.

2. **Quantum Physics** - Introduces the Lagrangian and Action. Gives a very brief description of the many paths formulation of quantum mechanics and directs students to QED [28]. Introduces the ‘Born rule’ and ‘complex amplitude’ Does not explicitly introduce the ‘Explore All Paths’ or ‘Superposition Principle’ concepts, but they are implied.

3. **Fundamental Physics** - Introduces symmetries and Noether’s Theorem. Suggests that action will have a key role to play in answering some of the big open questions in cosmology and fundamental physics.

4. **Ray Optics: Fermat’s Principle of Least Time** - Introduces Hero’s Principle of Least Distance and Fermat’s Principle of Least Time. Discusses the ray optics approximation and link’s Fermat’s principle to the least action principle. Discusses applications of Fermat’s principle to lenses and mirrors.

5. **The Principle of Stationary Action** - Introduces Hamilton’s Action Principle. Calculates the action for different paths in both free particle and constant gravitational force examples.

6. **Equivalence to Newtonian Mechanics** - Derives the Euler Lagrange Equation and shows that it is equivalent to Newton's 3rd law if $F = -\partial V/\partial x$. Mentions that Lagrangian mechanics cannot describe friction.

7. **Historical Interlude** - Describes the Historical significance of variational principles and the least action principles.

8. **General Coordinates** - Derives the Euler-Lagrange equations in generalised co-ordinates. Uses the Euler-Lagrange equation to find the equations of motion for 'a bead on a rotating hoop' and 'a spring pendulum' as a worked example.

9. **Symmetries and Conservation Laws** - Introduces cyclic coordinates and canonical momenta. Works through translational symmetry and rotational symmetry to obtain conservation of linear and angular momentum. Links this result back to Noether's theorem.

10. **Relativity** - Introduces proper time and the principle of maximal aging. Shows how this could be equivalent to the least action principle for certain Lagrangians. Resolves the twin paradox using the principle of maximal aging.

11. **General Relativity and Black Holes (non-examinable)** - Introduces the idea of curved space time, path parametrisation and the Schwarzschild metric. Goes on to explore black holes and introduces the event horizon, gravitational time dilation, light cones and Eddington-Finkelstein coordinates.

12. **Action Concepts** - A list of the action concepts that were tested in the Action Concept Inventory.

9.2 Action Lectures

A description of each of the 8 lectures given in the action unit is given below:

1. **Introduction** - Introduced the paths and stationary concepts. Students completed the Action Concept Inventory pre-test. After inventory students were invited to a gold coin donation BBQ and to help me shave my head for cancer.

2. **Fermat's Principle** - Introduced the Quantum mechanical principles, Noether's Theorem and Fermat's principle.

3. Stationary Action Principle - Reviewed Fermat's principle. Introduced Hamilton's Action Principle.

4. **Euler-Lagrange Equation (I took this lecture)** - Derived the Euler-Lagrange equation from the principle of least action.

5. **General Coordinates** - Reviewed Euler-Lagrange equation. Introduced generalised co-ordinates and solved 'a bead on a rotating hoop' as a worked example.

6. **Symmetry** - Continued 'a bead on a rotating hoop' example, discussed historical significance of variational principles and demonstrated the link between continuous symmetries and conserved quantities.

7. **Relativity** - Revisited symmetries and conserved quantities. Tied action to special relativity through the principle of maximal aging and resolved the twin paradox. Introduced general relativity (non-examinable) and showed that the principle of maximal aging is equivalent to the principle of stationary action.

8. **General Relativity** - Explored general relativity in more depth. Finished with a review of the entire action unit.

9.3 Action Tutorials

As a part of the action unit the students completed three tutorials. These tutorials focused on action calculations. The first tutorial was based on Fermat's principle, students were asked to show that mirrors and lenses had stationary time for rays which passed through the focal point. In the second tutorial students were asked to calculate the action for different paths for different masses and then show that a first order perturbation to the stationary paths causes no first order change to the action. In the third tutorial students were asked to use the Euler-Lagrange equations to find the equations of motion for a classical mechanics problem involving constraints (another bead on a hoop).

9.4 Action Computational Lab

The action unit also contained a computer based lab session. Half the laboratory session (1.5 hours) was spent on action exercises and the other half on relativity exercises. The action exercises consisted of two parts. The first part was a series of mathematica simulations, students were asked to find paths of minimum (and maximum) action for both a free particle and a particle in a gravitational potential. The second exercise was to complete the Action Concept Inventory post-test.

9.5 Action Homework

The students were given two sets of homework to complete as a part of the action unit. The first set of homework consisted of:

- Mathematica notebook where students work through 'Einstein's Mirror'
- Two extended response questions (see Sections 9.5.1 and 9.5.2).

The second set of homework consisted of:

- Mathematica notebook where students work through double pendulum.
- One extended response question (see Section 9.5.3).
- Attitude Survey (optional)

The three extended response questions are given for reference in the sections below:

9.5.1 Question 1 - Paths (Local and Global) Paradox

A friend of yours has just read an article on the principle of stationary action. he asks you to consider a ball traveling from event A to event B as shown in the diagram below:

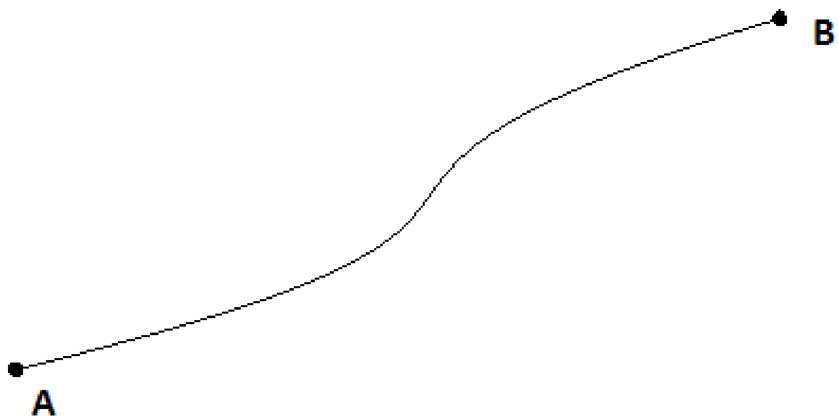


Figure 9.1

He says:
"Using forces that act on the particle at each instant we can make sense of this path. It responds to forcing acting on it at each moment along the path. However with the action approach, we somehow know that the ball ends up at event B and then we can work out the path. How could a particle know in advance that it is going to arrive at event B? Doesn't that violate causality?"

In 200 words or less provide a resolution of this problem.

9.5.2 Question 2 - Glass Block Paradox

One of your study group members has come up with an example that she claims proves Fermat's principle is not always correct for ray optics. Consider points A and B separated by a glass block with extremely high refractive index. See diagram:

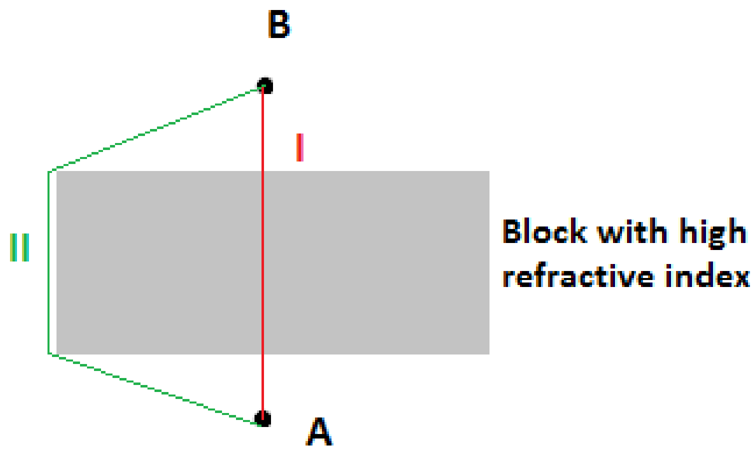


Figure 9.2

Because the block has such a high refractive index, path I takes longer than path II. Hence she claims that path I is not the least time path. Furthermore, she claims that as path II is made up of straight lines, it is a stationary path. However, she did the experiment and the light rays definitely follow path I, and not path II.

Is she right? If so, explain in 200 words or less why Fermat's principle fails in this case. If not, explain in 200 words or less the flaw in her argument.

9.5.3 Question 3 - Explore All Paths Paradox

In part 1 of Action tutorial 2, you were asked to calculate the action for two different paths for both an electron and a 1kg mass.

In that problem the straight-line paths are stationary and have a minimum action out of all paths. As shown in the tutorial solution the difference in action between the two paths is much greater for the 1kg mass than for the electron.

In your own words, explain the significance of this result for the physical behaviour of the 1kg mass and for the electron. The explanation given in the tutorial solution is a starting point, but it does not explain the physical consequences of the differences in action.

9.6 Student Survey

Students were asked to complete a survey as part of the 2nd Action homework. If students completed the survey the would receive 5 marks and their homework would be out of 25 instead of out of 20. The survey questions are contained below for reference:

Do you think action physics should be included in PHYS1201 next year?

Compared to classical mechanics in PHYS1101, rate from 1 to 7 how difficult you found each of the following topics:

- Electromagnetism (PHYS1101)
- Relativity (PHYS1201)
- Action (PHYS1201)

Compared to classical mechanics in PHYS1101, rate from 1 to 7 how interesting you found each of the following topics:

- Electromagnetism (PHYS1101)
- Relativity (PHYS1201)
- Action (PHYS1201)

Compared to classical mechanics in PHYS1101, rate from 1 to 7 how motivated you are to learn about each of the following topics:

- Electromagnetism (PHYS1101)
- Relativity (PHYS1201)
- Action (PHYS1201)

Outside of homework, lectures and tutorials, how many hours did you spend studying each of the following topics each week: (while studying this topic in class)

- Classical Mechanics
- Electromagnetism
- Relativity
- Action

In relation to other parts of PHYS1101 and PHYS1201, how has studying action improved your understanding of physics? (1 - not at all, 4 - about the same, 7 - much more than other topics)

What Aspects of Action Physics did you enjoy learning about the most? (Choose all that apply)

- A new perspective on physics
- The mathematics
- Seeing the link between different areas of physics
- Fundamental physics
- Noether's theorem and symmetry

What did you find most challenging about learning action physics? (Choose all that apply)

- Partial Derivatives
- Calculus of Variations and other mathematics
- Quantum Mechanical Concepts
- Using Mathematica

What changes do you think should be made to the action unit next year?

If you have any additional comments or feedback you can leave them here.

Chapter 10

Appendix 3: The Revised Action Concept Inventory

Instructions:

- *Each question has an additional confidence scale similar to the example below. For each question, mark on the scale how confident you are in your choice.*

Rate how confident you are in your answer:

◯ ⋯⋯ ◯ ⋯⋯ ◯ ⋯⋯ ◯ ⋯⋯ ◯ ⋯⋯ ◯ ⋯⋯ ◯

guessing unconfident neutral confident certain

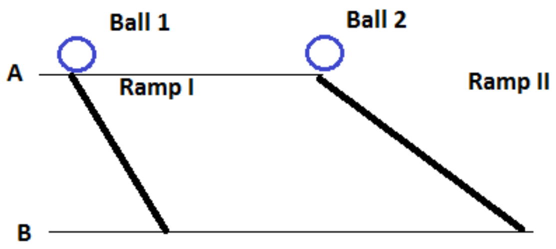
- *Answer all of the questions to the best of your knowledge.*

- *The quantity action is referred to frequently in this inventory. In this inventory action refers to Hamilton's Action and is defined as:* $S = \int_A^B (T - V)dt$ *where* S *is the action,* A *and* B *are the start and end events,* T *is the kinetic energy,* V *is the potential energy and* t *is time.*

- *This test commonly refers to 'arbitrarily small', 'extremely small' or 'very slight' displacements or movements. When this is stated you should assume that* $x^2 << x$*. Note arbitrarily small is distinctly different from infinitely small. An example which highlights the difference: if you give me an even number it can be arbitrary large but it is still finite (it is not infinitely large).*

If you have any questions about the study, please feel free to contact Lachlan McGinness (u4849410@anu.edu.au or 6125 8896), or Craig Savage (craig.savage@anu.edu.au or 6125 4202). If you have any concerns about the way the research is being done, please contact the Secretary of the Human Research Ethics Committee, Research Office, Chancelry 10B, ANU (human.ethics.officer@anu.edu.au or 6125 3427).

Two identical balls are originally stationary at the top of two frictionless ramps. Ramp I is twice as steep as ramp II. Both ramps start at height A and end at height B. When the balls are released at the same time, the ball on ramp I reaches the bottom first.



1. Compare the speed of the two balls when they reach the bottom of the ramp. (be as accurate in your comparison as possible)
   (a) Same speed
   (b) Ball 1 will be going twice as fast as ball 2
   (c) Ball 1 will slightly faster than ball 2
   (d) The speed of ball 2 will be greater than ball 1

2. Sam places a few perfectly elastic (no sound or heat is produced in collisions) balls in a can, closes it, and shakes them for a few minutes. While Sam is shaking the can, for which of the following is energy conserved?
   (a) The energy of each individual ball is conserved.
   (b) The total energy of all the balls is conserved.
   (c) The total energy of all the balls and the can is conserved.
   (d) None of the above.
   (e) Both a and b
   (f) both b and c

3. Cobar has found a point on his chair where a marble will sit at rest (assume there is no friction). The gravitational potential of each point on the chair is proportional to its height. Which of the following **must be** true? (choose all that apply)

I If the marble is moved very slightly to either side its potential energy will not change

II The marble is in a local minima or a local maxima of potential energy

III There is a small area around the marble where the chair is flat

IV The gradient of the potential at the point is zero

(a) I only

(b) IV only

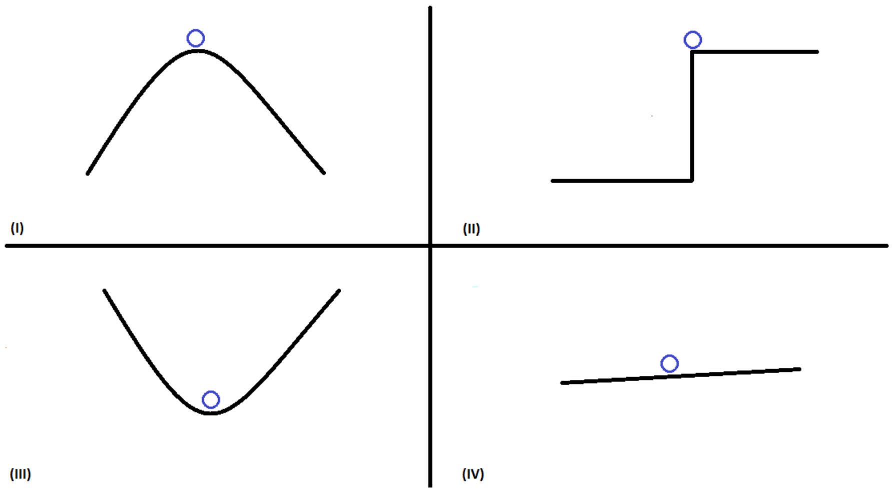
(c) I and IV

(d) I, III and IV

(e) I, II and IV

(f) I, II, III and IV

Identical balls are placed on several different idealized slopes. The gravitational potential energy is proportional to the ball's height. When the balls are originally placed on the slopes they are still. See diagram below:



4. Each ball is displaced an extremely small amount to the left. Order from greatest to least the change in magnitude of the height of the balls immediately after the displacement, **before the ball begins to roll**.

   (a) II > I > III > IV

   (b) II > III = I > IV

   (c) II > I > IV > III

   (d) II > IV > I = III

5. Sharon has constructed an experiment involving a potential. If she places an object exactly at rest at a certain point it will stay there. Which of the above shapes could describe the potential Sharon has constructed? (note the point does not necessarily need to be the location illustrated by the ball)

   (a) I only

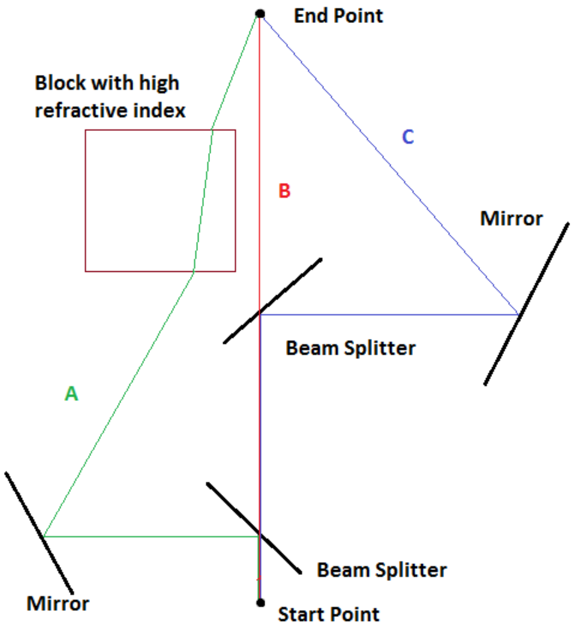
   (b) II only

   (c) III only

   (d) I and III

   (e) I, II and III

A light ray is split by beam-splitters and takes 3 different paths when moving from start to end point.



6. Which of these paths has stationary time compared to neighbouring paths?

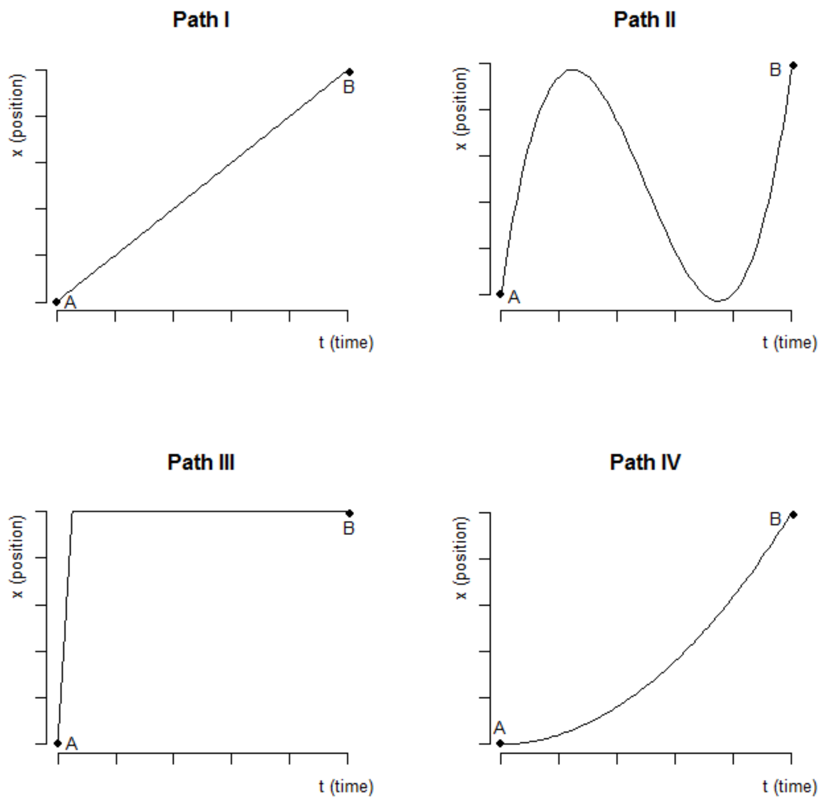
   (a) Only A
   (b) Only B
   (c) A and B
   (d) A and C
   (e) A, B and C

A particle is in a potential and starts at position x. There are infinitely many paths it could take to reach position y. For only three of these paths the action is stationary with respect to neighbouring paths. One is the shortest path, the second has a minimum action out of all paths, the third is neither.

7. Which of these paths make a significant contribution to the overall probability of the particle being detected at y? (choose all that apply)

   (a) I only
   (b) II only
   (c) I and II
   (d) I, II and III

An electron is measured at A and then measured at B some time later. Some paths are illustrated in the diagram below:



8. **Look carefully at the axes of the plots.** If there are no forces acting on the electron, which path did the electron take between the measurements at A and B? (Choose all that apply)

   (a) The electron took I only
   (b) The electron took all the paths
   (c) The electron could have taken any of the paths
   (d) The electron took the stationary path or a path near it.

9. If there is a constant (non-zero) force acting on the electron, which paths have stationary Action? (Choose all that apply)

   (a) I only
   (b) III only
   (c) IV only
   (d) I and IV
   (e) I, II, III and IV

10. Sharon is designing a trajectory for a space probe from Earth to Mars. Since the potential around Earth and Mars is well known she uses the stationary action principle. In order to find the path of stationary action she must specify:

I	The launching position	IV	The arrival position
II	The launching velocity	V	The arrival velocity
III	The launching time	VI	The arrival time

(a) I and IV

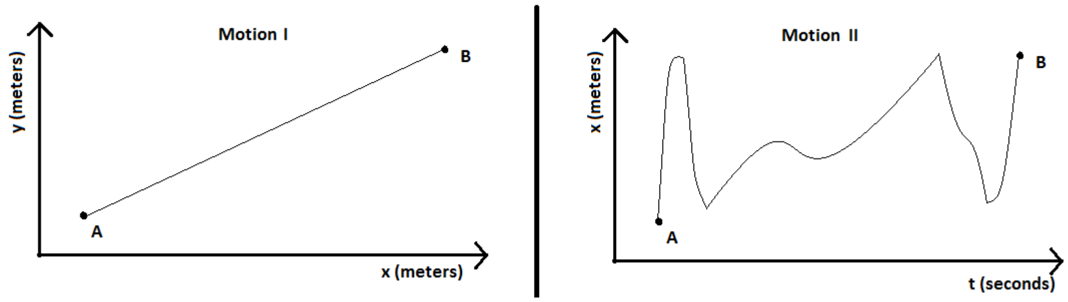
(b) I, II, IV and V

(c) I, III, IV, VI

(d) All of the above

(e) All of the above except V

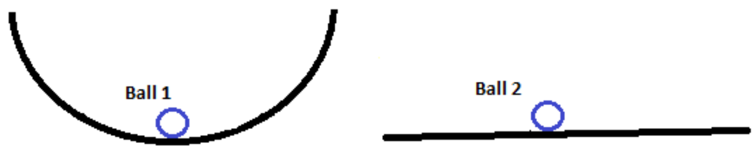
Two objects are placed in known potentials. Two different motions are illustrated diagram below.



11. Notice that motion I is a trajectory in two spatial dimensions but the timing of the trajectory is not specified. Motion II has one spatial dimension and is parametrised by time. For which of the motions is it possible to calculate the action?

    (a) The Action can be calculated only for motion I.
    (b) The Action can be calculated only for motion II.
    (c) Action can be calculated for both motions.
    (d) Action cannot be calculated for either motion.

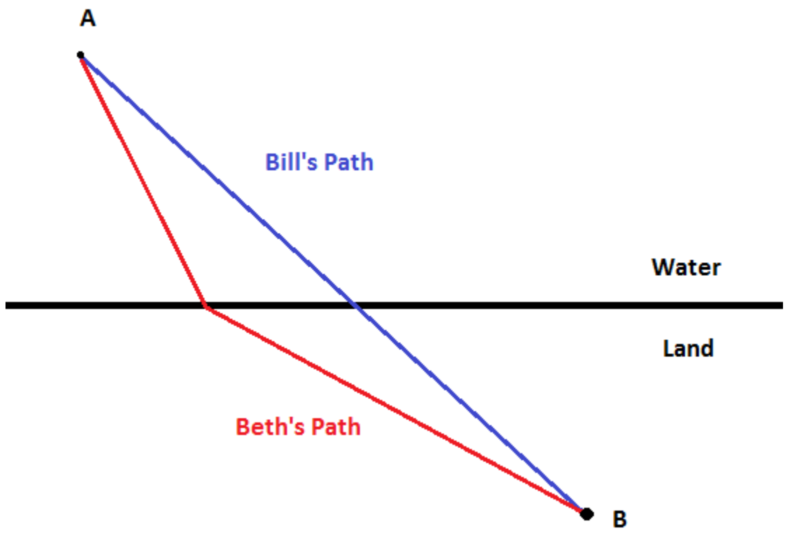
There are two balls that are still. One is at the bottom of a spherical bowl, the other is on an almost flat ramp. See diagram below:



12. Both the balls are moved by an extremely small amount to the right. Compare as accurately as possible how the heights of the two balls change.

    (a) There is no change in the height of ball 1, but there is a small change for ball 2
    (b) Both balls change height, but ball 2 changes by more than ball 1
    (c) Ball 1 will change height by more than ball 2
    (d) Depending on how small the displacement is, either a or b

Alice is drowning in a swimming pool at point A. Beth and Bill are both standing on the land at point B.



Both Beth and Bill rush to save Alice. Beth takes the quickest route to reach Alice and Bill takes the shortest route.

13. On the way both Beth and Bill stumble once causing them each to arrive at the water at a slightly different location from their original paths. How does this impact on their time to reach Alice? (obviously the stumbling itself would affect their time, but in this question we are only focusing on the change in path)

    (a) Both Beth and Bill could take longer or shorter. There is not enough information to determine whose time will be affected by more.

    (b) Beth will take longer, Bill could take a longer or shorter time. Both their times will be impacted by the same amount.

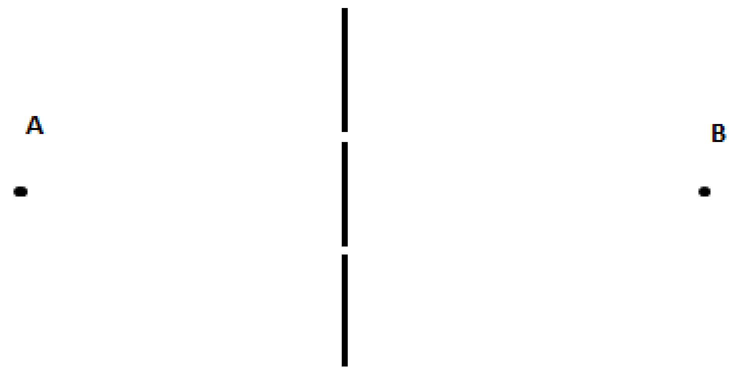
    (c) Beth will take longer, Bill could take a longer or shorter time. Both times will be affected, but Bill's time will be affected more than Beth's.

    (d) Beth will take longer, Bill could take a longer or shorter time. Both times will be affected, but Beth's time will be affected more than Bill's.

    (e) It will make no difference to Beth's time, Bill could take longer or shorter.

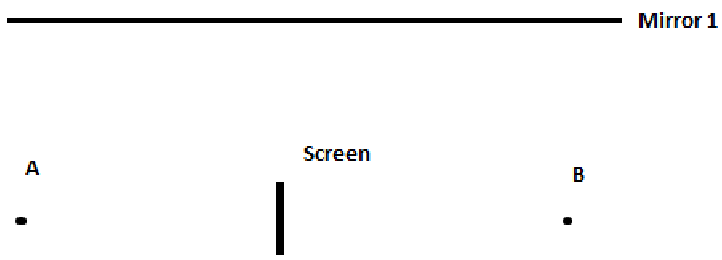
Questions 14 to 18 refer to an idealised photon source and detector. The ideal photon source is infinitely small and releases photons at a continuous rate with no preferred direction. Ideal photon detectors measure the presence of a photon within an arbitrarily small region of space. For the purposes of the questions in this inventory you may always treat the size of the source and detector as small compared to all other objects and length scales in the problem.

Photons (light particles) travel from a source at A to a detector at B. There is a blockout screen between A and B. Two narrow slits are placed in the screen both equally distant from A and B as shown in the diagram below:



14. Originally one slit is open and the other is closed. When the second slit is opened how does the probability of detecting a photon at B change?
    (a) The probability increases but not necessary double or fourfold
    (b) The probability doubles
    (c) The probability increases by a factor of 4
    (d) The probability is zero both before and after the second slit is opened
    (e) The probability is non-zero and it will remain the same

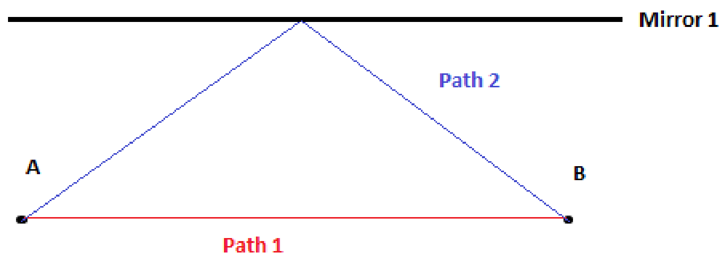
Photons (light particles) travel from a source at A to a detector at B. There is a blockout screen between A and B and a mirror placed above the screen. See the diagram below:



15. The mirror remains in the same position but the screen is removed. How does this affect the probability of a photon being detected at B?

    (a) The probability increases but not necessarily by double
    (b) The probability doubles
    (c) The probability stays the same
    (d) The probability could increase or decrease

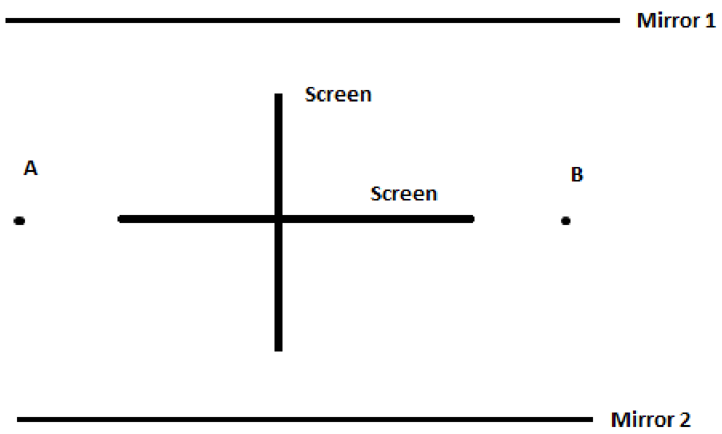
Two paths are shown in the diagram below.



16. Which of the paths have stationary action? (choose all that apply)

    (a) Path 1
    (b) Path 2
    (c) Both path 1 and path 2
    (d) Neither

A second mirror is added in such a way that it is **exactly the same distance** from A and B as the first mirror. A second screen is added to prevent light bouncing between the two mirrors. See the diagram below:



17. For this question ignore diffraction effects from the screens. How is the probability of detection changed by the addition of mirror 2? (compared to the case where there is mirror 1 and both screens)
    (a) The probability remains the same
    (b) The probability doubles
    (c) Probability increases by a factor of four
    (d) The probability increases but not necessary double or fourfold

18. Mirror 2 is gradually moved down. How does the probability of detection at B change as the mirror is moved?
    (a) Probability decreases slightly
    (b) Probability remains the same
    (c) The probability will fluctuate
    (d) The probability decreases gradually, approaching half the original probability as the mirror gets very far away
    (e) The probability will increase.