A cognitive model of axiom formulation and reformulation with application to AI and software engineering

Andy Clark, Simon Colton, Paul Crook,
Andrew Ireland, Alison Pease, and Alan Smaill

I. Previous track record

Alan Smaill, Andrew Ireland, Simon Colton and Alison Pease are members of the Mathematical Reasoning Group (MRG). Since the mid 1970s, the MRG has been engaged on the computational analysis, development and application of mathematical reasoning processes and their interactions. Its work is characterised by its unique blend of computational theory with artificial intelligence (http://dream.dai.ed.ac.uk/). Since the mid 1980s it has produced over 300 publications, trained 17 research fellows and 41 PhD students, two of whom have won the BCS/CPHC Distinguished Dissertations Award. It has been supported by over 40 research grants, including an EPSRC rolling funding grant (1982-2002), and it has just been awarded an EPSRC platform grant.

Andrew Ireland is a founding member of the Dependable Systems Group (DSG) at Heriot-Watt University, whose research focus is to improve the reliability and predictability of computer systems through the development and application of rigorous design, implementation and verification techniques. Current DSG research areas include: theorem proving, formal verification and synthesis; design of parallel and distributed functional languages; performance modelling and simulation of parallel and distributed systems. Since its inception in 1997, the DSG has grown to 2 Professors, 3 Senior Lecturers, 2 Lecturers and 5 Research Assistants and 19 PhD students. Group members currently hold 4 research grants.

Simon Colton has founded the Combined Reasoning Group at Imperial (www.doc.ic.ac.uk/crg), which currently consists of Dr. Colton and his five PhD students. Imperial College is a world-renowned institute, and the Department achieved a 5* rating in both previous research assessment exercises (RAE) with strengths across computing, and in Artificial Intelligence in particular.

Paul Crook is a member of the Institute of Perception, Action and Behaviour (IPAB), which researches how to link, in theory and in practice, computational perception, representation, transformation and generation processes to external worlds. Alan Smaill and Alison Pease are members of the Centre for Intelligent Systems and their Applications (CISA), which investigates how knowledge can be formally represented and reasoning can be automated. IPAB and CISA are in the School of Informatics at the University of Edinburgh. This school was one of only six computing departments in UK to have obtained a 5* ranking in the 2001 RAE. It returned the highest number of research active staff and was the only 5*A department. It contains world-class research groups in the areas of theoretical computer science, artificial intelligence and cognitive science.

Andy Clark is a member of the department of philosophy at the University of Edinburgh. Philosophy has been taught at the university since its foundation in 1583. The department was one of only six philosophy departments in the UK to obtain a 5* in the last RAE.

1 Dr. Alan Smaill

Dr. Alan Smaill has worked in the area of Automated Reasoning since 1986. He holds a D. Phil. from the University of Oxford in Mathematical Logic, and currently teaches in the School of Informatics in the University of Edinburgh. His research work centres around reasoning in higher-order and constructive logics, with applications in program construction and automated software engineering, and he has published widely in this area. He has been a grant-holder on an ESPRIT project on Logical Frameworks and a PI on EPSRC grant GR/M46624, on the mechanisation of first-order temporal logic via embedding into a higher-order representation.

2 Professor Andy Clark

Professor Andy Clark is Chair of Logic and Metaphysics in the School of Philosophy, Psychology and Language Sciences, at Edinburgh University. Having received his PhD from the University of Stirling in 1984, he has been Professor of Philosophy and Cognitive Science at the University of Sussex, UK, Professor of Philosophy and Director of the Philosophy/Neuroscience/Psychology Program at Washington University in St. Louis, Missouri, USA. and Professor of Philosophy and Director of the Cognitive Science Program at Indiana University, Bloomington, USA. He is the author of five books including Being There: Putting Brain, Body And World Together Again (MIT Press, 1997) and Natural-Born Cyborgs: Minds, Technologies And The Future Of Human Intelligence (Oxford University Press, 2003) as well as numerous papers and four edited volumes. Current research interests include robotics and artificial life, the cognitive role of human-built structures, specialization and interactive dynamics in neural systems, and the interplay between language, thought and action. He is currently P.I. (Edinburgh group) on a 3-year (2006-2009) project funded by AHRC as part of the ESF (European Science Foundation) initiative Eurocores: CNCCC (Consciousness In A Natural And Cultural Context). Project Title: Consciousness in Interaction: The Role of the Physical and Social Environment in Shaping Consciousness.

3 Dr. Andrew Ireland

Dr. Andrew Ireland obtained his PhD in Computing Science from the University of Stirling in 1989 (Carnegie Scholarship). After a period as a teaching assistant within the then Department of Artificial Intelligence (Edinburgh), he joined the Mathematical Reasoning Group (MRG) in 1990 as Research Associate. Within the MRG he worked on techniques for automating the search for formal proofs. In 1995, he became a lecturer in Computer Science at Heriot-Watt University, and was promoted to senior lecturer in 2005. His research focuses upon the role of formal proof within the development of dependable software systems. The applied nature of his work has led him into collaborative research projects with Praxis High Integrity Systems Ltd and QinetiQ. He has substantial research experience in the area of automated reasoning and automated software engineering on which he has published widely. Since joining Heriot-Watt he has been PI on 3 grants and CI on 5 grants. He has been involved in organising many national and international workshops and conferences. He is a member of the Steering Committee for the IEEE/ACM International Conference Series on Automated Software Engineering (ASE), and will be Programme Co-Chair for ASE-08. He has been a member of the EPSRC Peer Review College since 2003.

4 Dr. Simon Colton

Dr. Simon Colton is a lecturer and PhD. Admissions tutor in the Department of Computing at Imperial College, London. From October 2007, he will be a senior lecturer. He has published more than 70 papers on broad ranging topics in Artificial Intelligence, including automated theorem proving, machine learning, constraint solving, multi-agent systems and evolutionary approaches. His research has involved the introduction of a novel descriptive machine learning technique – automated theory formation – and the application of this to machine discovery projects in pure mathematics and bioinformatics. He has also been at the fore-front of the effort to combine AI systems so that the whole is more than a sum of the parts. His work has been recognised both nationally, with a BCS/CPHC distinguished dissertation award, and internationally, with a best paper award at AAAI 2000. His EPSRC-funded project post-Doc project was assessed as internationally leading. He has been/is a grant holder on 6 EPSRC funded grants. He was a member of the AISB committee (2001 - 2007), is on the Automated Reasoning Workshop committee, and serves on the steering committee for the International Joint Workshop on Computational Creativity. He has served on numerous programme committees for international conferences, and is currently on the AAAI, MKM, CADE, LPAR, MICAI and Discovery Science committees.

5 Dr. Paul Crook

Dr. Paul Crook obtained a PhD in Artificial Intelligence from the University of Edinburgh in 2007. He is currently employed as a Research Associate in the Institute of Adaptive and Neural Computation at the University of Edinburgh researching the application of machine learning to labelling animal behaviours. Prior to obtaining his PhD Paul worked as a project manager in the electricity industry where he obtained the professional qualifications of Chartered Engineer and Member of the Institution of Engineering and Technology (formally the IEE). He has also obtained an MSc with distinction in Artificial Intelligence at Edinburgh and a 1st class honours degree in Electrical Engineering from the University of Bath. His MSc and PhD were undertaken in the Institute of Perception Action and Behaviour where he gained considerable experience in the application of machine learning in embodied agents - his PhD looked at the learning of good policies by embodied agents whose environments were partially observable. The research he undertook in IPAB involved work with both simulated and real-world robots. He has presented his PhD work at a number of international conferences including European Conference on Machine Learning and IEEE International Conference on Robotics and Automation.

6 Dr. Alison Pease

Dr. Alison Pease obtained her PhD in 2007 after completing an MA in Mathematics/Philosophy from the University of Aberdeen and an MSc in Artificial Intelligence from the University of Edinburgh. Her PhD thesis was entitled *A Computational Model of Lakatos-style Reasoning* and described her implementation and evaluation of the theory of mathematical progress described in Lakatos’s *Proofs and Refutations*. She has authored various publications, and was an invited speaker to the European Conference on Computing and Philosophy in 2004. She was joint programme chair for the second and third Joint International Workshops on Computational Creativity, in 2005 and 2006, and has served on the programme committee for the same workshop series for the last four years and for the “AI and Creativity in Arts and Science” series in 2002 and 2003. Prior to obtaining her PhD Alison worked as a mathematics teacher for four years, teaching levels from Special Needs to A level students. She holds a PGCE specialising in teaching mathematics.

7 Suitability of the team

We have a considerable amount of relevant expertise. Alan Smaill, Simon Colton and Alison Pease, with John Lee, have already produced a computational representation and subsequent evaluation of Lakatos’s ideas and we anticipate that this model will form the initial basis of our model of axiomatisation. Simon Colton is an expert in automated theory formation, machine learning and CSP and, with Alison Pease, has carried out initial work on how automatically changing axioms may be of use to the theorem proving community. Simon has also worked with Alison on issues of Computational Creativity, in particular how to assign and assess notions of creativity in AI software. They have jointly written 11 papers, and in 2006 they co-chaired a workshop on computational creativity at ECAI. Additionally, Alison has both practical and theoretical experience in teaching mathematics, which we anticipate will be useful for the application to education. Paul Crook has seven years research experience in learning in embodied situated agents, in both simulated and real world settings. His work has looked at models of attention and learning to direct attention in order to complete tasks in partially observable environments. Andy Clark is world-renowned in embodied cognition. Andrew Ireland is an expert on automated theorem proving, software verification, and empirically successful automated reasoning and has strong industrial links which we intend to exploit.

II. Case for support

8 Motivation and overview

The building and transformation of mathematical theories has been pervasive through the sciences, engineering and medicine. We aim to provide a computational account of cognitively plausible mechanisms underlying these operations of theory formation and evolution. The axiomatic method has been a fundamental aspect of mathematical research since Euclid, and various axiom changes have led to revolutions in mathematics. For instance, relaxing the axioms defining numbers led to negative and imaginary numbers, and rejecting the parallel postulate opened up fascinating new areas of non-Euclidean geometry. We will build on our earlier work which provided a computational model of theory evolution based on the ideas of Lakatos [16; 17; 18]. It is perhaps surprising that mathematical thought has received relatively little attention from the viewpoint of cognitive science. This is lamented by Lakoff and Núñez, who claim that (prior to their work), “there was still no discipline of mathematical idea analysis from a cognitive perspective” [10]. The ideas from [10] will inform the computational theory that we will build.

In summary, we plan to make contributions to the computational philosophy of science, cognitive science and computer science by:

- developing a new theory of mathematical axiom formation and modification, and a computational model which implements a new approach to automating pre-axiomatic theory formation;
- applying the computational model to the development of background theories in problem solving in AI and software engineering.

9 A cognitive theory of mathematical axiomatisation

Lakoff and Núñez [10] and Lakatos [9] present cognitive and philosophical accounts of the origin and development of mathematical ideas. Both argue strongly against the “romantic” [10] or “deductivist” style [9] in which mathematics is presented as an ever-increasing set of universal, absolute, certain truths which exist independently of humans.

Lakoff and Núñez present the thesis that the human embodied mind brings mathematics into being. That is, human mathematics is grounded in bodily experience of a physical world, and mathematical entities inherit properties which objects in the world have, such as being consistent or stable over time. Exploring the physical world of object collection might lead to concepts like the empty collection and rules like “adding a collection of n objects to an empty collection yields a collection with n objects”. We then form grounding metaphors between the physical world and an abstract mathematical world, allowing us to project from everyday experiences onto abstract concepts, thus leading to the concept of zero and the axiom that $n + 0 = n$. Lakoff and Núñez posit that blending different mathematical metaphors leads to more complex ideas. We intend to strongly draw from these ideas to produce a computational model of grounded axiomatisation. Grounding a system of mathematics via embodied interaction with an environment relates to the symbol grounding problem and will enable us to provide an account of how concepts and axioms get their meanings, and what these meanings might be. The educational potential inherent in a better understanding of how we build and use representations of mathematical ideas is also a strong motivation.

Lakatos [9] emphasised the social nature of mathematical progress, describing how different mathematicians may have different interpretations of a conjecture, examples or counterexamples of it, and beliefs regarding its value or theoremhood. Through discussion triggered by counterexamples, concepts are refined whist conjectures and proofs are modified. Lakatos categorised these processes into various methods, which we believe can be extended to explain the evolution of axioms. For instance, “monster-barring” is when one reacts to a counterexample to a conjecture by barring it in a principled way rather than rejecting or refining the conjecture. This is illustrated with respect to Euler’s conjecture that for any polyhedra, the number of vertices minus the number of edges plus the number of faces ($V - E + F$) is equal to two, and the ‘counterexample’ of a hollow cube (a cube with a cube-shaped hole in it), for which $V - E + F = 16 - 24 + 12 = 4$. The monster-barring argument is that the hollow cube is not a valid polyhedron. The focus then turns to the meaning of the terms in a conjecture, rather than the truth of it, and a definition is sought which prevents the construction of the hollow cube.

The monster-barring method can explain changes in the axiomatisation of set theory. In his axiomatisation of arithmetic, Frege suggested a definition of number based on Cantor’s naïve set theory. This included the *comprehension principle*, which is the axiom that given any meaningful property P, it is possible to form the set of all sets which satisfy P. Cantor’s theorem holds that for any set, the size of that set is less than the size of its power set, *i.e.*, $o(S) < o(\pi(S))$. However, the set of all sets, s, is a counterexample to this theorem, since $o(\pi(s))$ is a proper subset of s and we know that the size of a subset is less than or equal to the size of the set of which it is a part, *i.e.*, $o(\pi(s)) \leq o(s)$. Rather than using the counterexample to refute Cantor’s theorem, the set of all sets was rejected as a monster (Russell’s paradox was another famous case in which this set was problematic), and the focus turned to finding a definition of set which prevented the construction of such a set. This resulted in a modification of the comprehension principle, which became known in Zermelo-Fraenkel set theory as the *axiom of subsets*: given the set S, and any meaningful property P, it is possible to form the set of all members of S which satisfy P. That is, in order to construct a new set, an *existing set* as well as a property is necessary. We have already developed a computational model of Lakatos’s theory and used our model to evaluate his theory. A principal contribution of our work was to show the generality of Lakatos’s theory, thus addressing a criticism often levelled at Lakatos that since his theory derived from only two case studies it was of limited generality and therefore limited interest. We expect to further extend the domain of application to pre-axiomatic theories.

There is a wealth of further literature on axiom formulation and reformulation. Boden [2] outlines the idea of a concept space which is mapped, explored and transformed by transcending mapped boundaries, where heuristics for transformation, such as *consider the negative* and *drop a constraint*, are suggested. Reverse mathematics is a programme in mathematical logic which seeks to determine which axioms are required to prove mathematical theorems. McCune [12] carried out work into automatically finding simple single axioms for groups and Abelian groups, in which he used the automated theorem-proving program OTTER to construct sets of candidate single axioms and to prove that these candidates were in fact single axioms. In the philosophy of science, the questions of what a law is, how science progresses by inventing subsidiary hypotheses when a law appears to have been falsified, and when laws might be rejected, have been well studied. In general, a set of axioms might be modified when: counterexamples have been discovered, for which there is no obvious repair to other hypotheses in the system; the model or search space has been fully explored and further models are considered dull; there are hidden assumptions in the axioms, which should be made explicit; we want to be able to prove a specific statement which we are currently unable to prove; or the axioms allow us to prove something which we want to be unprovable.

This work has largely been overlooked by AI researchers. By extracting aspects relevant to axiomatisation, we will formulate a theory consisting of two streams of research: embodied axiomatisation based on [10] and societal axiomatisation based on [9], and we will look at the interaction between these competing pressures. For each stream, we will identify:

(a) situations in which axiom formation/modification is necessitated;

(b) ways in which axiom formation/modification is effected,

(c) ways in which the effects of axiom formation/ modification can be assessed.

10 Automating pre-axiomatic mathematical theory formation

Automated theory formation
Despite forty years of research into automating the formation of mathematical theories, there is still no automated theory formation (ATF) system which works at the pre-axiomatic stage or takes cognitively plausible knowledge as input. In one of the earliest approaches, Lenat [11] claimed that his AM model could progress from prenumerical knowledge to that of an undergraduate mathematics student. One pivotal point in AM's theory formation was the invention of the concept of natural numbers, via the concept of "canonical bags", given only set-theoretic concepts as input. However, this has been criticised in [19; 4] as being too fine-tuned, since AM used an explicit production rule which specialised the concept of bags from sets. Hence, Lenat's work did not generalise to other domains and he failed to shed light on the processes by which humans form axioms. This has left the field of pre-axiomatic theory formation in a state of confusion and subsequent ATF systems have not attempted to address the issue. Lenat also confused the notions of concept, production rules and heuristic measures. One of our main contributions in [4] was to clarify and generalise these notions via the HR theory formation program. HR started with far fewer concepts (typically 6 or 7 as opposed to 115), had far fewer production rules (12 as opposed to 242) and was able to work in far more domains, including group theory, graph theory, number theory, finite algebras, Zariski Spaces, mutagenesis and puzzles. We would now like to further generalise ATF by clarifying ideas on axiom formulation and reformulation.

The interface between ATF and situated embodied agents
An important part of automating pre-axiomatic systems is determining where the axioms come from. Recent work has proposed approaches which can build up concepts and rules about the world based on experience gained from interacting with a stochastic domain [15; 21]. This work shows that the gap between systems which reason about the world and systems which interact with it is closing. Both approaches have been criticised on the grounds that they cannot both deal with a complex world *and* perform complex tasks. For instance, although Brooks' subsumption architecture framework has proven itself in allowing the creation of reactive robots that can deal with the natural complexity of the real world, the architecture has proved somewhat limited in the complexity of the tasks to which it can be applied. To overcome this and allow embodied agents to undertake more complex tasks, there has been a return to the older sense-model-plan-act approach, with the difference that the robustness to the natural world is built in at the modelling level through the use of powerful statistical techniques [23]. In [20] we successfully used our theory formation system HR to learn simple rules of a dice game from noisy data supplied from a vision system, and in [6] we showed that the HR system could invent black and white squares from only the concepts of grid coordinates and dividing two integers, which is a necessary concept in the solution of the mutilated checkerboard problem. Being able to reason at a high level about these rules and concepts would be a powerful tool for an embodied agent learning about its environment, especially if such reasoning resulted in testable hypotheses that the embodied agent could try out in its world.

10.1 Example case study in arithmetic

A computational model of Lakoff and Núñez's work must comprise both an embodied level where axioms can be seen as hidden rules which hold for, or are inspired by, a physical world, and an abstract level where these rules are explored and sometimes changed. Our abstract level will be based on Lakatos's theory of mathematical progress. We envisage a model in which the interaction between the embodied agent and the reasoning software would work in a simple arithmetic domain in the following way. An embodied agent with innate arithmetic capabilities is able to interact with its simulated environment, to move objects around into different piles and configurations, and to count the results. The agent may store the results of adding a first pile to a second pile, and the results of adding the second pile to the first. These would then be passed as input to the theory formation system (stage one). This system would abstract and generalise rules, or axioms, which are descriptive of the patterns it finds. For example, it might generate the *commutative axiom of addition* (for real numbers a and b, $a + b = b + a$). The system would then explore the search space which the axioms define, by generating further concepts, making conjectures empirically, such as *whenever we subtract 1 from a number then we get another number*, and *all numbers can be written as the sum of two numbers*, and passing these to a theorem prover. An important part of this stage would also be a consistency and independence check (stage two). Conjectures and theorems would then be passed back to the embodied agent for evaluation. For instance, the agent might evaluate relevance by testing whether a theorem can be instantiated within the world, or interestingness in terms of whether the theorem provides a new description of known behaviour or describes previously unknown behaviour. The agent might note that the two conjectures above hold for every collection of objects except for the collection of one object. It might then extend its concept of collection to including the empty collection, by performing the operation of removing one object from a collection of one object and labelling the result a collection (stage 3). Finally, the same theory formation program would be used to analyse the information about the theorems and axioms and used to modify the axiom set. If one axiom had only been used to generate uninteresting theorems then this may be rejected at this stage. Conversely, for instance, having the “number” zero in the system might suggest further conjectures which would justify its inclusion in the theory. If any of the theorems contradicted each other then the axioms used would need to be modified or rejected (stage 4). We would evaluate our model based on whether it could reinvent concepts such as zero or axioms in a cognitively plausible way, and whether it recognised the interestingness of such pivotal concepts. Other inspiring examples include reinventing Euclidean and non-Euclidean axioms in geometry, and set and group theoretic axioms.

Our implementation of pre-axiomatic theory formation will provide a general purpose tool which, starting from cognitively plausible innate abilities, can explore a simulated world and formulate, explore, evaluate and modify axioms which describe that world. It will be able to:

- interact with a simulated environment

- take results of this interaction as input knowledge into its theory formation system

- reason about these results, forming concepts and conjectures about its world

- modify and develop these concepts and conjectures via counterexamples, using Lakatos-style reasoning

- use the simulated environment to evaluate the modifications

10.2 Evaluating pre-axiomatic mathematical theory formation

Evaluating cognitive science models is a research question in its own right, which we explored in our work on modelling Lakatos’s theory. We will evaluate:

(i) our *computational model*, by (a) showing that the structure of the model reflects our theory; (b) demonstrating that it runs successfully on inspiring examples such as the simple arithmetic example above *i.e.*, we have a detailed, machine tested mechanical inferential route from pre-axiomatic state to arithmetic and other domains;

(ii) our *theory*, by running the model to (a) test hypotheses such as *Lakatos’s theory of monster-barring can be used to find and repair problems in the axiomatisation of set theory*; (b) determine those conditions under which it succeeds or fails to recreate inspiring and historical examples; (c) evaluate our extended general theory.

(iii) the *applications* of our model, described in the following section, by evaluating the resulting techniques: whether they are general, efficient and scale up to real world problems.

11 Applications to computer science

In order to develop a *general* theory of axiomatisation, we need to develop and test our theory in more than one domain. Mathematics is an obvious development domain, since two of the most important theories of axiom formulation and reformulation [10; 9] are in mathematics. To ensure the generality of our theory we will also develop it in two further domains: *(i)* software specifications requirements and *(ii)* AI problem solving domains. These will be explored by two PhD students, whose motivation and main tasks are described in §’s 11.1 and 11.2. The students will address the following twin questions:

1. Can the theory of mathematical axiomatisation inform the way in which background knowledge can be revised by software specifications and AI problem solving systems?

2. Can a general theory of background knowledge revision in AI problem solving and software specifications inform the theory of mathematical axiomatisation that is being developed for the main project?

11.1 A theory of software specifications modification

Finding ways of effectively handling the volatile nature of specifications is a key area in the sixth grand challenge identified by the UK Computing Research Committee in 2004: “Dependable Systems Evolution”. The level of complexity of modern systems, the fact that multiple development participants with different objectives, views, areas of concern, or languages may be involved in writing specifications, that specifications inevitably evolve, and that humans tend to be very bad at specifying exactly what they want, all contribute to the challenge. While some approaches tolerate inconsistency [14], and frameworks for representing and handling inconsistency have been developed [7; 13], very little work thus far has been carried out on automatically changing requirements specifications. Hoare [8] argues that one important, and frequently overlooked, characteristic of research into dependable systems is to maintain links with humans: the way we do things and the problems that we have. Modelling specification modification on theories which are based on the way in which humans think is one way of maintaining such a link. This should make interaction easier – full automation may be desirable as a long term research aim, but an interactive specification modifier is more realistic at this stage. The main tasks to be achieved, in this order, will comprise:

1. A survey of background knowledge re-engineering. Ways in which specifications are modified in software engineering, answering questions such as why specifications are modified and how success is measured will be described. The survey will also describe animation and support tools, and case study material will be gathered for experiments and evaluation.

2. An investigation into how to use Lakatos-style reasoning to evolving requirement specifications. Lakatos's work [9] is a veritable tour de force of ways in which mathematicians use inconsistency and exploit ambiguity to develop ideas. We will investigate using the possibility of using Lakatos-style reasoning to automatically generate modifications to specifications which are incomplete, incorrect or inconsistent. A formal specification language, for which animation tools exist that allow some specifications to be executed, will be used, where the results of execution correspond to the observations, or grounded perceptions in the general case, and the specifications are the axiomatisations that are under development. Lakatos's monster-barring method can be represented formally in terms of the conjecture $\forall x(poly(x) \rightarrow euler\_char(x, 2))$. Abstracting further, given the conjecture $\forall x(P(x) \rightarrow Q(x))$ and counterexample m such that $P(m) \wedge \neg Q(x)$, the monster-barring reaction is to argue that $\neg P(m)$. Each party must then revise the definitions of P accordingly.
In the specification case, given some initial functionality specification by an input-output relation R and a range of desired input values, the system may test a specification $\forall x \exists y(input(x) \wedge R(x, y))$ by providing examples of the form m, n where $input(m) \wedge R(m, n)$; the user may then decide that (m, n) are not in fact a desired input-output pair for the task at hand. The system can invoke one of Lakatos's methods, *e.g.* monster-barring, where it is decided that the input n should not be allowed in the range of desired input values. Then the system can suggest modifications to the definition of input range which exclude the problem case. Automatically modified specifications and concepts would be passed to the user as suggested improvements. If the specifications are written in the HOL system or the formal specification language Z, then the system can use existing tools to see how a specified system behaves and to reason empirically about the specifications.

3. Experimentation with axiomatisation system and theory. Using the initial model of our theory of mathematical axiomatisation, developed in the main project, we will devise and carry out experiments to support or falsify the hypothesis that our theory of mathematical axiomatisation can be of benefit to software specifications modification. The model will be tested on a library of case studies where varied and desirable specifications and properties in each case have been documented.

4. A revised theory of software specifications modification. We will analyse the experimental results and revise the theory of software specifications modification derived at the start of the PhD. The theory of mathematical axiomatisation in the main project will similarly be revised in light of the experimental results from the student using the system. Such an analysis will provide both an explanatory aspect to the theory, by explaining why certain results were or were not achieved, and a prescriptive aspect, by providing methods and methodologies by which a researcher in software specifications could identify ways in which to improve the background knowledge to problems.

We will work with Dr. Arthan, who is a well known expert in formal methods and has developed tools to support formal methods, particularly specification and proof in HOL and Z. He will provide technical support with respect to these tools. Additionally, he will provide case study material for the evaluation of this PhD project.

11.2 A theory of knowledge reformulation in AI problem solving

Knowledge based AI techniques such as planning, constraint solving, machine learning, expert systems and automated theorem proving have been major success stories in computer science, finding many important applications across society. Most techniques rely upon the intelligent task to automate to be presented as a problem to be solved, and the solution will involve aspects of a background knowledge-base, often presented in a logical formalism. The quantity and quality of this knowledge-base is crucial to the success of the problem solving technique, hence each method has its own ways of manipulating the background information it is given. For instance, in machine learning, theory revision is used to cope with inconsistencies between examples and definitions of concepts [1]; in constraint solving, constraints are automatically reformulated in order to more efficiently use solvers [24]; in theorem proving, axioms are selected from large sets in order to enable a proof of a theorem to be found in an acceptable time [22]. Drawing from these *ad hoc* approaches, the main tasks to be achieved, in this order, will comprise:

1. A survey of background knowledge re-engineering.

2. A general theory of background knowledge reformulation for problem solving tasks. We will generalise over the techniques implemented in the disparate domains to give a general theory of knowledge reformulation for problem-solving purposes. This will be based on describing the background knowledge in terms of: examples; concepts which categorise the examples; and hypotheses which relate concepts. These will be further sub-divided into knowledge which forms part of the axiomatisation of the problem (knowledge which is explicitly or implicitly taken to be correct, and which is required to describe the domain and the problem to be solved), a supporting knowledge base (additional knowledge which is suspected to be required for the solution) and procedural knowledge (which can help the solver more efficiently find a solution). The theory will address such questions as: in what ways can the various types of knowledge be described as incorrect or lacking; what general methods are available for attempting to fix faulty background knowledge; how will changes in one type of background knowledge (*e.g.*, definitions of concepts in axioms) affect the other types of knowledge; how will changes affect different AI search techniques?

3. Experimentation with the axiomatisation system and theory. As in task 3 above, we will carry out experiments to test the hypothesis that our theory of axiomatisation is useful to AI problem solving. This will involve both (a) using the system implemented in the main project for reformulation of the background knowledge of some standard AI problems, *e.g.*, theorem proving, and (b) using the theory of mathematical axiomatisation to improve existing methods of background knowledge re-engineering, *e.g.*, reformulating constraints in light of a re-axiomatisation of the domain. We will be particularly interested in questioning how one AI technique can be used to pre-process the background knowledge given to another technique. We have already looked at how machine learning can be used to reformulate a problem statement, which can be seen as part of the background knowledge, in both theorem proving [5] and constraint solving [3]. We will experiment, for example, to see whether theorem proving can be used to find inconsistencies in machine learning background knowledge, or constraint solving can be used to extract relevant axioms in theorem proving.

4. A revised theory of background knowledge reformulation. As in task 4 above, we will cohere this theory with that found in the main project.

12 Programme of research

There will be four major projects to be led by the two RAs and two PhD students employed for this research:
(i) embodied axiom formation/modification (based on [10]);
(ii) societal axiom formation/modification (based on [9]);
(iii) application to software specifications modification;
(iv) application to knowledge reformulation in AI problem solving.

12.1 Objectives

We have three main objectives:

1. Produce a general theory of axiom formulation and reformulation based on Lakoff and Núñez's account of embodied mathematics and Lakatos's theory of mathematical discovery and justification (§9).
2. Produce a computational model of the theory and use it to evaluate the theory (§10).
3. Explore the applications of the theory to AI and software engineering (§11).

12.2 Work plan and project management

Below we show our proposed work plan in terms of work packages and tasks, with deliverables and timescales for each work package. The scheduling and dependencies between these work packages is represented in our diagrammatic work plan for both RAs and both PhD students on the project. We will work together in the following way. Alan Smaill will direct and manage the project, and contribute in all areas; Alison Pease will develop the general theory of axiomatisation and construct the system which can generate and modify axioms (WP's 1,5). Paul Crook will determine the simulated world and agents, interface the simulated agents to the reasoning system, and perform initial investigations into the applications of the model, (WP's 2,3,4,6). They will work together to evaluate the system, analyse results and revise the system/theory, evaluate the revised system, and explore the applications of the system and model (WP's 6,7,8,9,10). Andrew Ireland will supervise a PhD student who will investigate software specification modification (WP 11). Simon Colton will supervise a PhD student who will investigate knowledge reformulation in AI problem solving (WP 12). Andy Clark will advise on the embodied cognitive and philosophical aspects of the project (WP's 1,2,4,5,10a,10e).

Thanks to the proximity of Heriot-Watt University and the University of Edinburgh, the RAs and Andrew Ireland and his student can enjoy frequent (weekly) interaction. There will be visits between Edinburgh and London at least every two months for meetings and talks, thus ensuring that all three projects co-evolve in conjunction. We will also hold tele-conferences between all members on the team.

WP1. Formulate a general theory of axiomatisation: (a) use the literature described above and historical case studies to develop a theory of mathematical axiomatisation; (b) incorporate results from WP's (11a) and (12a) to form a general theory.
Deliverables: journal paper including a review of the field and our general theory. *Timescale:* 6 months.

WP2. Determine simulated world(s): (a) decide upon complexity required; (b) select simulation software to use; (c) develop code base to utilise chosen software; (d) design and construct range of simulated environments.
Deliverables: Model of a physical world and our reasons for using it.
Timescale: 3 months

WP3. Construction of simulated agents: (a) decide physical form of agents; (b) decide required level of interaction with the world; (c) decide agents' motor abilities; (d) decide upon and design innate skills of agents.
Deliverables: simulated agents, and our reasons for using them.
Timescale: 6 months

WP4. Interface simulated agents to reasoning system(s): (a) design an interface such that: *(i)* reasoning system can abstract and generalise in order to form axioms, *(ii)* concepts formed by higher reasoner can be made available to simulated agent, *(iii)* simulated agent can test relevance of generalised theory through instantiating variables and testing if they fit observed or observable patterns; (b) examine possibilities of introducing abilities to reason with metaphor.
Deliverables: a set of simulated agents and environments capable of generating interesting patterns of data and testing relevance of abstracted theories to its perception of its world; conference paper on interfacing reasoning systems to embodied agents. *Timescale:* 3 months.

WP5. Construction of pre-axiomatic ATF system: (a) formalise Lakoff and Núñez's work as a series of algorithms and implement them; (b) formalise the general theory as a series of algorithms and implement them; (c) formalise notion of interestingness.
Deliverables: Conference paper on the implementation and preliminary results. *Timescale:* 12 months

WP6. Investigation into applications of the model: to be done in conjunction with WP's 11(c) and 12(c) below.
Deliverables: A theory of applications in at least one AI domain.
Timescale: 6 months

WP7. Evaluation of system: (a) show that grounded axiomatisation by computers is possible, by reproducing inspiring examples; (b) explore effects of varying agents, worlds, innate abilities, etc.
Deliverables: Workshop paper on initial experimental results. *Timescale:* 6 months

WP8. Analyse results and revise system/theory.
Deliverables: Revised theory and model, conference paper updating view on theory. *Timescale:* 6 months

WP9. Evaluation of revised system: (a) show that our model can produce interesting results in non-development domains
Deliverables: Experimental results, conference paper containing results.
Timescale: 6 months

WP10. Explore relevance of our system and model: to (a) mathematics education (6 months); (b) Lakoff and Núñez's work (whether our model

supports or refutes it) (3 months); (c) the philosophy of mathematics (6 months); (d) robotics (3 months)
Deliverables: (a) paper in Journal for Research in Mathematics Education; (b) Cognitive Science paper; (c) journal paper; (d) review of results specifically focused on supplying reasoning systems to embodied agents, journal paper. *Timescale:* 24 months

WP11. investigate specification modification: (a) literature survey on software specifications and support tools (6 months); (b) investigation into applying cognitive and philosophical theories to evolving requirement specifications (6 months); (c) experimentation with axiomatisation system and theory (9 months); (d) revised theory of software specification modification (9 months); (e) write up thesis (6 months).
Deliverables: (a) lit survey; (c) theory of techniques, (d) conference paper; (e) PhD thesis, journal paper. *Timescale:* 36 months

WP12. investigate knowledge reformulation in AI: (a) literature survey on dynamic problem solving in AI (6 months); (b) general theory of background knowledge reformulation for problem solving tasks (6 months); (c) experimentation with axiomatisation system and theory (9 months); (d) revised theory of background knowledge reformulation (9 months); (e) write up thesis (6 months).
Deliverables: (a) lit survey; (c) theory of techniques, (d) conference paper; (e) PhD thesis, journal paper. *Timescale:* 36 months

13 Relevance to beneficiaries

Subfields formed to study individual processes can both contribute to and benefit from a general theory of axiomatisation. It is a fruitful time for interdisciplinary collaboration in AI and we hope to contribute by forging new and important links. The cognitive science of mathematics is a new, multidisciplinary field, and it is therefore the ideal time to produce a computational perspective on this evolving field, since an evaluation will suggest refined versions of the theory. Interfacing robotics with a higher level reasoning system will be of immediate interest to the robotics community. Professor Aaron Sloman at the University of Birmingham suggests that future school children might be the single most important class of beneficiaries of this project, if their education is based on a deep understanding of the nature of mathematical discovery[1].

14 Dissemination and exploitation

Dissemination will be via journal and conference papers, presentations and the free availability of prototype systems. All deliverables will be available on the web as well as via traditional publication routes. In the second year of the project we intend to organise an AISB symposium on grounded embodied cognition. We will be working with Rob Arthan from Lemma 1, who provides consultancy and software tools supporting formal specification. This will give us the opportunity to exploit our methods in his support tools and ensure the industrial relevance of our work. We will also work with roboticists Professor Ruth Aylett at Heriot-Watt University and Dr. Joanna Bryson at the University of Bath, and with Professor Sloman on philosophy and cognition, and we expect them to be able to exploit our model in their work. Our project will provide the first general cognitive theory and computational model of how axioms are formulated and reformulated in mathematics. This will advance the state of the art in how automated theories are formed, AI problems are solved and software is specified.

References

[1] P. Baffes and R. J. Mooney. A novel application of theory refinement to student modeling. In *Proc. of the 13th National Conference on AI (AAAI-96)*, pages 403–408, 1996.

[2] M.A. Boden. *The Creative Mind: Myths and Mechanisms*. Weidenfield and Nicholson, London, 1990.

[3] J Charnley, S Colton, and I Miguel. Automatic generation of implied constraints. In *Proc. of the 17th European Conference on AI*, 2006.

[4] S. Colton. *Automated Theory Formation in Pure Mathematics*. Springer-Verlag, 2002.

[5] S. Colton and A. Pease. The TM system for repairing non-theorems. *Selected papers for the IJCAR'04 disproving workshop, Electronic Notes in Theoretical Computer Science*, 125(3), 2005.

[6] S Colton, M Shanahan, and J Charnley. Combining automated reasoning systems using global workspace architectures. *Submitted to Artificial Intelligence*, 2006.

[7] A. Finkelstein, D. Gabbay, A. Hunter, J. Kramer, and B. Nuseibeh. Inconsistency handling in multi-perspective specifications. *IEEE Transactions on Software Engineering*, 20(8):569–578, 1994.

[8] T Hoare, C. Jones, and B. Randell. Extending the horizons of DSE (GC6).

[9] I. Lakatos. *Proofs and Refutations*. CUP, Cambridge, UK, 1976.

[10] G. Lakoff and R. Núñez. *Where Mathematics Comes From: How the Embodied Mind Brings Mathematics into Being*. Basic Books Inc., U.S.A., 2001.

[11] D. Lenat. *AM: An Artificial Intelligence approach to discovery in mathematics*. PhD thesis, Stanford University, 1976.

[12] W. McCune. Single axioms for groups and abelian groups with various operations. *Journal of Automated Reasoning*, 10(1):1–13, 1993.

[13] R. Miarka. *Inconsistency and underdefinedness in Z specifications*. PhD thesis, University of Kent, 2002.

[14] B. Nuseibeh. To be and not to be: On managing inconsistency in software development. In *8th International Workshop on Software Specification and Design (IWSSD'96)*, 1996.

[15] H. M. Pasula, L. S. Zettlemoyer, and L. Pack Kaelbling. Learning symbolic models of stochastic domains. *Journal of Artificial Intelligence Research*, 2006.

[16] A. Pease, S. Colton, A. Smaill, and J. Lee. A multi-agent approach to modelling interaction in human mathematical reasoning. In *Intelligent Agent Technology Research and Development*. World Scientific, 2001.

[17] A. Pease, S. Colton, A. Smaill, and J. Lee. Semantic negotiation: Modelling ambiguity in dialogue. In *Proceedings of Edilog 2002, the 6th Workshop on the semantics and pragmatics of dialogue*, 2002.

[18] A. Pease, S. Colton, A. Smaill, and J. Lee. A model of Lakatos's philosophy of mathematics. *Proc. of Computing and Philosophy (ECAP)*, 2004.

[19] G. Ritchie and F. Hanna. AM: a case study in methodology. In D. Partridge and Y. Wilks, editors, *The foundations of AI: a sourcebook*, pages 247–265. CUP, Cambridge, 1990.

[20] P. Santos, S. Colton, and D. Magee. Predictive and descriptive approaches to learning game rules from vision data. In *Proc. of IBERAMIA 2006*. Springer-Verlag, 2006.

[21] M. Shanahan. Perception as abduction: Turninng sensor data into meaningful representation. *Cognitive Science*, 29:103–134, 2005.

[22] G. Sutcliffe and Y. Puzis. SRASS: a semantic relevance axiom selection system. In *Proc. of CADE*, 2007.

[23] S. Thrun. Probabilistic robotics. *Commun. ACM*, 45(3):52–57, 2002.

[24] R. Weigel and C. Bliek. On reformulation of constraint satisfaction problems. In *European Conference on AI*, pages 254–258, 1998.

---

[1] Personal communication