Probability, logic and the cognitive foundations of rational belief

John Fox

Cancer Research, UK

Abstract

Since Pascal introduced the idea of mathematical probability in the 17th century discussions of uncertainty and "rational" belief have been dogged by philosophical and technical disputes. Furthermore, the last quarter century has seen an explosion of new questions and ideas, stimulated by developments in the computer and cognitive sciences. Competing ideas about probability are often driven by different intuitions about the nature of belief that arise from the needs of different domains (e.g., economics, management theory, engineering, medicine, the life sciences etc). Taking medicine as our focus we develop three lines of argument (historical, practical and cognitive) that suggest that traditional views of probability cannot accommodate all the competing demands and diverse constraints that arise in complex real-world domains. A model of uncertain reasoning based on a form of logical argumentation appears to unify many diverse ideas. The model has precursors in informal discussions of argumentation due to Toulmin, and the notion of logical probability advocated by Keynes, but recent developments in artificial intelligence and cognitive science suggest ways of resolving epistemological and technical issues that they could not address.
Crown Copyright © 2003 Published by Elsevier B.V. All rights reserved.

Keywords: Argumentation; Decision theory; Knowledge representation; Medicine; History of probability; Agents; Artificial intelligence; Rationality

1. Introduction

"...uncertainty is an inevitable problem in the real world. ...Unfortunately, there are clear gaps in our understanding of how to incorporate uncertain reasoning into a general purpose agent...."

Artificial Intelligence: a modern approach
Stuart Russell and Peter Norvig, 1995 [23] (p. 843)

How we should make decisions in the face of uncertainty and arrive at rational beliefs have been at the centre of intellectual and philosophical thought for millennia. The need for a formal solution to these problems has been recognised as a major scientific and technical challenge for at least 400 years, and in the last 30 years discussions have been further stimulated by problems arising in the cognitive and computer sciences. Perhaps the most influential and admired work on these matters has been carried out by mathematicians and logicians, often informed by problems in some practical domain such as medicine or economics (e.g., [14]). This has led to a mathematical theory of probability having great power and intellectual depth. Yet despite this the field of probability theory has been dogged by philosophical and technical disputes. As Kyburg [18] has observed "Many proponents of many [different] views have argued that their interpretation of probability is the correct (or the most useful, or the only useful) interpretation".

Our goal is a scientific account of how autonomous "agents" (natural or artificial) can, do and ought to accommodate uncertainty in their reasoning and decision-making, drawing on insights from cognitive science as well as mathematics. This research programme adopts a more eclectic methodology than is usual in theoretical discussions of probability and logic. Among the methods used have been observation of naturally occurring behaviour (e.g. the behaviour of doctors making clinical decisions); computer simulation of software agents carrying out complex tasks, and empirical testing of the performance of such agents making decisions and plans in real medical settings. The paper concludes that our continuing failure to resolve well known issues about probability, and the new challenges raised by the cognitive sciences, point to the need for a fresh approach, and develops a system that has intuitive appeal, theoretical coherence and considerable practical versatility.

The paper is organised as follows. In Section 1.1 I put forward three independent lines of argument that there are important shortcomings in conventional accounts of how agents can accommodate uncertainty in achieving their objectives, *viz*: philosophical and technical issues come up repeatedly in discussions of uncertainty and do not look like being resolved (the historical argument); standard methods and technologies provide insufficient means for solving real-world decision problems (the practical argument), and challenges raised by the need for theories of autonomous functioning in artificial intelligence and psychology (the cognitive or "anthropic" argument). Section 1.2 presents a framework for resolving these difficulties through an account of reasoning about uncertainty and belief that explicates the role of "knowledge-based argument" in reasoning, decision-making and other cognitive processes. Section 1.3 summarises the main conclusions.

1.1. Three arguments for extending current conceptions of probability

1.1.1. Historical arguments

The standard history of probability begins with an ancient, pre-mathematical period which is poorly documented but about which historians are reasonably agreed on general points. In Ian Hacking's celebrated account *The Emergence of Probability* [12] we are told that ideas about uncertainty existed by Roman times and became increasingly explicit and diverse through the medieval period and the renaissance. But that they seem to have been at best vague and conceptually rather muddled until the modern concept of probability appeared in a correspondence between Pascal and Fermat, and the central ideas were pub-

lished in the *Port Royal Logic* in 1660. Hacking asks why it took so long for the modern notion of probability to develop, and whether particular technical and historical circumstances were necessary for it to appear. He provides a wide-ranging discussion of the many concepts that were competing for this conceptual space, from chance to fate, deductive to inductive reasoning, and from intuitive ideas about possibility to modern theories of probability.

Among the critical stimuli to a theory of probability may have been a growing practical need to quantify what would now be called social trends, in order to address problems in civic policy and private business. The collection of disease and death rates and other data (by de Witt, Wilkins, Petty, Graunt and others in the mid 1600s) stimulated quantitative ideas about uncertainty, risk and prediction. Once the basic idea of probability had taken root many famous figures got involved in its development (including Leibniz, Bernoulli, Laplace, Huygens, Poisson, to name just a few). In the period from the mid 19th century through the 20th techniques developed rapidly, leading to modern statistics and much other applied mathematics.

Along with the flowering of formal probability theory, however, philosophical disputes about the nature of probability have also emerged. Although we have a deep technical understanding of mathematical probability it is now generally accepted that it is a subject with many subtleties, and many philosophical questions about its relationship with ideas about human belief and rationality have been raised and debated. Recent discussions have involved some of the leading intellects of the modern period, including Ramsay, Savage, Russell, Popper and, a name we will hear more of, the economist John Maynard Keynes. Their attention was attracted because arguments about probability were not merely technically challenging, they often directly impacted on fundamental human concerns of politics and economics, belief, and the nature of mind.

The Hacking space. Hacking's contribution to these discussions is much more than a scholarly history of the subject in which he compares and contrasts the alternative positions. This would be valuable in itself, but he also tries to stand back from the details of the various disputes in order to understand the intellectual space in which the debate is being conducted. Although on the surface there seem to be many contradictions between the competing concepts and philosophies Hacking asserts that there is more historical continuity and coherence than is often recognised. Indeed he invites "...the reader to imagine ...that there is a space of possible theories about probability that has been rather constant from 1660 to the present" (p. 16).

Hacking does not seem to intend the term "space" in the formal sense of a mathematical space, but more as a related collection of ideas that recur in different contexts and with different vocabularies; a metaphorical space within which all the competing theories can be understood as variants of some underlying idea. I am not sure whether he would approve of an attempt to go further but it seems to me that we need to have something more formal. If we do not we cannot systematically compare theories of probability, to identify similarities and differences in their properties, to understand when different interpretations are helpful and so on.

A possible structure for the "Hacking space" is shown in Fig. 1. My hypothesis is that this space has three main dimensions, based on two classical distinctions which appear
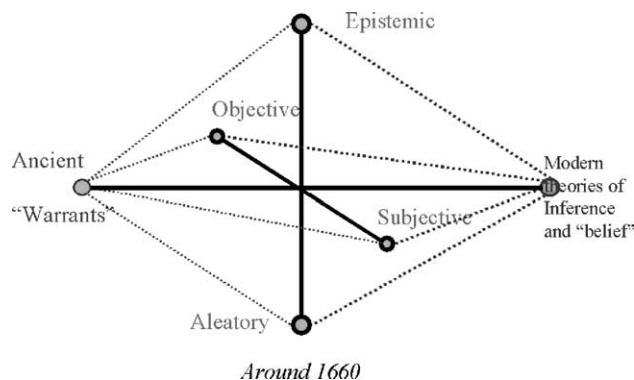
Fig. 1. The Hacking space.

repeatedly in the history of discussions about uncertainty and belief, augmented by a third dimension which has not been discussed much if at all, but reflects the observation that the intuitive phenomenology of belief and formal ideas about probability are not stable but constantly changing through time.

In Fig. 1 time runs from left to right, from some ancient and murky past to the present day with our modern corpus of scientific and other knowledge and our formal theories of reasoning, decision-making, representation of knowledge and the rest. We take the *Port Royal Logic* as the first point at which the *primary dimensions* of Hacking's space are identifiable. The first of these dimensions concerns a very old distinction between events that recur in random or apparently unlawful ways, and those that occur reliably, and for which there is some sort of developed theory. The fall of a fair die cannot be predicted while the position of a planet can. Writers speak of circumstances of the first kind with many words, such as "chance", "fate", "hazard" and so forth. These terms may involve subtle distinctions but they are generally grouped under the heading "aleatory", meaning "Dependent on chance, luck, or an uncertain outcome".[1] This is to be contrasted with an understanding of events in the world through some kind of scientific discourse that draws upon theories about the world, such as causal or geometrical theories and in which the usual modes of reasoning are categorical, logical and deductive. Such terms can be grouped under the heading "epistemic", meaning "Of, relating to, or involving knowledge".

The second dimension of the Hacking space concerns the distinction between objective probability and subjective probability. Hacking spends a considerable amount of time[2] getting to grips with a set of ideas that were around by the seventeenth century and perhaps earlier, which were concerned with "possibility" (as distinct from probability). These terms seem to have become a focus of dispute in the 18th century, with some scholars arguing that possibility is different from probability (e.g., Leibniz and Laplace), and others arguing that it is identical. Hacking relates this discussion to an earlier scholarly debate about *de re* and *de dicto* modes of language, where the former refers to statements about aspects of the

---

[1] http://www.dictionary.com.
[2] Particularly in Chapter 14.

world (e.g., "it is possible for Daniel to get to San Francisco by noon") and the latter about what can be said or known (e.g., "it is possible that Daniel is there now"). Hacking seems to come down on the side of those who regard possibility and probability as equivalent. Indeed he links the *de re* mode with the aleatory aspect of probability (chance, and hence possibility, is a property of the world) and the *de dicto* mode with the epistemic aspect (i.e., an aspect of our *knowledge* of the world) though to my way of thinking the objective-subjective distinction remains intact.[3]

Despite the apparent simplicity of the two primary dimensions of the Hacking space many believe that this is still too complicated and have argued for an even more parsimonious interpretation of probability since a simpler semantics and avoidance of psychological issues should yield practical tools that are both more general and easier to use. In the most influential view probability is viewed as a universal measure that can represent all uncertainty and belief. The "Bayesian" probabilists have been particularly active in this area due to their wish to move away from a strictly frequentistic concept of probability based on empirical observations.

The simplification is achieved in three moves. First, it is asserted that all uncertainty about a proposition can be represented by a single number representing a *degree of belief* in that proposition. This is conventionally a point in the $[0, 1]$ interval[4] whose properties are defined by the probability axioms. Second, the aleatory-epistemic dimension is collapsed by treating all empirical chances as probabilities, and treating epistemic reasoning (e.g., reasoning based on deductive logics) as a special case of probabilistic reasoning in which only two degrees of belief are permitted, 1.0 and 0. These points are viewed as equivalent to "true" and "false" in classical logic. Third, the philosophical complications of the objective-subjective distinction are finessed by assuming that subjective probabilities are technically no different from objective ones and therefore subject to the same theorems. Claims that human judgement and decision-making may take place without use of explicit probabilities is discounted by insisting that all decision makers have a personal probability for all possible states of the world, which can be revealed by forcing the decision maker to gamble on alternative options.

The Bayesian unification is parsimonious and elegant, and it has provided an influential account of the nature of uncertainty. It has also led to an impressive body of technical and practical work, and it provides an appealing framework within which to assess whether reasoning and decision systems are fundamentally *rational* or not by determining whether judgements comply with its prescriptions. Indeed it has been taken up as a foundation for discussion about rational cognition in many domains and its influence pervades much intellectual thought. "Bayesianism was involved in debates central to 20th century philosophy: debates about the ontology of decision-making, belief-revision ... the nature of

---

[3] It has been pointed out to me that the aleatory-epistemic distinction is also often equated with the objective-subjective distinction though I treat them here as orthogonal. In the present discussion I am trying to distinguish between aleatory or epistemic *representations* (e.g., statistical versus deterministic representations) of the world rather than whether or not our uncertainty is due to an inherent uncertainty *in* the physical reality as distinct from our personal ignorance *about* that reality.

[4] Some systems permit beliefs to have associated confidence intervals, upper and lower probabilities, and others allow of higher-order probabilities to express the idea that a degree of belief can itself be uncertain.

explanation, scientific progress, nature of belief and knowledge, rationality and practical reasoning" [22].

Credibility of an exclusively probabilistic position.   Yet many theorists doubt the universality of probability as an exclusive basis for assessing rationality. Doubts commonly arise on the grounds that ideas about uncertainty and belief are far more varied than a simple probability account permits. There are many different *formalisms* and calculi for representing uncertainty for example (e.g., [13,16,21]) and, more importantly, there may be different *kinds* of uncertainty (e.g., [24]). From an everyday point of view many humans claim a rich phenomenology of uncertainty, distinguishing between such notions as belief and doubt; suspicion and conviction; possibility, probability, plausibility, . . . vagueness, ignorance, . . . and so on.

If we follow those who reject the need for such distinctions then it seems to me that we get into deep water. From a psychological or linguistic perspective we should at the very least explain the remarkable fact that there is an enormous number of uncertainty terms in our natural language. Hacking discussed the historical distinction between "possibility" and "probability", but these are just two of many. The best known set of linguistic terms are the "p-modals". P-modals are terms that can substitute for the variable *P-ly* in sentences of the form:

"It is *P-ly* the case that *Sentence* is true"

where *Sentence* is also a variable, which might have values like "Daniel is in San Francisco" or "this patient is seriously ill". Among the most prominent p-modals are "possibly", "plausibly", "probably", "potentially", "provisionally", "presumably" and "perhaps".

Disregarding such linguistic distinctions is to ignore realities of human communication and the complex phenomenology of belief.

The distinction between possibility and probability reflects, it seems to me, the need to distinguish what *can* happen in principle (given known constraints about the world), from what *will* happen in practice (given the many competing influences on events). This has practical implications because a human or other agent will logically need to establish all the possible hypothetical events before establishing their relative probabilities. Furthermore agents may need to explain or discuss some hypothetical situation or event with another agent, distinguishing probable situations (for which there is evidence) from the mere *plausible* (whose existence may be theoretically consistent but for which there is no direct evidence). English includes yet more terms, like *presumably* (if there is no reason to exclude a possibility it should be assumed to be true) and *potentially* (it may or may not be the case now but unless we act it could become so in the future) and *perhaps* (there is at least one scenario in which the Sentence could be true).

Another collection of modalities is *cognitive* in nature, designed to capture a mental state, as in "it is {*conceivable, imaginable, supposable, suspected}* that . . . the patient has contracted Severe Acute Respiratory Syndrome while travelling". Furthermore all these modalities can be used in complex locutions involving lexical negation (e.g., not conceivable) and affixal negation (e.g., inconceivable). The very existence of such locutions suggests a rich and functional diversity in the cognitive states that lie behind human language and thought. Surely we possess this rich vocabulary for a reason? Among the

possible benefits of such terms are that they have both a representational role, capturing our confidence in some proposition, and also a communicational role, succinctly indicating something about the provenance of a proposition in terms of the logical and cognitive justifications on which it depends. We develop these ideas in detail later. My only claim at this stage is that the heterogeneous language of uncertainty indicates a phenomenological aspect of belief that is not merely some degenerate form of mathematical probability.

The Bayesian account ignores these different concepts of uncertainty it does not unify them. A true unification would suggest some conceptual structure that explains the phenomenology of belief rather than imposing an *a priori* technical view of how an agent ought to update some formal parameter. It is true that many Bayesians will argue that their program is normative (prescribing how we ought to reason) rather than descriptive (explaining the phenomenology of human reasoning). This is a perfectly valid restriction, but the objective of this paper is an account in which prescriptive and descriptive theories can be seen as species of some more general structure. We consider what this more general structure could be based on next.

"Warrants" and beliefs.   Hacking gives an important hint about how we might approach the development of a general framework that encompasses both the history of probability and the modern mathematical account. His conjecture about the existence of a space of "possible theories of probability" continues with "This space resulted from a transformation upon some quite different conceptual structure", though he does not identify the kind of structure that he has in mind, nor the specific transformation that took place through history. However, the full Hacking space shown in Fig. 1 is taken to have a further dimension, over and above the aleatory-epistemic and subjective-objective dimensions; this is the timeline which connects ancient modes of thought to modern theories of reasoning and uncertainty.

There do not seem to be many ways that early peoples could arrive at new beliefs about their environments or other circumstances. Prior to the development of language[5] these must be limited to personally witnessing events or situations. With the development of social groups and communication, however, a step-change could take place - an individual could arrive at new beliefs based on the testimony of others. As human culture developed it would quickly have become impossible for individuals to know everything about current affairs, the law or about good practice in agriculture, caring for the sick etc and dependency on third parties must have been unavoidable once communities reached even a few hundreds of individuals. This would quickly raise the question of trust, and the claims and opinions of others would need to be "warranted" in some way if they are to be accepted. In pre-technological societies notions like individual status (kings and chieftains) and organisational prestige (the church) would provide such warrants. Later people came to depend upon the opinions of specialists like doctors and lawyers whose judgement was warranted by their perceived (or claimed) knowledge and expertise. As the centuries passed and society's collective knowledge base grew, more and more abstract kinds of warrant would have to be accepted, in the end becoming based on disembodied theories of mathematics,

---

[5] A notional case of an "early people" unless one considers the possibility that other hominids, or even other primates might have cognitive states comparable to human "beliefs".

Table 1
Species of content and modes of reasoning in 20th century medicine

"Theories" within the knowledge base of modern medicine	Some of the modes of reasoning used routinely in modern medicine and other knowledge-based disciplines
☐ Immunology ☐ Physiology ☐ Anatomy ☐ Biochemistry ☐ Genetics ☐ Morphology ☐ Epidemiology ☐ Mental health ☐ Social dysfunction	☐ Causal ☐ Statistical ☐ Functional ☐ Structural ☐ Spatial ☐ Temporal ☐ Deontic

navigation, politics, engineering, and science, and in the modern era, the new "channels" of television and Internet.

Over time the "content" of our collective and individual knowledge has differentiated into ever more detailed conceptual systems.[6] By the time that Hacking's two-dimensional space of probability ideas had emerged in the 17th century the repertoire of theories that could justify individual decisions and organisational policies had clearly grown enormously. European lawyers had a well-established jurisprudential theory, doctors and apothecaries had theories of diseases and their proper treatments, and ecclesiastical law had reached a high level of refinement. (Other societies had different knowledge bases of course.)

The pace of differentiation continues to grow. Modern medicine, for example, draws upon many small theories, almost none of which were known 200 years ago. As the knowledge base grows new theories and modes of argument become articulated and refined allowing us to make ever more diverse and subtle inferences. In Table 1 the combination of knowledge of these different types yields the potential to provide a vast range of "warrants" for diagnosing, explaining and predicting clinical conditions.

Like Hacking, however, I suspect that although 21st century people are capable of formulating arguments with a degree of sophistication well beyond the capabilities of our forebears, our basic cognitive functioning is much the same. We just have a much larger knowledge base on which to ground a greater repertoire of arguments for what we believe. Warrants are the ancient conceptual structures that Hacking alludes to. They share an important feature with modern argumentation in that they are *grounded* in some body of prior knowledge (e.g., medicine) and some accepted mode of reasoning (e.g., authority or testimony, causal or statistical modes of inference). This suggests that any understanding of ideas like uncertainty, belief, doubt and so forth is incomplete without an account of the *reasons* for an agent's beliefs (or doubts); the "warrants" of earlier times and the "arguments" of today. We cannot understand how ancient or modern people may reason

---

[6] "Knowledge is fractal" Alan Rector.

under uncertainty (or how artificial agents ought to) without understanding the nature of knowledge and its role as warrant for beliefs and actions.

The position I shall take is that as a society's knowledge base grows and diverges over time the opportunities to develop and apply more sophisticated argument schemas also grow, though the basic format of the argument is more or less constant (Hacking). I shall suggest, in the terminology of computer science, that arguments should be viewed as *first-class objects* in any theory of uncertainty and belief, i.e. explicit objects of reasoning. If the grounds and arguments for an agent's beliefs are not explicit it cannot do many things, including reflecting on its beliefs and their provenance and explaining the reasons for particular propositions or claims to others. This idea is developed in more detail in part 3, but first we look at some of its practical benefits of a warrant-based view of reasoning and decision-making under uncertainty.

1.1.2. Practical argument

"As living and moving beings, we are forced to act . . . [even when] our existing knowledge does not provide a sufficient basis for a calculated mathematical expectation."

John Maynard Keynes

Probability theory has provided a powerful foundation for many important mathematical techniques, from statistical methods in science and social policy to technological risk analysis and economic decision-theory. The practical success of probability concepts have led many theoreticians to view it as much more than a useful mathematical technique and for many it has been elevated to the level of a touchstone for those wishing to "bring human judgement under the authority of mathematics" [14]. A notable example of this somewhat autocratic perspective is in the development of statistical decision theory:

"... there is essentially only one way to reach a decision sensibly. First, the uncertainties present in the situation must be quantified in terms of values called probabilities. Second, the consequences of the courses of actions must be similarly described in terms of utilities. Third, that decision must be taken which is expected on the basis of the calculated probabilities to give the greatest utility. The force of 'must' used in three places there is simply that any deviation from the precepts is liable to lead the decision maker in procedures which are demonstrably absurd" . . . "The first task in any decision problem is to draw up a list of the possible actions that are available. Considerable attention should be paid to the compilation of this list [though] we can provide no scientific advice as to how this should be done."

Dennis Lindley, 1985, p. vii

Notwithstanding Lindley's convictions expected utility models of decision-making have substantial practical limitations. One that has been widely discussed is the difficulty of estimating precise probabilities and utilities in real-world settings. In medicine for example epidemiological knowledge is surprisingly sparse; except in rare cases where there is reason to carry out national or regional epidemiological studies the critical statistics (the prior probabilities of diseases, and the conditional probabilities of symptoms given diseases) are

not known with any precision. Doctors often do not even have a basis for estimating these things for the community in which they live.

Lindley's unwillingness to give advice on how to determine the possible diagnoses and associated relevant clinical data makes his recommendations even more unhelpful for everyday medicine—where the determination of these things is at the core of practical clinical work. Not only do medical practitioners typically lack quantitative statistics for the decisions they are required to make they may even be undecided about the structure of the decision: the hypotheses that should be considered, the sources of evidence that are relevant, and even the decision to take.

The point of these observations, as Keynes saw, is that probability (and hence decision theory) presupposes a well-defined problem and a tight set of constraints for its use to be appropriate. In practical circumstances, problems and decisions are often ill-formed—a decision-maker may even know little about the logical structure of the task at hand yet *action is still needed*, perhaps urgently. Under these circumstances we need a theory of reasoning under uncertainty that tolerates the absence of quantitative data, provides an account of how to structure the decisions and modify this as circumstances change. In the rest of this section we address the first problem and turn to the second in the next.[7]

Three examples of medical "decision support systems" are now outlined. These illustrate some common decisions that doctors face and show how they can be addressed with a non-probabilistic approach (Fig. 2). The decisions do not depend upon the availability of precise probabilities and/or utilities but largely on qualitative or at most "semi-quantitative" rules. In these systems the specific argumentation processes are mathematically ad hoc yet the applications have proved to be surprisingly successful and robust; in Section 1.2 we shall describe a principled framework in which to build these and other argumentation systems.

Prescribing drugs for common conditions.   CAPSULE[8] was developed to assist general practitioners in routine prescribing decisions [27]. CAPSULE has a database of information about drugs, and a set of logical *if . . . then . . .* rules for deducing potential benefits and harms and other relevant attributes of candidate medications. The rules are formalised as rules in first-order logic. When the system is invoked the rules are instantiated with information from the patient notes and with knowledge from the drug database. The system then generates a list of candidate treatments based on its drug knowledge, and applies the rules in order to construct arguments for and against each option. For example, if CAPSULE is considering drug A and drug A is relatively cheap by comparison with the alternatives this is an argument in favour of A. On the other hand if the patient is already taking drug B and B is known to have an undesirable interaction with A then this is an argument against A. CAPSULEs rules cover 9 factors, including knowledge of drug efficacy, contra-indications, drug interactions, side effects and relative costs. Finally, CAPSULE simply counts up the arguments for and against each drug and presents the set of candidates in an order based on the ratio of pros to cons (Fig. 2). A controlled study with practicing doctors showed that

---

[7] These subjects can only be dealt with briefly here: They are developed in more detail in (Fox and Glasspool, forthcoming).

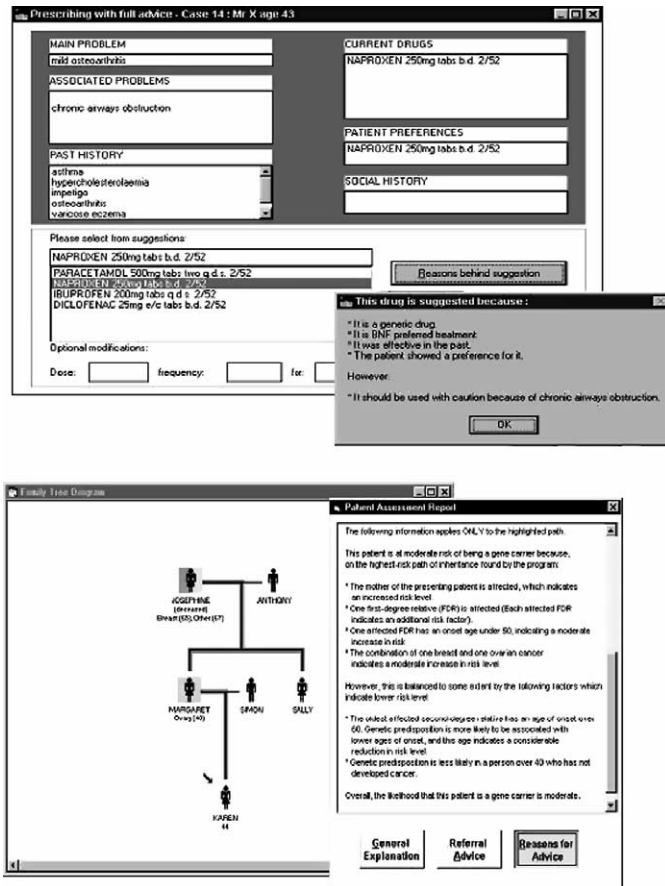[8] Computer Aided Prescribing Using Logic Engineering.

Fig. 2. Illustrations of CAPSULE (top) and RAGs (bottom).

CAPSULE produced a 70% increase in the number of times their decisions agreed with those of experts considering the cases, and a 50% reduction in the number of times that they missed a cheaper but equally effective medication.

Assessing genetic cancer risk.   RAGs[9] helps a healthy woman who has a family history of a disease, such as breast cancer, to systematically construct a family tree and record information about family members who are believed to have contracted the disease. For example, it will ask for information about the relationship between the woman and her affected relatives, what their approximate age was at diagnosis and so on. RAGs then uses a logical decision procedure to assess the risk that the woman is a "gene carrier" for the disease, assessing whether she is at population ("normal") risk, moderately elevated risk or high risk. RAGs does this by applying *if . . . then . . .* rules such as "If the person has

---

[9] Risk Assessment in Genetics.

two or more first-degree relatives diagnosed with breast cancer and both relatives were diagnosed under the age of 40 then this is a risk factor for genetic predisposition to the disease" (Fig. 2). Unlike CAPSULE RAGs rules are "weighted" to represent the relative importance of different risk factors. Each rule is weighted with an integer between 1 and 3 (low significance = 1, medium = 2, high = 3). The overall risk for the woman being a gene carrier is determined by establishing which rules are true in her specific case and adding up the associated weights. The risk classifications generated by the RAGS software was compared with that provided by the leading statistical risk assessment software for 50 families with known genetics; the two systems produced identical results for all the families [4].

Interpreting medical images.   CADMIUM[10] combines process scheduling and decision-making in order to assist medical radiographers carry out the tasks required in screening women for the presence of asymptomatic breast cancer. CADMIUM schedules and enacts the acquisition of mammographic X-rays and the tasks required for reporting on them. During image acquisition it automatically analyses image features and gives advice on whether any abnormalities are likely to be caused by cancer. As with CAPSULE and RAGs the decision-making advice is generated by translating input information into logical arguments for and against identified abnormalities being malignant or benign. All arguments were treated as having equal weight. In an evaluation of this decision procedure medical radiographers were asked to review a set of mammograms to find any abnormalities and make decisions about the diagnosis. CADMIUM demonstrated clear improvements in their ability to achieve this successfully [25].

Each of these three systems applies a body of medical knowledge to the decision-making process. CAPSULE knows about drugs and their uses, RAGs is equipped with knowledge about genetics and rough statistics, while CAPSULE incorporates knowledge of disease processes and their effects on structural and morphological abnormalities.

In Fig. 2 the CAPSULE prescribing system takes patient information (problem, symptoms, current drugs), generates a set of possible medications (bottom left) and constructs a set of arguments for and against each (inset box). It uses the arguments to lay out the options in order of preference. RAGs takes in information about the patient (Karen) and her family history and constructs a family tree. It then constructs a set of arguments for and against her having a genetic predisposition to breast cancer based on this information (right).

Medicine has been an important area for developing decision support systems of many kinds, culminating most recently in the development of knowledge-based expert systems that emphasise the use of logic and human-oriented representations of knowledge.[11] CAPSULE, RAGs and CADMIUM lie in this tradition, but were developed with the objective of investigating the practical strengths and weaknesses of decision-making procedures that can use qualitative and semi-quantitative inference procedures when traditional quantitative methods are impractical or inappropriate. None of these systems makes use of classical

---

[10] Computer Aided Decision Making and Image Understanding in Medicine.

[11] See http://www.openclinical.org which has a major repository of system descriptions and results.
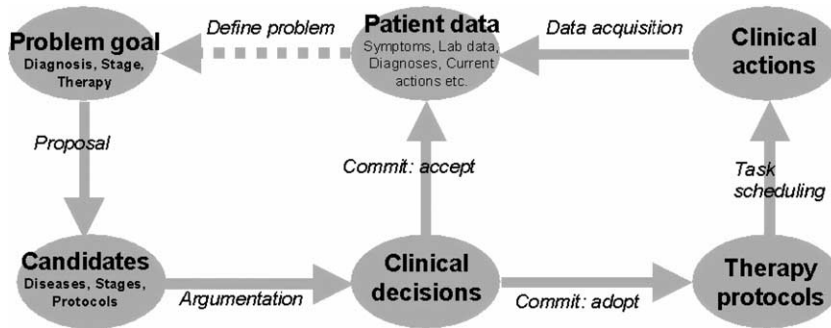
Fig. 3. A general medical decision procedure.

decision-making techniques or probabilities but all demonstrate a good level of competence on the medical tasks they are designed for.

This is not to say that it would not be possible to incorporate standard decision-making methods in the applications. However, there is a practical price to be paid for using quantitative techniques (e.g., the costs entailed in parameter estimation) which may not be justified by the improvements in decision quality that are produced, if any [9]. The general conclusion from this is that probability based decision procedures are not necessarily and exclusively the technique of choice for practical decision-making.

Studies of this kind have allowed us to develop a general model of the decision and other processes that are carried out in complex medical domains (Fig. 3). The model has been described in detail elsewhere (e.g., [7]) so suffice it to say here that according to this account clinical thinking can be well described as a collection of logical processes that reason over and update cognitive representations of general medical knowledge and specific clinical situations.

In Fig. 3 given information about a patient the clinical goals are established (top left) and then the options for each goal (bottom left). Arguments for and against each option are constructed and used as the basis for a decision about what to believe (accept) about the situation, or what to do (which plan to adopt). Plans decompose into collections of tasks which may yield new information and hence new goals in a cyclical process carried out over time [7].

1.1.3. Cognitive or "anthropic" arguments
"Modelling the human is central to logic"
Dov Gabbay, *de Morgan Workshop on Combining Probability and Logic*, 2002.

Formal investigations of the nature of uncertainty have been carried out in many fields, from the higher realms of logic and statistics to the "low sciences" of medicine and commerce. In the last 40 years or so we have seen the rise of a new strand in western scientific thought, that of "cognitive science". Cognitive science is an umbrella term for a number of disciplines, a number of which share the scientific objective of understanding intelligence, as exemplified by the human mind or by robots and other artificial agents that implement complex cognitive functions like reasoning, decision-making, natural language and the

possession of knowledge. Cognitive science has become a new driver for discussion of ideas about uncertainty and belief. Among the most prominent areas of cognitive science where novel ideas are emerging, are psychology and AI.

In the 'fifties and 'sixties psychologists accepted orthodox ideas about logic and probability as the standard against which human judgement and rationality were to be assessed in many respects. But it quickly became evident that orthodox frameworks had surprisingly little to say about how people actually make decisions and reason under uncertainty, culminating in H.A. Simon's concept of a bounded rationality which seems to conform poorly with conventional economic theories of "rational" decision-making. This work, which some claim led to a revolution in micro-economics, was awarded the Nobel prize in 1978.

In the 'eighties and 'nineties a new research agenda began to emerge, largely stimulated by Kahneman and Tversky's programme of research into the "heuristics and biases" that underpin human reasoning under uncertainty [29]. Few psychologists now accept the (exclusive) jurisdiction of normative probability and decision theory as the basis for understanding human decision-making.[12] Indeed, many doubt that they even represent a gold standard against which judgement *ought to* be assessed. It now seems likely that biological, environmental and other demands on mammalian cognitive function created a wider range of needs and constraints than purely mathematical ones (e.g., [11]). Natural agents, like humans and animals, must operate in a world in which environments are unpredictable and even capricious, time is of the essence, mental effort and computational resources are limited. Decision processes that can meet these difficulties must be optimised over more parameters than those recognised in the axioms of probability theory.

Simon was also one of the founders of another vigorous branch of cognitive science: Artificial Intelligence. AI is also driving new lines of thought in logic and probability for related, though different, reasons to the trends in psychology. Where psychologists are finding that human judgement and decision-making depart in significant ways from the prescriptions of logical and probabilistic notions of rationality, designers of software agents and robots have also encountered challenges that do not arise in conventional discussion of logic and mathematical probability and will demand new capabilities. Orthodox probabilistic theories of uncertain reasoning do not provide enough representational power for designing and constructing "intelligent agents" that can operate successfully in complex and unpredictable environments. Why is this?

There is no universally accepted definition of an "intelligent agent" but it is widely accepted that the notion can be captured in terms of a small number of behavioural and cognitive characteristics. One influential summary is due to Wooldridge and Jennings [28] who suggested that agents are characteristically:

- *Proactive*—showing the ability to exhibit goal-directed behaviour.
- *Reactive*—having the ability to be able to respond to changes in the environment, including detecting that its goals are at risk.
- *Social*—interacting, cooperating and negotiating with other agents.

---

[12] It may be interesting to note that the only Nobel prizes awarded to cognitive scientists to date were to Simon and Kahneman.

- *Autonomous*—making decisions and taking actions independent of others.

Agents with these features raise fundamentally new challenges to our ideas about judgment, belief and similar matters.

Proactive.   Reasoning and decision-making do not take place in a vacuum, but are grounded in the agent's *goals* that arise in response to events in the context of its ongoing needs and prior beliefs. Complex goals (e.g., finding food or a mate; planning and executing a successful medical procedure or business campaign) require extended plans and actions that often need to be synchronised to be successful. Goal-directed tasks may also need to address multiple objectives. A general theory of rational inference and decision-making must be grounded in this setting, where beliefs, goals and intentions may interact in complex ways, and in which decisions may impact on any number of current goals and tasks.

Reactive.   Despite the need to plan and execute behaviour in a coordinated way an agent must be able to respond to unexpected circumstances that represent threats or opportunities, as and when they happen. In familiar situations the agent may respond by simply applying a pre-programmed response, but in unfamiliar situations it will need to adapt existing strategies and plans to meet its goals. In some cases the agent may need to completely abandon a current strategy in response to changed beliefs.

Autonomous.   When new goals are raised an agent must be able to solve problems and take decisions by itself. It cannot be dependent on an external programmer or "decision analyst" to set up the decision process. It must be able by itself to identify candidate solutions to achieving its goals, identify relevant knowledge and criteria for choosing between decision options and implement procedures to obtain required information. Indeed, at a higher level it may need to reason about the decisions it needs to take, when they need to be scheduled against other tasks and so forth. "Should I attempt a diagnosis? Or just make a risk assessment? Is it sufficiently urgent that I should go straight for a treatment decision? Or should I refer the patient to a colleague who is more experienced?"

Social.   An agent may not have sufficient resources to permit it to solve problems or make decisions by itself, and for certain tasks it may need to be able to communicate and collaborate with other agents which have access to resources which are not directly available. To do this it will need to be able to engage in dialogues with other agents, to inform them of its beliefs, goals and intentions, and to explain its reasons for these "mental states" if these are questioned or challenged.

The ability to implement such capabilities requires cognitive capacities that are well beyond computing engines that execute simple logical rules or arithmetic functions. It entails adaptation to the unexpected and, particularly, a capability for the agent to *reflect* on its beliefs and intentions. Since an agent's environment will be undergoing constant change past decisions and commitments may cease to be valid. It would therefore be desirable for the agent to be able to reason about its commitments and their justifications, question its
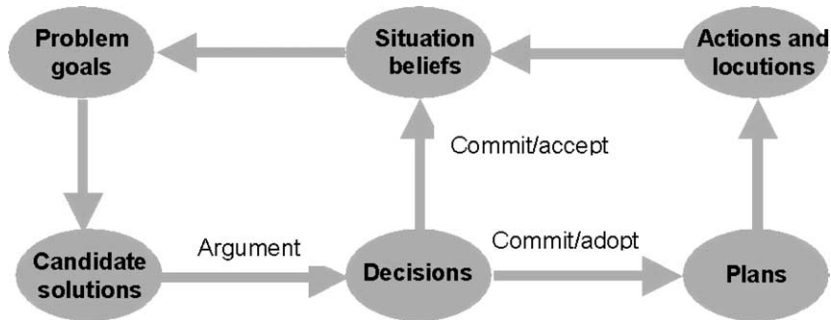
Fig. 4. The generalised "domino" agent.

assumptions, reverse previous decisions and abandon earlier goals or plans. If agents are to work together to achieve collective goals they will need to be able to communicate and discuss their beliefs and the provenance of those beliefs. Such meta-cognitive capabilities cannot be captured by simple rules or other conventional logical machines and require formalisms with meta-logical expressiveness, such as specialised first- and higher-order logics (see the extensive collection of papers edited by Abramson and Rogers, [1]).

Traditional accounts of uncertain reasoning make no provision for the four "anthropic" features summarised above, or the meta-logical capabilities they entail. This implies a context in which to investigate ideas about reasoning and rationality that is radically different from the context in which discussions of logic, probability and decision have traditionally been carried out.

In recent years AI research has set about developing practical architectures and associated theories of intelligent agents that have these anthropic characteristics and meta-cognitive capabilities. An important class of such agents are so-called Belief–Desire-Intention or BDI agents [10,20]. Fig. 4 is an example of an agent that falls within this general class but has been extended to support reasoning and decision-making under uncertainty. It is a generalised version of the clinical process model in Fig. 3 earlier and has emerged from our efforts to build general-purpose anthropic agents based on this model [8]. Here, the ellipses can be thought of as data-structures; the arrows as inference systems. The model has a formal semantics and is the basis of a practical development system [7].

The theoretical pivot around which this agent system reasons and makes its decisions is a generalised procedure based on the application of knowledge and logical argumentation. In the next section we explain how this serves the anthropic capabilities discussed above.

1.2. Knowledge, argument and belief

"... whenever we pass to knowledge about one proposition by contemplation of it in relation to another proposition of which we have knowledge ..., I call it an *argument*"
                                        Maynard Keynes, *A Treatise on Probability* 1921 (p. 14).

To recap on the main points so far traditional ideas about rational judgement and decision have put *quantified degrees of belief* or probability at the centre but in Section 1.1 we

have argued that qualitative or semi-quantitative inference techniques are often more flexible and practical. Many theorists find it hard to accept that simple qualitative arguments could be adequate for practical applications that people find challenging, yet this is frequently the case. In a domain like medicine, which is rich in different kinds of knowledge, logical arguments are commonly all that is required for building useful decision systems. This frees the designer from the rigid constraints of the probability calculus and greatly simplifies the technical problems of developing such applications.

An important exception to the usual position was J.M. Keynes who in his *Treatise on Probability* [15] tries to interpret the idea of probability from a logical as well as a quantitative perspective, observing that "in its most fundamental sense [the term *probability*] refers to the logical relation between two sets of propositions" (p. 11) and "... whenever we pass to knowledge about one proposition by contemplation of it in relation to another proposition of which we have knowledge ... I call it an *argument*" (p. 14). This seems close to my own position. Keynes saw some of his own limitations in developing these ideas, saying for example that he did not "wish to become involved in question of epistemology which I do not know the answer".[13] Because of technical developments that have taken place in the last eighty years, however, it is possible to go further than he could, particularly in developing a formal account of argumentation and the "knowledge" that arguments exploit. In order to make my proposals credible we need to provide a clearer description of exactly what an argument is and its relationship to "knowledge". I shall do this in two stages: first to present an intuitive account of the processes of argumentation, and then to provide a more formal treatment.

1.2.1. Arguments and probabilities, the intuition

In a classical logic an argument is a sequence of sentences (the premises of the argument) from which we can derive another sentence, the conclusion, under some set of inference rules (the logic). Normally, the conclusion of the argument is assigned the value "true" if it can be derived by mechanically applying the inference rules of the logic. As a method of reasoning classical logics, such as the standard propositional and predicate logics, are powerful, but do not directly represent any uncertainty that may be encountered in practical domains. Classical logics do not represent *degrees* of truth or accommodate *changes of mind* or *relative* truth (when a proposition is viewed as true from one point of view but false from another). Our approach is based on a more familiar type of reasoning than has traditionally been studied by logicians. Here arguments are not proofs but *reasons to believe* (in some statement) or *reasons to act* (in some way). Individual arguments are not generally conclusive, so decision making may require us to assess *collections* of arguments, weighing up the "pros" and "cons" as in everyday decision making processes.

The distinction between informal argument and formal theories of reasoning was also recognized by the philosopher Stephen Toulmin. In *The Uses of Argument* [26] he explored the question of why traditional theories of reasoning, notably classical deduction and probabilistic reasoning, have little apparent relevance to everyday dispute and debate. He concluded that informal argumentation is a reasoning procedure that is different from

---

[13] J.K. Galbraith is said to have observed that there are two types of economists, those who do not know; and those who do not know they do not know.

```
     Data ─────────────────→ Qualifier, Claim
                  │                      │
                  │                      │
              Warrant                 Rebuttal
                  │
                  │
              Backing
```
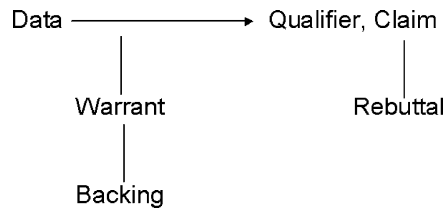
Fig. 5. Toulmin's argument schema.

both mathematical traditions, and characterized it by means of the well-known schema reproduced in Fig. 5.

Toulmin explains the schema with the following example:

"In support of the claim that Harry is a British subject, we appeal to the datum that he was born in Bermuda, and [the claim is warranted by a sentence such as] 'a man born in Bermuda may be taken to be a British subject': since, however, questions of nationality are always subject to qualifications and conditions we shall have to insert a qualifying 'presumably' in front of the conclusion and note the possibility that our conclusion may be rebutted in case it turns out that both his parents were aliens or he has since become a naturalised American. Finally, in case the warrant itself is challenged, its backing can be put in: this will record the terms and the dates of enactment of the Acts of Parliament and other legal provisions governing the nationality of persons born in the British colonies" [26, p. 104].

Two points stand out here: first the idea that *conclusions are not always certain* (hence the qualifier 'presumably') and that *practical argumentation frequently entails contradictions* (the notion of rebuttal). When we look at decision-making in domains like medicine we come across similar problems to those noticed by Toulmin. First, medical professionals have to make decisions in situations where they are not certain about information and its evidential significance, but it is impractical to precisely state their degree of uncertainty. They may suspect that a patient is at risk but have little basis for quantifying the likelihood of the hazard. Second, in many settings we must deal with apparent inconsistencies, as when we are faced with a patient who is clearly ill but a routine lab-test shows nothing abnormal, or in one clinician's opinion the patient is suffering from one condition but has a different diagnosis in the opinion of another.

Contradiction does not have a place in classical logic or probability theory; something cannot be both true and false nor have a probability of 0 and 1. Toulmin's approach anticipated the current interest in symbolic representations of uncertainty in AI and logic, as well as the current interest in accepting and managing contradictions in commonsense reasoning and paralogical systems. However, there are important aspects of argumentation that are not dealt with by Toulmin. These are illustrated in Fig. 6.

The large ellipse at the top left of Fig. 6 represents the knowledge base of a hypothetical agent. This knowledge base is partitioned into a number of different "theories". These may be "commonsense" theories, capturing general assertions about the world (dealing with time, space, properties of physical objects and so forth) or specialist knowledge in some
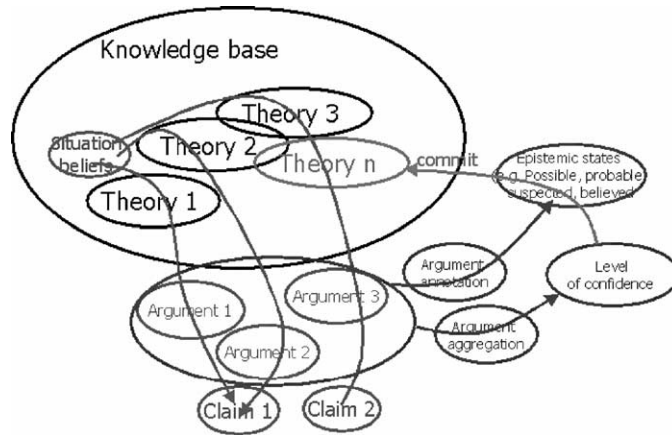
Fig. 6. Visualisation of an argumentation process.

technical domain like medicine (e.g., knowledge of anatomy, biochemistry, physiology, immunology, symptomatology as in Table 1).

Generally, I shall view an agent's knowledge base as a collection of sentences in some formal language. Some sentences assert general facts about the world (e.g., "all cancers are diseases") while others represent assertions about the agent's specific situation (e.g., "this patient may be suffering from breast cancer"). AI research has shown that knowledge can be formalised in many ways, but we shall view it as a database of sentences in first-order logic.

Now suppose that an agent of this kind acquires some information about a particular situation, such as an abnormal medical condition, and the agent has a general goal in such circumstances which is to find the most likely cause of such an abnormality. Let us suppose the agent is presented with a patient who has suddenly lost weight, so it wishes to know why. The agent might formulate several candidate explanations for this (such explanations or hypotheses are called "claims" in Fig. 6), including the possibility that the loss of weight might be caused by a gastric ulcer, which can discourage normal eating and digestion (argument 1), or perhaps the patient is covertly dieting (argument 2). The process of constructing such arguments can be viewed here as a conventional proof-procedure where "the patient has lost weight due to an ulcer" and "the patient has lost weight due to anorexia nervosa" are derived mechanically in the same way as a logical theorem can be deduced from premises, but here the conclusion is uncertain. Another way of putting this is that an argument has the *form* but not the *force* of a proof.

Once we have a set of candidates the agent can bring to bear any number of theories from its knowledge base. For example the agent could argue for the hypothesis of gastric ulcer on the grounds that the patient has pain after meals, using *causal knowledge* that gastric acids irritate pain receptors in the lining of the stomach which are exposed by the ulcer. On the other hand it may argue on *statistical* grounds that peptic ulcer is unlikely because the patient is only 20, and a peptic ulcer in a patient under 50 is very rare. In fact we can develop any number of arguments for and against the alternative claims, drawing
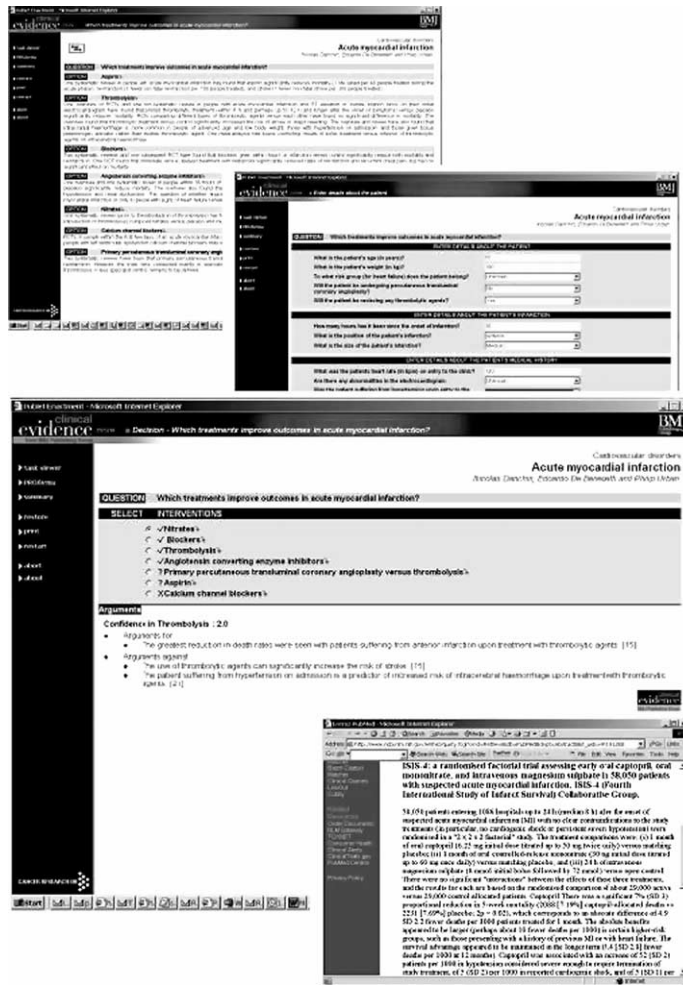
Fig. 7. The use of argumentation in medical decision-making. Clinical Evidence is a standard reference text that doctors and others use to review the potential benefits and harms of possible clinical interventions for their patients (top). This version of the publication has a decision support process embedded in it, which accepts patient data (second panel), and then constructs arguments for and against the options in order to arrive at recommendations for action (large panel). Each option is a "claim" and each argument has a "backing" which can be followed up by finding the relevant paper in the Pubmed database (bottom).

on different subsets of knowledge and applying different modes of reasoning such as those listed in Table 1.

We can also see from Fig. 6 that the construction of arguments is just the first step in a process that results in new information being added to the knowledge base. In the generalised agent model in Fig. 4 this step is to *commit* to a new belief. For example a medical decision agent may consider the various arguments about a patient who has

lost weight and commit to a particular diagnosis. This is then added to its current set of situational beliefs.

Fig. 6 illustrates a circumstance where there are two claims and three arguments. In fact there can be any number of claims and any number of arguments for and against each claim so how may we select the most plausible claim? In standard probability theory we would resolve the competing claims by assessing the posterior probability of each claim to decide which was correct once all the data are in. However, as we have noted above, in many decisions we do not know what the probability parameters are. Ordinary logic cannot help because it does not have any way of representing uncertainty.

Argumentation provides an alternative. Unlike logic we can have any number of arguments with respect to a possible claim, and we can simultaneously entertain any number of tentative conclusions. Furthermore we can compare the persuasiveness of our tentative conclusions by comparing the arguments for and against the competing claims.

This leads to a simple decision procedure. If we have several independent arguments for one claim, but only one argument for the other, then even if we cannot assign "strengths" to the individual arguments it is reasonable to have more confidence in the first claim than in the second. This process of argument, "aggregation", allows us arrive at a rational assessment of the overall persuasiveness of the competing claims even if we cannot establish the absolute or relative strength of supporting arguments. An example of an argumentation and aggregation process in medical decision-making is illustrated in Fig. 7.

1.2.2. A formal model of argumentation

In this section and the next we give a formal description of the processes of argument construction and aggregation.

A basic schema for a "logic of argument" (LA) is summarised in (1):

$$\text{Knowledge base} \cup \text{Situation} \vdash_{LA} (\text{Claim} : \text{Warrant} : \text{Qualifier}) \tag{1}$$

We start with a *knowledge base*, a collection of propositions and rules about a domain, and a collection of propositions that describe a specific situation. The turn-style stands for a set of inference schemas that define valid argument constructions based on this information. There are several options for these axioms here but, whatever the choice of axioms, the *argument term* is a triple of elements (on the right) consisting of

- the *Claim* which the argument deals with;
- the *Warrant*[14] is a representation of the theories, facts and rules drawn from the knowledge base instantiated with information about the specific situation in which the argument has been constructed;
- a *Qualifier* that represents the confidence in the claim as warranted by the argument.

A set of inference schemas for LA is given in Table 2. The Greek characters φ, ψ, σ represent propositions (which can play the role of the assumptions or the claim of an argument) and $\wedge$, $\vee$, $\neg$ are the usual logical connectives, *and*, *or* and *not*. The schemas

---

[14] Elsewhere we have sometimes used the term "grounds" for this component of the argumentation term.

Table 2
A logic of argument

Introduction rules		*Elimination rules*	
(∧I)	$\varphi : Q \quad \psi : Q'$ ——————— $\psi \wedge \varphi : \min(Q, Q')$	(∧E)	$\psi \wedge \varphi : Q \quad \psi \wedge \varphi : Q$ ——————— $\varphi : Q \qquad \psi : Q$
(→I)	$\varphi : \overline{++}$ ⋮ $\psi : Q$ ——————— $\psi \rightarrow \varphi : Q$	(→E)	$\varphi : Q \quad \varphi \rightarrow \psi : Q'$ ——————— $\psi : \min(Q, Q')$
(∨I)	$\varphi : Q \qquad \psi : Q$ ——————— $\psi \vee \varphi : Q \quad \psi \vee \varphi : Q$	(∨E)	$\varphi : \overline{++} \quad \psi : \overline{++}$ ⋮ ⋮ $\varphi \vee \psi : Q \quad \sigma : Q' \quad \sigma : Q''$ ——————— $\sigma : \min(Q, Q', Q'')$
(¬I)	$\varphi : \overline{++}$ ⋮ $\bot : Q$ ——————— $\neg \varphi : Q$	(¬E)	$\varphi : Q \quad \neg \varphi : Q'$ ——————— $\bot : \min(Q, Q')$
		Weakening	$\varphi : ++$ ——————— $\varphi : +$

have the *form* of conventional premise-conclusion rules, but they may not have the *force* of a classical deductive rule in which the conclusion can be only true or false. According to [1] an argument has three elements, the claim, the warrant and the qualifier. For clarity the warrant is omitted from the schemas in Table 2.

Arguments affect confidence in a claim but may not be categorical. It may not even be possible to say by how much an argument affects confidence, just perhaps that it is increased or decreased. The representation for confidence is the qualifier. This can take many possible forms. For example, it can merely say that an argument "supports" a claim or "opposes" it, meaning that the argument increases or decreases our confidence in a claim but does not indicate by how much. In Table 2 we use two other qualifiers, "+" and "++", where + indicates the argument increases confidence while ++ is the *top* symbol, meaning the argument increases confidence in the claim to a maximum (though again without giving a quantitative interpretation for this maximum). Qualifiers can also be represented by variable symbols, Q, Q′, Q″ etc.

The rules are presented in the Gentzen style with a set of introduction and elimination schemas. In each schema the assumptions of the argument are above the horizontal bar and the conclusion (claim) below. For instance the and-introduction schema (∧I) can be read as: if we have some level of confidence Q (say +) in a claim φ and confidence Q′ (say ++) in ψ, then confidence in the conjunction of φ and ψ is the *minimum* of Q and

Q′ (i.e., + as the "weakest link" in the argument). The or-elimination schema (∨E) should be read as: if by assuming φ is the case in the current situation we can conclude σ is the case with confidence Q′, and by assuming ψ is the case then σ is the case with confidence Q″, and we know either $\varphi \lor \psi$ with confidence Q, then we can conclude σ with a level of confidence that is the minimum of Q, Q′ and Q″.[15]

The usual interpretation of classical logic includes the axiom of the "excluded middle". In probability theory we have an analogous constraint that if there is evidence that increases the probability of hypothesis H then this necessarily implies a reduction in the probability of ¬H. In the version of LA in Table 2 the not-introduction schema serves an analogous function. If we assume that some claim φ is the case, but this leads to a contradiction with some confidence Q then this entails the existence of an argument with confidence Q in the complementary claim ¬φ. In the clinical systems described above this axiom is not enforced. This is because when eliciting knowledge from clinicians there is a strong tendency to describe positive associations between symptoms and diseases, while symptom absence is not interpreted. Such systems are therefore more similar in spirit to systems based on intuitionistic logic, which also excludes the EM axiom, than classical logic though formalisations of LA like that above can have standard axioms and theorems.

Just as LA does not impose a unique set of axiom schemas the argumentation approach does not commit us to a particular representation of confidence. The representation in Table 2 uses just the qualifiers {+, ++}. However, in many real world applications it is natural to extend this set of symbols, to include the symbols {−, −−} where – indicates that the argument "opposes" a claim and – "rebuts" it [5].

Quantitative uncertainty representations are also possible. For example, qualifiers can be integers, as in the RAGs genetics risk assessment system described above, where each domain rule had an integer from {1,2,3} to represent low, medium and high confidence in the argument. Alternatively we may take a more orthodox approach, attaching a value from the [0, 1] interval, which could be taken to represent a conditional probability of the truth of the claim conditioned on the "evidence" embodied in the premises of the argument. Simon Parsons and I have argued that under appropriate technical constraints an argumentation proof procedure can be constructed that is equivalent to a standard Bayesian decision procedure [6].

Another important observation is that there is no objection to the use of "linguistic" qualifiers in an argumentation system if we can give a clear meaning to such terms. Toulmin's example used the qualifier "presumably" though he relies on our intuitions as English-speaking readers rather than providing us with a definition of its meaning. Krause and Clark [16] discuss a range of more formal proposals for linguistic qualifiers and Elvang-Gøransson et al. [2,3] formalize a set of "linguistic" qualifiers based on the consistency relationships between arguments for and/or against a claim. We return to the subject of linguistic qualifiers in a moment.

LA is a well-defined system on which to base a mechanical process for generating arguments, using a knowledge base formalised as a theory in an appropriately extended

---

[15] I am indebted to Paul Krause for providing this formulation of LA.

first-order logic. Krause et al. [17] describe a theorem-prover that implements the logic in Table 2.

1.2.3. Argument aggregation and epistemic states

We saw from the informal presentation of argumentation (Fig. 6) that the production of arguments is only part of an argument-based reasoning system. By *aggregating* over collections of arguments we can induce an ordering over a competing set of claims. In the situation where arguments have attached confidence coefficients, such as *ad hoc* or probabilistic weights, we can apply an appropriate aggregation function to yield quantitative degrees of confidence in the competing claims [16]. Bayes' rule for revising probabilities in the light of evidence for competing hypotheses is one such aggregation function, the expected utility model advocated by Lindley is a further extension, and there are many other possible aggregation methods.

Other aggregation functions can be employed when there is no information about the absolute or relative weights attached to arguments. Since we can construct arguments that are purely symbolic under the rules of LA (e.g., arguments that increase confidence in a claim but do not indicate by how much) the aggregation process is less obvious. However, suppose we have three supporting (+) arguments for one claim, two supporting arguments for another, and just one argument for a third then there is an obvious natural order over the claims *even though we have provided no quantitative interpretation of the level of support provided by each argument*. This aggregation is justified by the "principle of insufficient reason" introduced by Keynes, which says that if an agent has no reason to assign different levels of confidence to competing hypotheses it may reasonably assign equal levels of confidence. A minor variant of this principle warrants assigning equal levels of confidence to all arguments if there is no reason to do otherwise.

A general model for aggregation is

$$\big\{(\text{Claim} : \text{Warrant} : \text{Qualifier})\big\} \rightarrow (\text{Claim} : \text{Commitment}) \qquad (2)$$

Which defines a mapping from the set of arguments for/against a claim into a "commitment" about the claim. In this model arguments are tentative and revisable, while commitments are states of knowledge that are *entrenched,* which is to say the agent may be unwilling to give them up in the face of counter-arguments. Such entrenchment seems irrational but may in fact be quite understandable in real-world settings. For example an agent may deny or argue against counter-evidence for its claim in preference to accepting it which could incur a risk of being held liable for consequences of the "error" or exposure to charges of incompetence.

1.2.4. Annotation of arguments and the concept of belief

Aggregation is not the only function that can map from collections of arguments onto epistemic states. Another is called "annotation" because it enhances an agent's ability to describe its current view or representation of its beliefs.

As natural language users we routinely use annotations. We do not just say "I'm starting the flu" but also "it is {conceivable, possible, likely, pretty certain, . . . } that I'm starting the flu". Your doctor does not just say "you have got athlete's foot" s/he will just as often say "I {assume, suspect, doubt, believe, am certain/uncertain} that you {have, may have, could

have} the flu" and so on. This vocabulary is augmented by lexical and affixal negation (e.g. not possible, impossible), hedges (e.g., {quite, very, highly} plausible ...), elaborations (e.g., "it appears to be the case that ...", "there are persuasive reasons to think ...") and many other everyday constructions.

The existence of this large sub-language of English and other natural languages has long puzzled linguists, psychologists, philosophers and others. Logicians view some of the terms as *modal operators*, while others see them as mere stylistic decorations, of no formal significance. Social scientists may explain them in terms of conversational conventions, signalling a particular emphasis or serving an auxiliary social or pragmatic function.

I would add a further possible function for the many linguistic terms that we use to talk about our beliefs. If I say "its possible that (I'm starting a cold)" I intend to communicate something like "I have grounds to think I may be getting a cold (such as a sore throat, sneezing etc.) but there is also at least some reason to believe that I may not (e.g., (1) I already had a cold just last week; (2) I had the same symptoms yesterday but nothing happened, (3) my throat is only *slightly* sore)". In everyday communication we have not the time to go into detail so if our language allows us to communicate our epistemic state with a built in summary of the reasons for that state so much the better. In short, I suspect (sic!) that our large lexicon of uncertainty terms has a direct communicative function.

We see exactly this kind of phenomenon in medical and other guidelines. For example a guideline published by the International Agency for Research on Cancer sets out a standard terminology for talking about categories of risk associated with chemical compounds: a carcinogen is *confirmed* if there is epidemiological data and/or an established causal relationship between cancer and the compound; *possible* if a potential hazard has been recognised; *probable* if there is better evidence than merely recognition of possible carcinogenic activity; *improbable* if there is possible carcinogenic activity, but strong evidence against), and the risk is *equivocal* if a hazard recognised and there is evidence both for and evidence against.

This might be nothing more than a minor linguistic observation, with no formal interest except that it is possible to develop a logical system based on this idea. Elvang-Gøransson et al. [3] define a set of logical "acceptability classes" for talking about belief in some proposition P based solely on the logical properties of the set of arguments for and against P. These acceptability classes define a completely ordered set of predicates for expressing confidence in P from low to high:

P is open

if it is any well-formed formula in the language of the logic

P is supported

if an argument, possibly using inconsistent data, can be constructed

P is plausible

if a consistent argument can be constructed (we may also be able to construct a consistent argument against)

P is probable

if a consistent argument can be constructed for it, and no consistent argument can be constructed against it.

   P is confirmed

if it satisfies the conditions of being probable and, in addition, no consistent arguments can be constructed against any of the premises used in its supporting argument

   P is certain

if it is a tautology of the logic (meaning that its validity is not contingent on any data in the knowledge base).

In the foregoing we have explored a wide range of phenomena concerned with ideas about uncertainty and belief, and suggested that these and other notions like probability, possibility can be derived by analysing patterns of argument. A common view of such terms in logic is to treat them as modals, basing their semantics on a possible-world formulation. This is a well-developed position but expensive in the sense that modal logics have well known difficulties compared with standard predicate logic, and proving such logics as mathematically sound and complete is time-consuming and demanding. This may be an important thing to do if you are a logician, but not if you are trying to understand the use of modals in natural language or trying to build an agent that can use modal terms in its reasoning and user interface

My position is that these so-called modal terms are "epistemic states", predicates that evaluate properties of *collections of arguments* about claims. The use of epistemic states may be mathematically *ad hoc* yet they provide a useful vocabulary for an agent to *reflect* upon its beliefs and their justification, and *communicate* succinctly with other agents about the confidence to be attached to its claims. In this respect we follow John McCarthy in his article *"Modality, Si! Modal logic, No!"* where he remarks "Human practice introduces new modalities on an *ad hoc* basis. . . . Introducing new modalities should involve no more fuss than introducing a new predicate. In particular, human-level AI requires that programs be able to introduce modalities when this is appropriate" [19].

Mathematicians and logicians are rightly suspicious of anything that is *ad hoc,* by which they mean unprincipled, particularly where there is already some well established alternative like modal logic. An implication of Elvang-Gørannson's work is that despite appearances we can construct a sound and useful set of logical predicates for talking about beliefs and other cognitive states whose semantics can be put on a sound footing.

1.3. Summary and conclusions

The modern concept of mathematical probability is widely taken to be the normative standard for reasoning under uncertainty and the formulation of rational beliefs. However this position has often been questioned on a variety of grounds and the issues never seem to be quite settled. The rise of the cognitive sciences in the last quarter century has led to an explosion of new questions about the sufficiency of probability theory for dealing with problems in artificial intelligence, psychology, linguistics and other areas concerned with the nature of cognitive agents. Three lines of argument (historical, practical and anthropic) have been presented which support the view that traditional formalisations of probability

leave important questions about uncertainty and belief unresolved. An account of knowledge and decision-making based on logical *argumentation* has been presented that appears to unify diverse intuitions about the nature of probability and belief. These discussions are substantiated by examples from clinical medicine. However, there are few truly new ideas in this field and previous discussions by Keynes, Hacking, Toulmin, Simon and others suggest that the problems and proposals discussed here are not specific to the medical domain. The central claim is that some general form of argumentation underpins all the competing ideas about probability, and developments in artificial intelligence and non-classical logics suggest ways of formalising this class of system in order to resolve some of the epistemological and technical issues that they were not in a position to address.

References

[1] H. Abramson, M.H. Rogers (Eds.), Meta-Programming in Logic Programming, MIT Press, Cambridge, MA, 1989.

[2] M. Elvang-Gøransson, P.J. Krause, J. Fox, Dialectic reasoning with inconsistent information, in: D. Heckerman, A. Mandani (Eds.), Uncertainty in Artificial Intelligence. Proceedings of the 9th Conference, Morgan Kaufman, San Mateo, CA, 1993.

[3] M. Elvang-Goransson, P.J. Krause, J. Fox, Acceptability of arguments as logical uncertainty, in: M. Clarke, R. Kruse, S. Moral (Eds.), Symbolic and Quantitative Approaches to Reasoning and Uncertainty; Proc. European Conference ECSQUARU '93, Springer, Berlin, 1993, pp. 79–84.

[4] J. Emery, R. Walton, M. Murphy, J. Austoker, P. Yudkin, C. Chapman, A. Coulson, D. Glasspool, J. Fox, Computer support for interpreting family histories of breast and ovarian cancer in primary care: Comparative study with simulated cases, British Medical J. 321 (2000) 28–32.

[5] J. Fox, P.J. Krause, M. Elvang-Goransson, Argumentation as a general framework for uncertain reasoning, in: Proc. Int. Conf. On Uncertainty in Artificial Intelligence, Morgan Kaufman, Washington, 1993, p. 428.

[6] J. Fox, S. Parsons, Arguing about beliefs and Actions, in: A. Hunter, S. Parsons (Eds.), Applications of Uncertainty Formalisms, Springer, Berlin, 1998.

[7] J. Fox, S. Das, in: Safe and Sound: Artificial Intelligence in Hazardous Applications, Cambridge MA, MIT Press, 2000.

[8] J. Fox, M. Beveridge, D. Glasspool, Understanding intelligent agents: Analysis and synthesis, AI Communications 16 (2003) 139–152.

[9] J. Fox, D. Glasspool, J. Bury, Quantitative and qualitative approaches to reasoning under uncertainty in medical decision making, in: S. Quaglini, Pbarahona, S. Andreassen (Eds.), Proc. AI in Medicine Europe, Springer, Heidelberg, 2001, pp. 272–282.

[10] M. Georgeff, B. Pell, M. Pollack, M. Tambe, M. Wooldridge, The Belief-Desire-Intention model of agency, in: Proceedings of Agents, Theories, Architectures and Languages (ATAL), 1999.

[11] G. Gigerenzer, P.M. Todd, ABC Research Group Staff, Simple Heuristics That Make us Smart, Oxford University Press, Oxford, 1999.

[12] I. Hacking, The Emergence of Probability, Cambridge University Press, Cambridge, 1975.

[13] A. Hunter, S. Parsons, Applications of Uncertainty Formalisms, in: Lecture Notes in Comput. Sci., Vol. 1455, Springer, Berlin, 1998.

[14] I. Johnston, . . . And Still We Evolve: A Handbook on the History of Modern Science, Chapter 4. The Beginnings of Modern Probability Theory, 2000, http://www.mala.bc.ca/~johnstoi/darwin/sect4.htm.

[15] J.M. Keynes, Treatise on Probability, Macmillan and Cambridge University Press, London, 1921.

[16] P. Krause, D. Clark, Representing Uncertain Knowledge: an Artificial Intelligence Approach, Intellect Books, Oxford, 1993.

[17] P.J. Krause, S. Ambler, J. Fox, A logic of argumentation for uncertain reasoning, Comput. Intelligence 11 (1) (1995) 113–131.

[18] H. Kyburg, Science and Reason, Oxford University Press, Oxford, 1990.

[19] J. McCarthy, Modality Si, Modal logic no, Stanford University Department of Computer Science, 1997, http://www.formal.stanford.edu/jmc/modality/modality.html.

[20] A.S. Rao, M.P. Georgeff, Modelling rational agents within a BDI-architecture, in: R. Fikes, E. Sandewall (Eds.), Proceedings of Knowledge Representation and Reasoning, KR&R '91, Morgan Kaufmann, San Mateo, CA, 1991, pp. 473–484.

[21] S. Parsons, Qualitative Methods for Reasoning under Uncertainty, MIT Press, Cambridge, MA, 2000.

[22] M. Risse, Review of D. Corfield, J. Williamson (Eds.), Foundations of Bayesianism, 2002, at http://ksghome.harvard.edu/~.mrisse.academic.ksg/Papers/Papers%20-%20Decision%20Theory/FoundationsOf Bayesianism.pdf.

[23] S.J. Russell, P. Norvig, Artificial Intelligence: A Modern Approach, 1st Edition, Prentice-Hall, Englewood Cliffs, NJ, 1995.

[24] M. Smithson, Ignorance and Uncertainty, Springer, Berlin, 1989.

[25] P. Taylor, J. Fox, A.T. Pokropek, The development and evaluation of CADMIUM: A prototype system to assist in the interpretation of mammograms, Medical Image Anal. 3 (4) (1999) 321–337.

[26] S. Toulmin, The Uses of Argument, Cambridge University Press, Cambridge, 1957.

[27] R. Walton, C. Gierl, P. Yudkin, H. Mistry, M.P. Vessey, J. Fox, Effects of decision support on prescribing in general practice, British Medical J. 26 (1997).

[28] M. Wooldridge, N.R. Jennings, Intelligent agents: Theory and practice, Knowledge Engrg. Rev. 10 (2) (1995) 115–152.

[29] D. Kahneman, P. Slovie, A. Tversky (Eds.), Judgment under uncertainty: Heuristics and Biases, Cambridge University Press, New York, 1982.