Group II

DECISION SUPPORT FOR PRACTICAL REASONING:

a theoretical and computational perspective

Rod Girle, David Hitchcock, Peter McBurney, Bart Verheij

CONTENTS

1 Introduction

When faced with difficult decisions about what to do, decision makers benefit from good advice. Good advice comes most reliably from advisors with relevant expertise. As well, good advice has at least three other essential features. First, the advice should be presented in a form which can be readily understood by the decision maker. Second, there should be ready access to both the information and the thinking that underpins the advice. Third, if decision making involves details which are at all unusual, the decision maker needs to be able to discuss those details with their advisors.

Computer based systems are being increasingly used to assist people in decision making. Such systems are known as *decision support systems*. As well as the appropriate expertise, it is vital that decision support systems have the three features above. As for the first, it has been pointed out that, “In order for any beings, human or machine, to talk to each other, they must share the same mental structures.”[1] Decision making is field dependent. Advice must be presented in the appropriate conceptual framework. Decision makers need advice to be given in terms which come from the decision making domain. This requirement leads to the second feature, that the basis for advice should be clear. The basis includes not only the information drawn on by an advisor, but also the reasoning which leads to the specific advice about what to do. Reasoning about what to do is known as *practical reasoning*. The third feature draws the decision maker into the process, so the advice is not “over against” the decision maker, authoritarian rather than authoritative, final rather than part of a process. There needs to be the possibility of joint *deliberation* about what should be done. If the decision maker is involved, by supplying “local” information, questioning the rationale behind advice, and discovering the deeper reasons behind the advisor’s recommendations, then the final decision will be more considered and more likely to be correct.

In this chapter we propose a theoretical basis and make general recommendations for the design of decision support systems which have the three features mentioned. In particular, we discuss the nature of practical reasoning, the deliberative interaction between advisor and decision maker, and the related questions concerning sensible computer system implementation. There is considerable expertise about practical reasoning in the general area of argumentation theory. Work in artificial intelligence will indicate the limits and possibilities for implementation. We draw upon work in both argumentation theory and artificial intelligence.

In the next section of this chapter we explore the rich and complex area of practical reasoning, of the agents who do things, and of the variety of domains in which decisions are taken. In the third section we describe the variety of schemes for rational argument, of demonstrative and dialogical argument, and of the contexts of argument.

The fourth section is devoted to considerations about the actual processing of material in producing advice. In the fifth section we consider the resource constraints under which such processing is performed. These two sections, fourth and fifth, bring us face to face with the reality of processing in actual machines. The sixth section contains proposals for a wide ranging research programme about how one might integrate moral considerations into advice.

In the seventh section we argue for deliberative interaction between system and user or users. There is a proposal about how deliberative interaction could be modelled by formal deliberation dialogue. Overall, such deliberation would change the advice coming from the decision support system, acknowledge the autonomy of the user, and facilitate deliberation amongst a group of users.

The eighth section sets out principles which should guide the building of an interactive interface for any decision support system for practical reasoning. The concern is for high-level principles rather than for details about computer screens, speech recognition, or

---

[1] Michie, D. and Johnston, R. 1985. *The Creative Computer*, Penguin Books, Harmondsworth, pg. 72

virtual reality interfaces. The ninth section contains proposals for systematic feedback about the effectiveness of the decision support system. This can be seen both as a way of revising the operation of the system to improve it, and as a way of facilitating research into argumentation-based decision support systems. A summary follows the ninth section.

2 Practical Reasoning

Practical reasoning is reasoning about what is to be ***done***. Doing something includes as the most elementary case (1) simple physical actions such as raising an arm. More complex cases are (2) a series of simple actions and (3) adoption of an intention to initiate a sequence of actions later (a “plan”, which may be only partially elaborated at first). Plans may be logically complex, including for example, disjunctions or conditions. (4) More general than a plan is a policy, which is the carrying out of a certain type of plan whenever specified conditions obtain (e.g. whenever I want to walk across a street, do so only when it is safe). More complex still are (5) cases where the agent is not an individual human being but an organization of human beings–an academic department, a municipal government, etc.

Actions of all these types include intentional omissions, i.e. deliberately not (now or later) undertaking a certain physical action, defeating a resolution to undertake some initiative, etc. Thus, generically, practical reasoning is reasoning directed to the adoption of a policy by some agent, where policies include as limiting cases plans (policies adopted for just one occasion) and actions (plans with only one component), and the agent adopting the policy may or may not be identical to the set of individuals carrying out the reasoning.

Philosophers are interested in practical reasoning from two points of view: explanation and guidance; for the distinction, see Raz’s introduction to his (1978). The explanatory interest leads them to consider such questions as whether a belief-desire-intention model is the correct explanation of intentional action, whether reasons are causes, how *akrasia*[2] is possible, what is the difference between *akrasia* and (possibly self-deceptive) hypocrisy, and whether all reasoning-produced motivation is partly derivative from motivation already present in the reasoner ("motivational internalism"). Although the investigation of questions to do with explanation is sometimes relevant to questions related to guidance, and indeed it is sometimes difficult to tell whether a given philosophical contribution is oriented to explanation or guidance, we will focus on guidance-oriented questions–i.e. those which bear directly upon the general issue of how an individual or group might, if it were rational, arrive at and implement decisions about what to do.

A comprehensive guidance system for human action would assist any human being or group of human beings, given any situation, to decide what to do in any given respect at any time, assuming possession of the required factual knowledge and other inputs. Such a system would fall under the general category of a decision support system. The system we would be interested in describing would be one in which the deep structure was based on argumentation and developed on the basis of argumentation theory.

Practical reasoning is often seen as domain-dependent reasoning. A decision support system for any given domain would have to take account of salient features of the domain in which the reasoning takes place. A completely general system would, therefore, have to "model the world." Such modelling is beyond our remit. We would require, rather, that our decision support system should give expression to high-quality practical reasoning in each of the domains to which it is applied.

There are so many ways to engage in practical reasoning, and they interact in such complex ways, that the task of formulating a comprehensive system is daunting. The underlying architecture for good practical reasoning will in general be complex (in the ordinary sense which contrasts with "simple," not in the computer science sense of having a certain degree of difficulty).

---

[2] "weakness of will"

Good practical reasoning is complex with respect to the argument schemes it can use. Only in limited and well-defined domains of decision-making does it make sense to use a single scheme of practical reasoning. An example of such a limited, well-defined domain is a physician's decision as to which drug to prescribe for a given condition. The factors relevant to such a decision are known, finite and in fact few: efficacy, side-effects, interaction with other drugs being taken by the patient, contra-indications, patient's past experience with the drug, cost, recommendations by authorities, local policy, whether a drug is proprietary or generic (Fox & Das 2000: 40). Given that only a small number of drugs are possible candidates for a given condition, it is possible to list the pros and cons for each candidate drug with respect to each relevant factor, to arrange the candidates in a hierarchy of decreasing net number of supporting considerations, and to present this information to the prescribing physician. Contrast such a well-defined domain, in which the relevant considerations are few and well-known, to decisions in less well-defined domains, such as the decision of a high school student on what to do immediately after graduation, a government decision on macroeconomic policy, or deciding how to manage global temperature. There is no known limit to the number of considerations that could be relevant, positively or negatively, to an individual or group making such decisions. The goals in such a decision-making situation may be multiple and even conflicting. Further, some of the goals may be of questionable "validity," requiring scrutiny to see if they are rationally justified, for example in terms of how well those affected by the decision will like the situation which results if the goal is achieved: not everybody who gets what they want likes what they get. Further, the goals may be unclear or incomplete, requiring clarification or elicitation. There may be incomplete knowledge of the outcomes of the various options under consideration, a situation which Hansson (1996) refers to as "delimitation uncertainty" and which is sometimes referred to in artificial intelligence as "possibilistic risk." (Krause *et al.* 1998) And so on.

Some idea of the complexity required for a comprehensive system for rational guidance for human decision-making can be gathered from John L. Pollock's computational architecture for an autonomous rational agent, which he dubs OSCAR (Pollock 1995, 1999). Pollock's model of practical rationality assumes a much richer psychology than the belief-desire psychology which he traces back to David Hume.[3] Pollock argues (1995: 12-35) that practical reasoning, understood as having the function of making the world more to its possessor's liking, requires seven distinct types of states: beliefs, situation-likings, feature-likings, intentions (construed as adoptions of plans), and three kinds of desires (primitive, instrumental, present-tense action). *Token situation-likings* are feelings about how good the agent's present situation is; they are the basic starting-point for working out rationally how to make the world more likable for the agent. The agent also needs to know what features or combinations of features are causally relevant to situations' being liked or disliked. Although such knowledge could in principle be acquired as a result of empirical investigation, time and resource constraints dictate quicker methods, which in human beings are provided by the ability to introspect whether they like or dislike an imagined situation; since such an imagined situation is a type rather than a token; such (dis)likings are *feature-(dis)likings*. In order to focus its planning activities, a rational agent needs to adopt goals whose achievement it thinks will make its situation more likable; such an adopted goal is a *primitive desire*. Humans also have primitive desires from built-in and conditioned optative dispositions, e.g. the disposition to want to eat when one feels hungry. Adoption of goals can trigger planning, which can result in the adoption of a plan, encoded as an *intention*. Since such adopted plans are often partial, further planning can be required to execute them; such planning takes its start from components of the partial plan which the agent must work out how to realize, for which the agent has an *instrumental desire*. Finally, there must be some state which actually initiates an action at a certain time, whether or not this action is part of a previously adopted plan; such states are *present-tense action desires*. (A computational support system for decisions will not need such present-tense action desires, since it will not actually implement plans it proposes.) Practical reasoning involving these states must obviously appeal to *belief*s about what is the case. Thus the belief-desire-intention model of practical reasoning (Wooldridge 2000) is overly simple as a general account of good practical reasoning, although its ease of computational implementation makes it acceptable in applications where intentions and desires do not need rational scrutiny.

Pollock's psychology is the basis of a complex and subtle computational architecture for a rational agent, for whose construction much sophisticated thinking and refinement of originally attractive but over-simple ideas was required. Complex as it is, Pollock's OSCAR is incomplete in at least three important respects. First, it is solipsistic, in the sense that there is no provision for verbal input from, or verbal output to, other autonomous rational agents, still less for back-and-forth discussion, whether argumentative or non-argumentative. Second, it is egoistic, in that the function of the entire system is to make the world more to the liking of that system itself, without regard (except instrumentally) to whether its actions make the world more or less to the liking of other systems which have situation-likings and situation-dislikings; morally speaking, Pollock's OSCAR is a monster. Third, it is unsocial, in that it does not (and cannot) belong to any group of autonomous rational agents with governance structures for making decisions about the actions of the group; it is a citizen of no country, belongs to no professional associations, owns no shares in any joint-stock company, has no immediate family, does not belong to a recreational bridge-playing group, etc. A really comprehensive system of rational guidance for human action would have to remedy all three of these lacks. In particular, it would have to include decision support systems for multi-agent decision-making.

Good practical reasoning is complex in another respect, namely with respect to the

---

[3] see Pollock 1995: 33

number of types of argumentation schemes which are directly constitutive of it. Walton (1996), for example, lists 15 argumentation schemes which always or sometimes involve practical reasoning to a conclusion about what is to be done. An additional source of complexity in the application of these schemes is that they give only defeasible support for a course of action whose wisdom is always open to rebuttal on the basis of further information. We discuss these and other argument schemes in the next section.

Further, practical reasoning typically includes as subordinate argumentation a great deal of epistemic reasoning, directed at determining what to believe. For example, deciding in an environmental risk assessment whether a proposed expressway should be built, and if so under what constraints, will require determining many facts about the present state of the area through which the expressway is to be built, about the likely consequences of building it, and of the way in which those consequences would be different if various changes were made to the expressway design. In fact, Pollock claims that “the epistemic reasoning required by practical reasoning is probably much more complicated than the structure of the practical reasoning itself.” (1995: 265) Like Pollock, we will ignore in this chapter all the difficult questions about epistemic reasoning which need to be answered as part of the computational implementation of a general theory of good practical reasoning, and will simply note that they too must be part of the wider, relevant research agenda.

The complexity and open-endedness of good practical reasoning are a powerful reason for restricting computational applications to decision support rather than decision-making, and for building into such computational applications user interfaces which display in an understandable way the reasoning by which the program has arrived at its recommendations, allowing the user to “second-guess” and even alter the program. The need for informed and experienced judgement in many situations of practical reasoning is another reason pointing in the same direction.

3 Argument Schemes and Defeasibility

There are patterns or schemes of argument which occur frequently in practical reasoning. We saw earlier that Walton (1996), among others,[4] has proposed a list of 15 such schemes. They are said to be presumptively valid, giving defeasible support for courses of action. A central question in the study of practical reasoning is whether these, or any other patterns can be considered as 'practically valid'. One could say that a reasoning pattern is considered to be practically valid if its application legitimately leads to plans and intentions to act, and thus can be legitimately used to guide one's behavior.[5]

In most theories of practical reasoning, it is assumed that such practically valid reasoning patterns exist. (The one extreme exception is nihilism about practical reasoning, according to which such patterns do not exist, and neither does practical reasoning. Cf. Milgram 2000) For example, one type of reasoning that is often accepted is means-end reasoning, in which plans and intentions to act are selected because they serve one's goals. For instance, the intention to go out for a walk might be justified by the fact that it can fulfill one's desire to get some fresh air. Some practical reasoning is what one might call "specification reasoning," where one has a justified partial plan (e.g. to eat lunch) and one specifies it (by deciding to have a ham and cheese on rye); Richardson (1995) has explained in detail how a rational agent can use such reasoning in deliberating even about a final end, i.e. a goal which the agent is pursuing for its own sake and not for the sake of anything else.

Another argument scheme for practical reasoning subsumes a particular case under a general principle, e.g. reasoning that I ought not to cross the road right now because it is not safe to do so. Arguments from consequences reject an option on the ground that its

---

[4] For other such lists see also Perelman & Olbrechts-Tyteca 1971/1958, Kienpointner 1992, and Grennan 1997.
[5] Obviously there are a lot of strings attached to the notion of practical validity, as defined here. We hope that the definition here suffices as a first indication. Below, when the idea of defeasibility is discussed and in the section about deliberation and dialogue, we briefly return to the notion of practical validity.

consequences are unacceptable. In various contexts, reasoning about what to do may also involve reasoning from a verbal classification, from commitment, from analogy, from precedent, from expert opinion, and so forth

One argumentation scheme concerns the weighing of the pros and cons with respect to a certain plan or intention. Though this scheme is particularly relevant for practical reasoning, since it often is the case that a plan or intention has both favourable and unfavourable aspects, it is occasionally overlooked (e.g., by both Walton 1996, and Grennan 1997). Naess's (1966/1947) work on pro-et-contra and pro-aut-contra schemes is an early example of an analysis of this kind of reasoning, though Naess is not particularly dealing with practical reasoning. Wellman (1971) also analyzes this type of reasoning, and uses the term *conductive reasoning* for it, in opposition to deductive and inductive reasoning (cf. also section 5 below). Govier (1987, 1999) has further analyzed the notion of conductive reasoning. In these analyses, weighing is taken mostly as a *qualitative* primitive. There is normally no calculus, e.g., in terms of numbers or weights, that determines the result of weighing. An example from a logically styled analysis of the weighing of reasons is Reason-Based Logic, as initiated by Hage (see, e.g., his 1997) and further developed in cooperation with Verheij (see, e.g., his 1996). In Reason-Based Logic, the statement that certain pros outweigh certain cons is treated on a par with all other statements, and can be derivable from a given set of premises. The only calculus built into the system has for instance the effect that adding pros to a set of pros that already outweigh a fixed set of cons, does not change the result of weighing. Analyses of legal case-based reasoning (e.g., Ashley 1990) also contain ideas that are related to the weighing of reasons.

Two points are striking when one consults the lists of argumentation schemes, as they have been proposed in the literature. The first is that the lists have a rather *ad hoc* character. For instance, in Walton's book (1996), there are schemes dealing with a great variety of relevant topics, such as the consequences, one’s commitment, expert opinion, analogy, and precedent. However, the *ad hoc*-ness of the lists of argumentation schemes might be only apparent. What one should look for is a principled basis for devising and testing the lists. Such a principled basis can be found in criteria like empirical adequacy and expressiveness. According to the criterion of empirical adequacy, the schemes should make explicit how practical reasoning goes on in a specific domain, like the law or medicine. The criterion of expressiveness can be used to select argumentation schemes in terms of which other schemes can be expressed. A braver approach, and one with more fundamental consequences, would be to consider the selection of the argumentation schemes to be itself a topic of practical reasoning: in order to determine what is to be done, one should also determine which argumentation schemes can be used under the circumstances of the situation at hand (cf. also McBurney & Parsons 2000). For instance, the maximal expected utility approach to decision-making is sometimes considered inappropriate for modeling risk averse decision making: an investment of 1000 Euros that gives a 1% chance on a return of a million Euros is supported by the maximal expected utility approach, but may not be wanted if one does not accept the possible consequence of losing the investment. For another example, in the law, it can occur that the parties disagree about the use of an argument scheme, such as argument by analogy, which in some jurisdictions is generally not allowed in criminal cases. The debate may then focus on the acceptability of this argument scheme.

This brave approach to the selection of argumentation schemes would require a rethinking of the basics of logic. It challenges the often uncritically assumed primacy of logical semantics that serves as an external criterion to determine the validity of argumentation schemes in terms of truth preservation. Current orthodoxy in classical logic holds that a form of argument is valid if and only if it never leads from true premises to a false conclusion. If one accepts that the validity of argumentation schemes depends at least in part on the context (as suggested by the context dependent lists of argumentation schemes), such an external criterion does not exist and logical semantics loses its primary role (cf. in this respect Verheij 1999c). Moreover it is commonly thought that the truth preservation approach does not work for recommendations. For instance, the suggestions given by the sort of guidance system we are considering may not even have a truth value in the classical sense.

This brings us to the second striking point concerning argumentation schemes for practical reasoning: they are normally only *presumptively* (or *defeasibly*) valid, in the sense that the schemes do not always or directly lead to their conclusion when the premises obtain. They merely lead to the presumption that the conclusion of the scheme obtains when the premises obtain. This presumption can be defeated in case there are certain exceptional circumstances. The critical questions that can be asked with respect to argumentation schemes, as they occur in the argumentation literature do not only help to determine whether the conditions of the scheme are met, but also can be regarded as ways to establish such exceptional circumstances.[6] For instance, critical questions that can be asked with respect to the scheme according to which an action should be undertaken in case an expert advises one to do so, are for instance: did the expert lack any relevant knowledge, and are there experts with different, contradictory opinions? In both cases, a positive answer indicates an exceptional circumstance that means that the presumptively valid scheme should not be applied. Clearly, the circumstances can give rise to different ‘standards’ for answering the critical questions. For instance, when asking the way to the train station in Glasgow, one’s critical standards will differ from those when looking for medical advice concerning the

[6] There are two ways of thinking about argumentation schemes and the role of critical questions. The first is that argumentation schemes are types of reasoning, of which the presumptive validity in a particular situation is subsequently determined by answering some of the critical questions. On this descriptive conception, argumentation schemes are not necessarily presumptively valid, since the presumption that the conclusion is to be accepted follows only after establishing the answers to some of the critical questions (which ones being a function of the particular situation). The second way, adopted above, has it that argumentation schemes are indeed presumptively valid, but the presumptive validity can be defeated under exceptional circumstances, e.g. those discoverable by answering the critical questions. Whether one chooses the first descriptive conception or the second normative conception may not have computational implications.

simultaneous use of possibly interacting drugs.

The idea of presumptive validity is in stark contrast with logic in its classical forms. For instance, the classical deduction rules, like *Modus ponens*, are presented as unconditionally valid. One might conclude from this that computers - since they are 'logical' machines - are not built for the use of presumptively valid argumentation schemes. However, though presumptively valid schemes are computationally more difficult to compute than the schemes of classical logic, artificial intelligence has provided several approaches that enable reasoning on the basis of presumptively valid argumentation schemes. Relevant notions are for instance the defeasible rules of inference, as they have been formalized by Reiter (1980) in his non-monotonic logic of defaults rules (cf. also Gabbay *et al.* 1994), and the more recently developed logics of defeasible argumentation, in which arguments can be defeated when they are attacked by counterarguments (see, e.g., the overviews by Prakken & Vreeswijk, *to appear*, Loui *et al.*, *to appear*). In defeasible argumentation, arguments do not under all circumstances justify their conclusion, but can be defeated by counterarguments. Here a game-theoretical approach has proven useful: an argument can be considered as justifying when the proponent of the argument has a winning strategy against attacks by the opponent (cf. in a different connection Hintikka & Sandu 1997 on logical semantics in terms of games).

The distinction between classical validity and presumptive validity can be made clearer by Pollock's (1995: 199) contrast between *warranted* as opposed to *justified* conclusions. A conclusion (of practical or epistemic reasoning) is justified if and only if the reasoner has reasoned correctly up to the time of reaching it. A conclusion is warranted if and only if reasoning which could proceed without limit would reach a point where the conclusion was justified and would never subsequently become unjustified just as a result of further reasoning. The terminology is a useful way of marking a distinction which is important for systems (like human beings and computers) with limited resources for reasoning (cf. also section 6).

At all stages of computation, a system's current conclusions are justified, in Pollock's sense. Only when (and in case of limited resources: if) computation has been completed, the system’s conclusions would be warranted. Limited resources may imply that warranted conclusions are never reached. An example can be found in automated chess. Computers can find good chess moves by searching large parts of the space of possible moves. A search of the whole space would result in finding the best possible moves. Those moves would be warranted in Pollock's sense. Due to the enormous size of the space of chess moves, it is impossible to search all of it within a reasonable time. Therefore the moves found can only be justified in Pollock's sense. Continued search may always result in finding that another move is better. Interestingly, even the merely justified moves as found in today's computer chess, result in grandmaster level play.

In defeasible argumentation, the kinds of attacks between arguments that can lead to argument defeat have been discussed. For instance, Pollock (1987, 1995) claims that there are two kinds: rebutting and undercutting defeaters. A rebutting defeater is a defeating reason that pleads for a conclusion opposite to the conclusion supported by the reason it attacks. An undercutting defeater is a defeating reason that merely attacks the connection between the attacked reason and its conclusion. In Verheij’s (1996, 1999b) CumulA model, it is possible to distinguish types of defeat in terms of the structures of the arguments involved in an attack. For instance, Pollock’s rebutting and undercutting defeaters are *step-type defeaters*, since they involve single reason-conclusion steps in arguments. Verheij also distinguishes *sentence-type defeat*, in which a statement attacks another statement, and *composite-type defeat*, in which composite argument structures are involved. Examples are defeat by sequential weakening, according to which arguments can become defeated by containing an ever weakening series of steps, and defeat by parallel strengthening, according to which an argument is defeated by an argument for the opposite conclusion since the latter contains several accruing reasons for its conclusion. Defeat by sequential weakening can be used in order to analyze the sorites paradox, and defeat by parallel strengthening to analyze cases of accrual of reasons, where a group of reasons for a conclusion outweigh reasons against, while the reasons individually would not suffice. The discussion concerning types of defeat is not yet completed. Pollock (1995) has for instance argued against defeat by sequential weakening and by parallel strengthening.

4 Decision Calculi

A common form of reasoning about what to do is to "weigh up the pros and cons" of a projected policy. A comprehensive exercise in such "weighing" would take into account every consideration which is relevant, whether positively or negatively, to the decision to be made; the right decision would be the one which the various considerations on balance favour (Fox and Das 2000). The argumentation-based medical decision support systems developed at London's Imperial Cancer Research Fund (ICRF), for instance, typically provide a human decision-maker with a list of the arguments for and the arguments against each suggested course of action (Krause *et al.* 1995, Fox and Thomson 1998, Carbogim *et al.* 2000). One such system, CAPSULE, a drug prescription advice system, provides the doctor using it with a list of suggested drugs for each presented set of patient symptoms, along with arguments for and against the use of each drug. One drug may be more efficacious than another, but may cost more or may interact with other drugs being taken by the patient. The final decision as to which drug to prescribe is left to the doctor.

Arguments involving such reasoning have been variously called "good reasons" arguments, "balance-of-considerations" arguments, "pros-and-cons" arguments and "conductive" arguments. Such reasoning arises whenever no consideration or combination of considerations is decisive one way or the other – that is, when any accumulation of considerations supporting a certain decision is subject to rebuttal by citing further considerations pointing in the opposite direction.[7] We could define such reasoning as

[7] Conductive reasoning occurs in support of a variety of types of conclusions, not just decisions or recommendations to adopt a certain course of action or policy. In a recent defence against scepticism about the very existence of conductive arguments, Govier includes among 10 quoted examples (1999: 160-166) some whose conclusions are, or are construed as, causal claims (that rape is not due to natural psychological impulses, that punishment will render a criminal more morally sensitive, that the main beneficiaries of programs to combat global warming will be developing countries).

reasoning based on consideration of arguments for and arguments against each particular course of action, where both premises and conclusions of the arguments concern a single practical reasoning decision problem. Such arguments may include the reasons justifying or denying each action-option, their (positive or negative) consequences, and/or the (qualitative or quantitative) costs and benefits of acting or not acting according to each option. The process of consideration of the pertinent factors may be simple or sophisticated. The ICRF drug prescription system, for example, allows the doctor to evaluate the arguments listed by the system for each prescription option in whatever manner he or she desires. Likewise, an operationalization of conductive reasoning exemplified by Dutch tort law on liability in endangerment cases provides another example, involving a central case on endangerment, the so-called trapdoor case (Netherlands Supreme Court, November 5, 1965, NJ 1966, 136, in which someone fell into a café's basement because the trapdoor was left open by the supplier of soda drinks). The case lists a number of considerations that must be considered when deciding on the wrongfulness of endangerment. Among the considerations are the difficulty of taking precautionary measures, the proportions of the possible damages, and the chance that such damages occur. Though not prescribing how the resulting factors need to be weighed, the trapdoor case narrows the decision space that is open to the judge's discretion. The process of reasoning about the relevant considerations in order to reach a course of action may also be more complex, and we discuss such decision calculi later in this section.

There are many open questions about this form of reasoning. What does it mean to say that a consideration is relevant to a proposed policy? How does one discover a consideration which is relevant to a proposed policy? This question has received some attention from students of risk and decision. As mentioned in Section 2, for example, Hansson, in a typology of uncertainty in environmental risk decisions (Hansson 1995), called uncertainty arising from the lack of complete knowledge of possible outcomes "Delimitation

Uncertainty", and in the Artificial Intelligence community it is known as Possibilistic Risk Assessment (Krause *et al.* 1998, Fox 1999). Further, given a proposed consideration, how can one check whether it is relevant, positively or negatively?

Hitchcock has proposed a method of refutation by logical analogy of claims to relevance of a consideration in conductive reasoning (Hitchcock 1994). For example, if someone argues that a patient should not be told that he has terminal cancer, on the ground that telling him will upset him emotionally, one might object that on that reasoning no teacher should give a student a failing grade if doing so will upset the student emotionally. The person arguing for concealment of a terminal cancer condition could reply that there are external and generally accepted rules for determining when it is justified to give a student a failing grade, but there are no such rules for determining when to tell patients that they have terminal cancer.

These contrasting appeals to what is relevant show the difficulties with the method of refutation by logical analogy. Even though this method may be the best one can produce, it turns out to be quite difficult to show that any proposed consideration is irrelevant, since the defender of its relevance can always object to a parallel case where the consideration seems irrelevant that there is, in the parallel case, an overriding consideration with the opposite kind of relevance which explains our judgement about that case.

How can one tell that one has exhausted all the relevant considerations, or at least all those which are of sufficient "weight" to make a difference to one's overall judgement? How can one reconcile disagreements among different individuals as to whether a given consideration is relevant? A recent study in Britain of attitudes to Genetically-Modified foodstuffs, for example (Stirling and Mayer 1999) found irreconcilable differences in which issues different experts – all rational, knowledgeable, co-operative and well-intentioned - considered salient to public policy decisions on the issue, and in how much weight each issue should be given. How can we cash out the metaphor of "weighing" in a way which enables us to determine which policy the competing considerations "on balance" favour? Is this form of reasoning best modelled in terms of deciding whether or not to adopt a specified policy, or of choosing among a number of specified options? If the latter, how does one assign each consideration to each option? Stirling and Mayer (1999) use a quantitative method of scoring different options on different criteria and then weighting these according to an agreed relative weighting scheme, a method long used for multiple-criteria decision-making in business. But such quantitative scores and weights are, by their very nature, subjective and thus, in public policy domains, highly contested.

The literature in argumentation theory about conductive support is suspicious of any attempt to produce a calculus which could be applied so as to generate a judgment by some sort of quantitative reasoning. Benjamin Franklin proposed a rough-and-ready calculus of this sort in a letter to Joseph Priestly in 1772.[8]. At least one contemporary undergraduate textbook in critical thinking (Ennis 1996) has incorporated the "Ben Franklin method," with lots of caveats, as an informal approach to decision-making of this sort. The method, as described by Ennis, involves listing the pros and cons in opposite columns and crossing out competing considerations judged to be of roughly similar "weight;" sometimes, for example, two pro considerations might be judged to be jointly of equal weight to one con consideration. When all the considerations on one side are crossed out, and some considerations remain uncrossed out on the other side, the latter side is the one to adopt. A weakness in applying this rough-and-ready approach is a poverty of imagination and lack of background knowledge required to generate a full enough range and detail of competing considerations.

The most widely-taught calculus for decisions under uncertainty is classical decision

[8] Marius Vermaak mentions this letter in his (1999: 829).

theory (e.g. von Neumann and Morgenstern 1944, Raiffa 1968), which first identifies a range of decision options and possible states of nature (scenarios), identifies outcomes for each option under each scenario and then assigns quantitative probabilities to these scenarios and utilities (or consequential losses) to the outcomes.[9] A decision rule is then used to choose between different outcomes on the basis of the probabilities and utilities. Which decision rule is used may depend upon the degree of risk aversion of the decision-maker. Classical decision theory generally uses the rule which selects that option with the maximum expected utility, where the expected utility of an option is defined as the total utilities summed across all the scenarios weighted by the probability of each scenario.

There are many criticisms that have been made of this approach. Firstly, one has to define and agree possible decision options and states of nature, and this may be problematic. The possibilistic risk discussion in Artificial Intelligence, mentioned above, has arisen in response to this issue. Secondly, there is good reason to be sceptical about the possibility of measuring the well-being of an individual in terms of cardinal utilities, and even less so for groups of people. A careful discussion of the difficulties can be found in Pollock (1995: 14-18), whose theory of rational agency requires that agents can assign a cardinal measure to token situation-likings, in order to be able to calculate the expected likability of a situation type. His route to this cardinal measure assumes that human beings can introspect a preference relation among four situations which obtains when they prefer having B to having A more than they prefer having D to having C, that certain "reasonable assumptions" hold of the binary preference relation defined in terms of this quaternary preference relation and of the set of possible situation tokens, and that there are constraints on the quaternary preference relation which guarantee the existence of a cardinal measure. The complexity and

[9] Decision theory is often attributed to John von Neumann and Oskar Morgenstern (1944), although earlier work in this vein was published by Jerzy Neyman and Egon Pearson (1928), Abraham Wald and von Neuman himself.

tenuousness of these assumptions illustrate the difficulties in the way of applying classical decision theory under risk to human decision-making. Page (1978), for instance, noted that in domains of environmental risk assessment, the consequences of different regulatory options may so differ in the populations impacted, and in the likelihood, timing, duration, magnitude, extent and severity of impact, that meaningful comparison of (negative) utilities between options becomes in practice impossible. Thirdly, in many real-world domains it is not possible to assign quantitative probabilities or utilities, or it may only be possible to do so on assumptions which are contested. This is typically the case in decisions involving more than one participant, especially public policy decisions, where agreement between stakeholders on probabilities and utilities is rarely achievable.[10] Hansson (1999) noted that the difficulty of reaching agreement on probabilities of scenarios in public policy decisions often leads regulators and decision-makers to focus on that single possible scenario judged to have the highest probability of occurrence, to the exclusion of all other possibilities and to the potential detriment of the decision. On the other hand, there is some tendency among political activists to focus on the worst-consequence scenario. For these contrasting reasons, practitioners of scenario-planning techniques in the business world often oppose the assignment of probabilities to decision scenarios.[11]

A fourth difficulty arises with the maximum expected utility decision rule used in classical decision theory. The claimed superiority of the theory over alternative approaches is usually based on considerations of the asymptotic performance of this rule.[12] However, this rule is essentially an average (albeit weighted by probabilities), and, as such, it is not necessarily robust against small deviations in the probabilities or utilities used in its

[10] Jamieson (1996) has observed that stakeholders in public policy decisions may also have political reasons to establish scientific and other uncertainties and to resist their resolution.

[11] See the discussion on the internet at The Global Business Network (www.gbn.com).

calculation (Huber 1981). Small errors or changes in the input probabilities or utilities may lead to large changes at the other end – i.e. very different suggested decisions. For example, Banerjee and Sen (2000) have shown that a different rule is superior to the maximum expected utility rule for agents contemplating partnerships where the number of interactions with potential partners is finite, small and known in advance. The theory does countenance the use of other rules which may be more robust, such as those which accord with higher degrees of risk aversion (e.g. choosing that action with the least-worst possible outcome), although such rules may not have the asymptotic properties of the maximum expected utility rule. However, once again, the challenge of finding interpersonal agreement arises, as different people demonstrably have different degrees of risk aversion. The recent debate in the environmental risk domain over the use of the Precautionary Principle is evidence of these differences becoming manifest in public policy decision-making (Hansson 1999, Millstone *et al.* 1999, Sandin 1999).

It is possible to adapt classical decision theory to a Bayesian perspective (Raiffa 1968, Lindley 1985), where probabilities are treated as subjective. This approach merely reinforces the difficulties mentioned above of reaching inter-personal agreement in any business or public policy decision context. Moreover, when applied to group decision-making, Bayesian decision theory has been shown to be incoherent, in the sense that the decision option suggested by the theory may be not the option preferred by every participant (Seidenfeld *et al.* 1989). An associated issue is the fact that human decision-makers typically do not conform to classical normative models of decision-making, whether Bayesian or otherwise (Nisbet and Ross 1980, Kahneman *et al.* 1982, Schneider 1998). Such human decision-making is not necessarily irrational, but may be a rational response to limited computational or other resources, and/or limited time in which to make a decision. Moreover, as Rehg

---

[12] In other words, a decision-maker using the maximum expected utility rule will always eventually outperform a decision-maker using any other rule, in an infinite sequence of repeated decisions made using the rules.

(1997) has demonstrated, it can be rational for human decision-makers to incorporate rhetorical elements, such as arguments from authority or epideictic arguments (e.g. arguments that draw attention to themselves), into a decision. A key question then confronting designers of decision support systems will be to what extent the systems support the decision-making styles of the person or team (or agent) taking the decision, possibly in contravention of normative models of decision-making (McBurney *et al.* 1999).

Driven by the practical difficulties of implementation of classical decision theory, researchers in Artificial Intelligence have sought practical means of eliciting probabilities (e.g. van der Gaag *et al.* 1999) and utilities (e.g. Boutilier *et al.* 1997, Ha and Haddawy 1997), work which has a counterpart in the earlier development by applied marketing theorists of techniques for preference elicitation in purchase decision contexts (e.g. via conjoint analysis, as in Green and Krieger 1993). A second approach within Artificial Intelligence has been the development of qualitative approaches to decision calculi, including the use of logics of argumentation (see Parsons 2000 for a review). Fox and Parsons (1998), for example, propose a decision logic which explicitly represents the argument for a possible action along with the action, and the decision-maker may assign a value-label to this pair. However, this logic only requires the value-labels to be elements of a qualitative dictionary (e.g. {"positive", "negative"}), rather than numerical utilities or losses. The development of such qualitative decision calculi is still in its early days, and we agree with Vermaak (1999) who states that the question of evaluation of conductive arguments is the outstanding research problem in this area.

If successful, this development of new qualitative theories of decision on the basis of lessons learnt in applications will repeat the experience of Artificial Intelligence in applying probability theory, an experience which has led to the development of non-probability uncertainty formalisms (Hunter and Parsons 1998). We believe the same potential exists to extend argumentation theory in the process of developing operational decision support systems using argumentation. An obvious example will be the development of argumentation frameworks for deliberative dialogues, which are still to be developed.

Finally, it seems that "weighing up the pros and cons" involves a great deal of sensitivity to the particular decision-making situation, a great deal of background knowledge about potential consequences of possible courses of action (or policies) in that situation, the ability to entertain imaginatively and appreciate emotionally what these consequences would be like, and the extent to which they would be likely to occur, and considerable judgment. For decision-problems involving multiple stakeholders, interpersonal agreement on all these issues is a further requirement for practical decision-making. Thus computational modelling of this form of reasoning presents real challenges, both to argumentation and decision theory and to computer science.

5 Reasoning Under Resource Constraints

A key issue for real-world decision making and for guidance systems designed to assist such decision-making is the fact of limited resources, both of computational resources and of time. There are two aspects to this issue: resource limitations as part of the practical reasoning problem, for example, the allocation of limited medical treatment resources to competing patients, and resource limitations within the guidance system itself, for example, constraints on the time within which a course of action must be suggested. The former aspect can be included as part of the domain model which is a necessary part of the task of designing effective guidance systems, and so is not further discussed here. The latter aspect – constraints on the operations of the guidance system – can involve constraints on the input data available for processing, on the processing and computational power available, and on the time permitted for processing.

Such operational constraints were an early concern by one of the founders of Artificial Intelligence, Herbert Simon, and, within the discipline, the issue has received sporadic attention since then. Arguably, Simon's key notion was that of *satisficing*, that is of finding a solution to a problem which is good enough, rather than optimal. (Simon, 1957/1982) It is interesting that the two disciplines which have arguably devoted the most attention to the question of reasoning under resource constraints – Artificial Intelligence and marketing – both confront practical problems of modeling human rationality without the luxury of only theorizing. Much of the focus of marketing theory over the past thirty years has been the development of realistic and testable models of consumer decision-making. An early observation was that most consumers in most purchase decision situations do not consider all the alternatives open to them, but only a subset of these. They do so because they typically face non-zero costs of thinking and often time constraints in purchase decisions. This subset is called the Consideration Set by marketers (Lilien *et al.* 1992), and sophisticated quantitative models have been developed and tested to predict the selection of products or brands into the set, and, once inside, the evaluation of its elements to choose a particular one to purchase. Damasio (1994) has argued that such behavior by consumers can be understood in terms of an emotions-based model of reasoning and is a necessary part of being human.

Within Artificial Intelligence, Russell and collaborators (Russell and Wefald 1991, Russell and Subramanian 1995) developed models for decision-making under resource constraints using modifications of maximum-expected utility approaches, with a decision-maker repeatedly re-assessing the costs and benefits of acting now versus undertaking further deliberation. Perhaps because of the difficulties in calibrating the models involved, widespread implementation of these approaches has not occurred in Artificial Intelligence (see Schut and Wooldridge 2000b for a review). As discussed in the Artificial Intelligence community, this issue is related to that of intention reconsideration - the question of whether, when and how often to reconsider one's intentions or goals. This has been a concern of philosophers of action and intention (Bratman 1987, 1999) and of the agent-design community (Wooldridge and Parsons 1998, Schut and Wooldridge 2000a), especially within the Belief-Desires-Intention (BDI) paradigm. Within the latter, appropriate operationalization of these ideas is still at an early stage of research development.

Interestingly, dialectical argumentation may provide a means to address the problem of limited resources. If one thinks of group decision making, one need not address all conceivable counter-arguments to a claim (e.g., all those based on conflicts of interest), but argumentation and deliberation may be limited to only those counter-arguments raised in the debate. In this way, a more efficient use of resources can result, since one can quickly focus on the extant differences of opinion instead of on all possible differences. In cases where information is lacking, one can do as well as possible by arguing on the basis of only that information which is available. This approach to argumentation has received some attention within the Artificial Intelligence community, for example in the work of Loui (1993) and Vreeswijk (1997), and we believe it has considerable further potential in applications intended for practical reasoning.

A related question for implementation of effective decision support systems will be that of the computational complexity of any decision calculi and inference algorithms used within the systems. Recent work in Artificial Intelligence has addressed the issue of the complexity of different classes of acceptable arguments arising from the Toulmin argumentation scheme (Parsons and Wooldridge 2000), and similar analyses will be necessary for the argumentation engines of the practical reasoning systems discussed here.

6 Integration of moral considerations

An outstanding problem in developing a comprehensive theory of good practical reasoning is how to integrate moral considerations into an overall theory. One approach, worked out in an influential book by David Gauthier (1986) and explored further with computational modelling by Peter Danielson (1992, 1998), is to account for the evolution of morality among initially purely self-regarding individuals through a kind of "social contract" of the sort originally postulated by Thucydides and other thinkers of the classical period in ancient Greece (cf. Plato's *Republic* II.358e-359b), then taken up in the modern period by Hobbes and other thinkers in the social contract tradition. Much interesting work has been done in working out the implications of various assumptions about the starting-point for the kind of morality that arises from practical reasoning iterated through generations. The results of this work may however reveal the limitations of a contractualist understanding of morality. Since it is a paradigm of moral evil to take sadistic pleasure in torturing non-human animals, an adequate moral theory must imply a robust condemnation of such a practice. But "morals by agreement" can base condemnation of cruelty to animals only on the slender reed of offence to the feelings of some of the contracting humans, since non-human animals cannot themselves enter into contracts; cruelty practised beyond the ken of oddly sensitive human beings must therefore be morally unexceptionable on a contractualist account. A contractualist approach is one way to deal with the self-defeating character of purely self-regarding practical reasoning, as revealed by non-constant-sum games like the prisoner's dilemma. But it produces a truncated morality.

If one seeks to graft morality on to a system of practical reasoning which is directed at making the world more to the liking of the agent and those for whom the agent personally cares, then the first question will be what sort of morality to graft. Roughly speaking, contemporary moral theories divide into three main types: consequence-based, rule-based, and virtue-based.

Consequence-based theories seek to justify all moral claims by appeal to the consequences of abiding by them. For example, Singer's (1979) principle of the equal consideration of interests enjoins that a morally responsible agent act in each situation in a way likely to maximise the equally weighted interests all those affected by the agent's decisions, including non-human animals.

Rule-based theories are motivated by standard objections to purely consequence-based theories, that they countenance injustices for the sake of the general good (e.g. punishing an innocent person as an example to the rest) and that they give implausibly indirect rationales for such past-directed obligations as keeping contractual commitments and reciprocating past favours. Although rule-based theories can incorporate consequentialist principles as part of an overall moral theory, they are distinguished from consequence-based theories in that they assign moral status to certain types of actions independently of their consequences; a good example of such a rule-based theory is that of William Frankena in his (1987).

Highly general consequence-based and rule-based moral theories turn out to have implausible implications in particular situations, or to be difficult to apply at all, for example because they are too defeasible to be useful. Such problems have led to an explosion of philosophical interest in virtue-based moral theories, which give priority to the judgement which moral agents develop through a combination of skill and experience; a big stimulus to this interest was Alisdair McIntyre's (1985). Recent philosophical work in practical ethics has also exposed the limitations of top-down moral theorizing which starts from highly general abstract principles.

At least two recent moral theories, however, seem to combine generality with a precision which allows application to particular situations, and thus computational application. Richard and Val Routley[13] in their (1980) identify concentric spheres of items of

[13]Later they wrote under the names Richard Sylvan and Val Plumwood.

moral considerability, with a different type of moral responsibility corresponding to each sphere. Individuals with whom one can enter into contracts form the class represented by the innermost sphere; to them one can have obligations. Sentient creatures form a wider class. And so on, up to items like species and wilderness areas. Bernard Gert (1998) understands morality to be an informal public system by which the behaviour of moral agents is regulated; he has articulated what he takes to be the core of morality so understood which any rational agent must accept. His theory, which differs both from standard consequence-based and from standard rule-based theories, includes a list of ten moral rules, a set of moral ideals, and a conception of moral virtues.

In our opinion, it is an important problem to explore computationally the results of incorporating moral theories of these various types into systems for practical reasoning.[14] Integration of a comprehensive moral theory into a computational model of practical reasoning would not be simply a matter of imposing absolute prohibitions on policies which violate moral requirements, since in most sophisticated contemporary moral theories general moral principles are not absolute but defeasible. Further, they can be defeated not only by other moral principles which bear on a particular situation but also by non-moral, prudential considerations; for example, we take killing in justified self-defence to be an exception to the strong general prohibition against killing other human beings. Less sharply defined moral responsibilities, such as the responsibility of human beings to preserve and enhance biodiversity on earth, are even more obviously defeasible.

A particularly difficult problem is how to provide for discussion of intrinsically conflicting values in group decision-making, especially when the values are moral values. In extreme cases, a party to the discussion may refuse to entertain a fundamental question as a

[14] Note that, with the development of multi-agent systems, issues of trust and obligation between interacting software agents have assumed increasing prominence, especially in e-commerce applications. Such ideas may be implemented through the use of deontic logics, as outlined for example in the recent collection of McNamara

subject for discussion, because of a strong and inflexible attachment to a certain position on the question. This problem is one version of the problem of “deep disagreement.” Some theorists, such as Ackerman (1989), recommend that, in cases where two people differ on some dimension of the moral truth, they should say nothing in their conversation with one another about what divides them. To this recommendation of “conversational constraint,” Vermaak (1999) replies that such constraint is undesirable and unnecessary in what Kant called “public argument,” i.e. argument addressed to the world at large rather than to an audience restricted by role or function, and that unrestrained argument is the best way to handle deep disagreement. He cites a “very promising” account of deep disagreement by Henry Richardson (1995), in which Richardson adds to the list of obvious barriers to overcoming deep disagreement (stupidity, ignorance, obstinacy, arrogance, bias, prejudice) barriers requiring more effort to identify and remove, which are due to the facts that “(1) much learning is tacit, (2) much of what is learned is seen as *a priori* or definitional, and (3) inculcation of a form of life or a set of specialized practices typically takes for granted a rough characterization of the ends that are treated as final within that endeavor.” (Richardson 1995: 260; quoted in Vermaak (1999: 832)) It is a challenge to provide computationally for dialogue directed at removing such barriers. In such interactive situations, decisions could be made more effectively if there were some well established form of interactive deliberation. We turn to this in the next section.

7 Deliberation dialogue

Up to now, we have been speaking about practical reasoning as if it were carried on by a single individual thinking out what is to be done. But practical reasoning occurs also in interpersonal conversation, in what Walton and Krabbe (1995) call a ‘deliberation dialogue.’

---

and Prakken (1998). These approaches may provide examplars for the operationalization of the moral considerations discussed in this Section.

This is a form of interactive reasoning. A comprehensive computational approach to argument-based practical reasoning must obviously include theorizing about such interactive reasoning, both for its own sake and for the sake of providing computational support to groups making decisions.

Hamblin's ground breaking work in formal dialectic (1970) initiated much of the modern study of interactive reasoning (see Rescher 1977 on dialectics, Lorenz & Lorenzen 1978 on dialogical approaches to logic, Barth & Krabbe 1982 on formal dialectics and Van Eemeren and Grootendorst 1984 and 1992 on pragma-dialectics). Most of the research that has followed Hamblin's original work has concentrated on what Walton and Krabbe (1995) call 'persuasion dialogues', i.e., dialogues in which one party tries to get the other to accept a thesis or, conversely, the other party tries to refute the first party's thesis. The thesis in question in a given discussion may be a factual claimabout what to believe, or a claim about what is to be done formulated as an indicative sentence (e.g. that a certain course of action is best). For examples of the latter type of thesis, see Walton and Krabbe (1995) and van Eeemeren and Grootendorst's (1992) **"**incitive propositions". Persuasion dialogues are distinctive in requiring at least one party to advance a thesis at the beginning; initially open dialogues, whether about what to believe (inquiry dialogues) or what to do (deliberation dialogues), differ in this respect.[15]

But, the general approach to dialogue has drawn attention to other forms of dialogue which would be most useful in designs for decision support systems. The study of deliberation dialogues has received increasing attention both in argumentation theory and computational models of argumentation (Walton 1999, Hage/Leenes/Lodder 1994, Prakken & Sartor 1996). Deliberation dialogues typically do not begin with one participant proposing a course of action, which must then be justified to the rest. More commonly, the question of what to do is open at the beginning of the discussion. The question might be: Where shall we go for lunch? Or: What targets should we adopt for reduction of greenhouse gas emissions?

To the best of our knowledge, there are no formal systems of such open-ended deliberation dialogue analogous to the formal systems of persuasion dialogue referred to above. Some hints as to the structure of a formal system for deliberation dialogue can perhaps be gathered from Hitchcock (1991). There are 18 principles mentioned in Hitchcock, principles to be incorporated in a formal system for inquiry dialogue, i.e. dialogue about what to believe with respect to an open question. They make sense for formal systems of deliberation dialogue, including in particular externalization (formulation of rules in terms of observable linguistic behaviour), mutuality (nothing becoming a commitment of any participant until all accept it), orderliness (permissible locutions in a turn opening up at most one choice-point for the other interlocutor), staging (division of the dialogue into an invariant sequence of stages), logical pluralism, rule-consistency, realism, retraceability, provision for data collection, tentativeness, tracking and allocation of burden of proof.

The 'risk agora' of McBurney & Parsons (2000), although it is a formal system for persuasion dialogue, also provides a helpful parallel, in that it accommodates discussion of what rules of inference or argument schemes to use.

The dialogical approach can give insight into the central relevance of specification in practical reasoning (Bratman 1987, 1999, Richardson 1995). Especially Bratman has stressed that practical reasoning does not merely consist in selecting actions as they serve fixed goals, but involves specification of plans and ends. Often one's ends are not sufficiently precise to determine one's actions. For instance, if one wants to plan what to do on Saturday afternoon, there can be conflicting ends (like cutting the grass and going to the grocery store), and there are many ways to further specify these. The problem then is to reach a specification of possible plans and their relation to one's ends. Richardson has argued that such specification can in the end lead to the resolution of the initial conflicts. The dialogical approach provides

[15] For a taxonomy of dialogue types see Walton and Krabbe 1995.

a natural setting in which such specification takes place: specification occurs in response to the conflicts or underspecification as it arises in a discussion, either internally in oneself, or externally interacting with others. Similarly, it provides a starting point to address the problem of the reconsideration of plans, as stressed by Bratman. Plans tend to be relatively stable, which makes sense since otherwise the rational planning might take too much time in relation to the ends served. In decision support systems, one speaks of the necessity of *real time* operation. One can think of the exaggerated example of an uncontrollable plane going down while the pilot's decision support system asks to wait a minute since it is busy computing its advice.

A related issue has been raised by Wohlrapp (1995, 1998), and goes by the name of *retroflexive argumentation*. In retroflexive argumentation, given a heterogeneous group of considerations, a common ground is interactively sought by going back and forth between premises and conclusion while allowing changes to both of them. He illustrates the sort of dialogically situated frame confrontation, frame shifting and frame unification he has in mind through a couple of examples. It is an interesting question whether one could model computationally the sort of back-and-forth retroflexive argumentation/discussion as recommended by Wohlrapp.

Dialectical argumentation has recently begun to be applied in the design of multi-agent systems, where a group of intelligent software agents interact to achieve some common or separate goals. Parsons et al. (1998) proposed an argumentation framework to model negotiation dialogues between two agents, collaborating to achieve possibly-separate objectives which cannot be achieved without support or resources from each other. Subsequent work in the same vein, such as that of Amgoud and her collaborators (Amgoud, Maudet & Parsons 2000, Amgoud & Perrussel 2000) has explored more general argument frameworks, including persuasion dialogues and dialogues over differing preferences.

However, to our knowledge, no work in agent design has used a formal model of a deliberation dialogue, although systems such as Loui *et al.*’s Room 5 (1997) and Zeno System (Gordon & Karacapilidis 1997), and the Risk Agora of McBurney and Parsons (2000a) seek to support human deliberations in (respectively) law, land-use planning and environmental regulation. Verheij’s ArguMed (1999a) is called a system for automated argument assistance, and is meant for individual users as a tool for drafting and analyzing their argumentation. Aakhus (1999) has made critical remarks about the design of this kind of product. We consider that there is an urgent need to develop a general theoretical framework for formal systems of deliberation dialogue.

A point of attention with the kind of guidance systems under discussion involves the respect for and enhancement of the user’s (or users’) autonomy. There are many pitfalls. Automated systems tend to be attributed authority uncritically. Also persistent use of such a system can lead to insensitivity to the system’s peculiarities.

8 Interface Design

It is vital that any decision support system should be constructed in accordance with the principle that it is to give advice to people in a way that facilitates the consideration of the advice. Annoying and difficult features of the human to system interface could defeat the whole point of the system. One can also think of the effects of the (in)directness of communication. For instance, saying ‘Step aside!’ as opposed to ‘Excuse me’ invokes different reactions from the one addressed.

As long ago as 1985 the importance of interface design was discussed in Michie and Johnston (1985:56-75). There is an excellent discussion which is as timely now as it was then. They take a series of cases where the behaviour of machine systems baffles and defeats human efforts to solve problems ranging from the Three Mile Island disaster to the machine defeat of chess Masters. The machine systems are not always computer systems. But they all have information interfaces with human users, and the designers seemed not to have had the slightest understanding of human reactions to the information flows presented by these systems, information flows which were intended to support decision making, but which did precisely the reverse.

From their discussion it follows that there are at least four important aspects of good interface design. These apply to the decision support systems we have been discussing. It is possible that the interface will have to be different from domain to domain in order to conform to the principles we discuss below.

First, the rhetoric of guidance presentation should be appropriate for the audience. The advice presentation should encourage the audience to take notice and to focus on the important and salient issues. The presentation should neither confuse the audience with detail nor drive it up the wall with dysfunctional operation. The advice should be posed in such a way that final decisions are clearly left to people.

One way of doing this is to open up various options for action, but not too many options. For example, some decision-making situations are best conceptualized in terms of a matrix whose rows are options (alternative plans, which may be partial), whose columns are (really possible) states of the world which affect the outcome of each option, and whose cells are the outcomes for each option-state combination. Some rules of thumb for human decision-making are:

1. Make sure that your options are mutually exclusive.
2. Restrict yourself to at most 3 or 4 options.
3. Options do not have to be jointly exhaustive.
4. The options should include what strike you initially as the most promising.
5. The states should be mutually exclusive and jointly exhaustive.

6. Find out as much as you need about the outcome in each option-state combination.

7. Do as much as is reasonable to check the quality of your information.

8. It is often enough just to have a rough rating or ranking of the outcomes.

9. If none of the options seems to have a satisfactory outcome, rethink your options.

10. If you still don’t like any of the options, consider delaying your decision.

11. If one option strikes you as clearly satisfactory, choose it.

The last four of these open up the question of negotiation or deliberation when two or more parties are involved in considering options. This is addressed elsewhere in this essay.

Second, there should be transparency. People should be able to understand the advice given. It should not involve notions which are ambiguous, too technical, of the wrong degree of precision, or too vague. Since the decision support system is to be based on argumentation theory, the arguments should be available If options are presented, the arguments for each option and their structures should be available.

The structures of argument could be made available by both visual representation and textual representation. There is evidence (see M. Ford, 19xx) that some users are liable to find visual representation difficult to understand, but to find textual description easy to understand. Other users will be amenable to visual representations, but will find textual representations difficult to cope with. The interface should have both interaction modalities available. User modelling should indicate which is appropriate for each type of user.

Third, the possibility of reconsideration by and discussion with the system should be available, as has been indicated in the section above on deliberation dialogue. The interface should facilitate such interactions and not make them discouragingly difficult.

Fourth, the conceptual structures involved in working out the advice should be analogous to those used by people. The decision support system and the people seeking guidance should share the same conceptual frameworks.

9 Evaluation

A key issue for any decision support system which will be applied in practice is that of the evaluation of the system. There are a number of dimensions relevant here and most of these will require considerable further thought and analysis.

The first issue is what precisely is being measured by any evaluation. A high-quality decision support system would presumably offer relevant, complete, timely, appropriate, reasonable, suitably-precise and easily-understood advice, which would be taken into account (and not ignored), by the decision-maker and be supportive of the process of decision-taking. Do we measure quality by the calibre of the advice given, or by its completeness, timeliness, etc? These dimensions are not equivalent, and different dimensions may be appropriate in different domains of application. Measuring system performance on only one or several dimensions may lead to erroneous understandings of the totality of performance-in-use. Do we assess the system by the extent to which decision-makers take the advice and/or embrace the technology represented by the guidance system? Use of a technology may be affected by factors unrelated to any narrow technical or design aspect of the system concerned. Emery *et al.* (1999), for instance, report a trial of an intelligent decision support system to assist medical personnel in assessing the need for screening for genetic factors as a potential cause of breast cancers in patients. They report that some doctors involved expressed a preference for a system which worked off-line, rather than being able to present the recommendation to the doctor immediately, so that he or she would have time in which to formulate an appropriate form of words to inform each patient of the system's recommendations.

There are several associated philosophical questions around the issue of evaluation which do not admit straightforward responses. How does one measure the quality of the system's advice if the suggested advice is not taken, or only partly taken? How does one measure this quality if the advice is taken, but the world changes in a salient way between the giving of the advice and its execution? Most management consultancy advice, for instance, is not assessable or assessed for these two reasons. How can the system be evaluated for extreme situations or rare events? For example, a system designed to support water-flow management through a dam will only be required to recommend actions for 200-year floods[16] on average once every 200 years. There may be insufficient data to design the system or to predict its performance in these circumstances, and possibly only one case every 200 years on which to base an assessment of that performance. Moreover, if a system is designed for an entirely new activity, or if it completely replaces earlier means of decision-making, how does one assess the adequacy of its advice? The various AI systems currently being deployed by NASA for control of autonomous space-craft are an example of systems which support completely new activities. How is it possible to rate their performance in any other but crude terms, such as overall mission success versus non-success? These issues are related to the notions of delimitation uncertainty and possibilistic risk mentioned in Section 5 above.

The NASA example demonstrates that there may also be domain-specific issues involved in measurement of the quality of decision-guidance systems. In a recent review of the research on medical decision-making, for instance, Schneider (1998) notes that it is almost impossible to assess the quality of medical treatment decisions. Selecting one procedure or course of treatment for a patient usually precludes the selection of alternatives, and so comparison of results of alternative decision options for an individual patient are impossible. The diversity and complexity of individual circumstances and medical aetiologies make comparisons at an aggregated level also highly problematic. For intelligent guidance systems, we might assess the performance of a group of system-guided medical decision-makers with a control group not so guided, but ensuring that each group is presented with a matched set of patients, with equivalent case histories, is likely to present challenges to

[16] i.e. floods which are expected to occur once every 200 years.

the statistical design of such an evaluation test. Moreover, as argued in McBurney *et al.* (1999), effective, efficient and useable decision support systems would need to be tailored to the possibility of very different decision-making styles of the human agents taking the decision. This may add another level of complexity to the issue of evaluation, since each deployment of a system may differ according to the decision-maker or team using it.

This brief discussion demonstrates the many philosophical and practical questions to be faced in evaluating decision support systems. We believe these issues are yet to be resolved, even in principle, and their resolution would greatly benefit from collaborative research from philosophy, artificial intelligence, statistics and the various decision application domains.

10 Summary

We have recommended that a decision support system be used for the guidance of agents, singly or in groups, in deciding in a wide range of domains what is to be done. Whatever the system, it should be based on argumentation, and transparent in that respect to any user. The basis for the system should not be spartan in its use of argumentation schemes and techniques. It should reflect the richness of quality argumentation, and should use the techniques appropriate to the domain in which it gives advice. There should be an open ended approach to advising, and users should be able to deliberate jointly with the system about advice and how it is generated. The interactive interface between agents and the machine should facilitate the giving of advice and the joint activities of system and agents.

References

Aakhus, Mark (1999). ‘Reconstruction games: Assessing the resources for managing collective argumentation in groupware technology.’ In: van Eemeren et al., 1-7.

Ackerman, Bruce (1989). ‘Why dialogue?’ *The Journal of Philosophy* 86: 5-22.

Amgoud, L., N. Maudet and S. Parsons (2000). “Modeling dialogues using argumentation.” In: E. Durfee (Editor): *Proceedings of the Fourth International Conference on Multi-Agent Systems (ICMAS-2000),* Boston, MA, USA: IEEE Press.

Amgoud, L. and L. Perrussel (2000). “Arguments and contextual preferences.” In: G. Vreeswijk (Editor): *Proceedings of the ECAI Workshop on Computational Dialectics.* Berlin, Germany: ECAI.

Ashley, K. (1990). *Modeling Legal Argument. Reasoning with Cases and Hypotheticals.* Cambridge: MIT Press.

Banerjee, Bikramjit and Sandip Sen (2000): "Selecting partners." In: Piotr Gmytrasiewicz and Simon Parsons (Editors): *Proceedings of the Second Workshop on Game-Theoretic and Decision-Theoretic Agents, Conference on Agent Technologies, Architectures and Languages* (ATAL2000), Boston, MA.
The proceedings of GTDT2000 are available on-line at:
http://www.csc.liv.ac.uk/~sp/events/gtdt/gtdt00/proc.html

Barth, Else M., & Erik W. C. Krabbe. (1982) *From Axiom to Dialogue*. New York: de Gruyter.

Blair, J. Anthony (1999). ‘Walton’s argumentation schemes for presumptive reasoning: A critique and development.’ In: van Eemeren et al., 56-61.

Bourinot, John George (1995). *Bourinot's Rules of Order*, 4th ed., ed. Geoffrey H. Stanford. Toronto: McClelland & Stewart.

Boutilier, C, R. Brafman, C. Geib and D. Poole (1997). ‘A constraint-based approach to preference elication and decision making.’ *Stanford Spring Symposium on Qualitative Preferences in Deliberation and Practical Reasoning.* Menlo Park, CA, USA: AAAI Press.

Brandom, Robert (1994). *Making it Explicit: Reasoning, Representing and Discursive Commitment.* Cambridge, MA: Harvard University Press.

Bratman, Michael (1987). *Intention, Plans and Practical Reason*. Cambridge, MA/London: Harvard University Press.

Bratman, Michael (1992). ‘Planning and the stability of intention.’ *Minds and Machines* 2: 1-16.

Bratman, Michael (1999). *Faces of Intention: Selected Essays on Intention and Agency*. Cambridge: Cambridge University Press.

Carbogim, D.V., D. Robertson and J. Lee (2000). “Argument-based applications to knowledge engineering.” *Knowledge Engineering Review* (in press).

Chesnevar, Carlos I., Ana G. Maguitman and Ron P. Loui (2000). “Logical Models of Argument.” *ACM Computing Surveys.* (In press).

Cohen, L. Jonathan (1977). *The Probable and the Provable.* Oxford, UK: Clarendon Press.

Damasio, Antonio R. (1994). *Descartes’ Error.* New York, NY, USA: G. P. Putnam’s Sons.

Danielson, Peter A. (1992). *Artificial Morality*. London: Routledge.

Danielson, Peter A., ed. (1998). *Modeling Rationality, Morality, and Evolution*. New York: Oxford University Press.

Eemeren, Frans H. van, and Rob Grootendorst (1984). *Speech Acts in Argumentative Discussions*. Dordrecht: Foris.

Eemeren, Frans H. van, and Rob Grootendorst (1992). *Argumentation, Communication and Fallacies*. Mahwah, NJ: Lawrence Erlbaum Associates.

Eemeren, Frans H. van, Rob Grootendorst, J. Anthony Blair & Charles A. Willard, eds.

(1999). *Proceedings of the Fourth International Conference of the International Society for the Study of Argumentation* (Amsterdam: Sic Sat).
Eemeren, F.H. van, Grootendorst, R. and Kruiger, T. (1987). *Handbook of Argumentation Theory. A Critical Survey of Classical Backgrounds and Modern Studies*. Dordrecht, The Netherlands: Foris Publications. Translation of van Eemeren *et al.* (1981).
Ehninger, Douglas, & Wayne Brockriede (1963). *Decision by Debate*. New York: Dodd, Mead & Company.
Emery, J., R. Walton, A. Coulson, D. W. Glasspool, S. Ziebland and J. Fox (1999). "Computer support for recording and interpreting family histories of breast and ovarian cancer in primary care (RAGs): qualitative evaluation with simulated patients." *British Medical Journal*, 319: 32--36.
Ennis, Robert H. (1996). *Critical Thinking*. Upper Saddle River, NJ: Prentice-Hall.
Fisher, Roger & William Ury, with Bruce Patton, editor. (1983). *Getting to Yes: Negotiating Agreement Without Giving in*. New York: Penguin Books.
Fox, John (1999). "Will it happen? Can it happen? A new approach to formal risk analysis." *Risk, Decision and Policy*, 4(2) (August) 117-128.
Fox, John and Subatra Das (2000): *Safe and Sound: Artificial Intelligence in Hazardous Applications.* Menlo Park, CA/Cambridge, MA: AAAI Press/MIT Press.
Fox, John and Simon Parsons (1998). "Arguing about beliefs and actions." In: A. Hunter and S. Parsons (Editors): *Applications of Uncertainty Formalisms*. Berlin, Germany: Springer Verlag (LNAI 1455), 266—302.
Fox, John and Richard Thomson (1998). "Decision support and disease management: a logic engineering approach." *IEEE Transactions on Information Technology in Biomedicine*, 2 (4): 217-228.
Frankena, William (1987). *Ethics*, Second edition. Englewood Cliffs, NJ: Prentice-Hall.
van der Gaag, L.C., S. Renooij, C. L. M. Witteman, B. M. P. Aleman and B. G. Taal (1999). "How to elicit many probabilities." *Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence (UAI99).* (Editors D.Geiger and P. Shenoy) Morgan Kaufman: San Francisco, CA, USA.
Gabbay, D. M., Hogger, C. J., and Robinson, J. A. (Editors) (1994). *Handbook of Logic in Artificial Intelligence and Logic Programming. Volume 3. Nonmonotonic Reasoning and Uncertain Reasoning.* Oxford, UK: Clarendon Press.
Gauthier, David (1986). *Morals by Agreement*. Oxford : Clarendon Press.
Gert, Bernard (1998). *Morality: Its Nature and Justification.* New York: Oxford University Press.
Gilbert, Michael A. (1997). *Coalescent Argumentation*. Mahwah, NJ: Lawrence Erlbaum Associates.
Goodman, Nelson (1954). *Fact, Fiction, and Forecast*. London: Athlone Press.
Gordon, Thomas F. (1993). "The Pleadings Game." *AI and Law,* 2 (4): 239-292.
Gordon, Thomas F. and N. Karacapilidis (1997): "The Zeno argumentation framework." *Proceedings of the Sixth International Conference on AI and Law*. New York, NY: ACM Press, 10-18.
Govier, Trudy (1987). *Problems in Argument Analysis and Evaluation*. Dordrecht: Foris.
Govier, Trudy (1999). *The Philosophy of Argument*. Newport News, VA: Vale Press.
Green, P.E. and A. M. Krieger (1993). "Conjoint analysis with product-positioning applications." In: J. Eliashberg and G. L. Lilien (Editors): *Handbooks in Operations Research and Management Science, Volume 5: Marketing.* Amsterdam, The Netherlands: North-Holland, 467—516.
Grennan, Wayne (1997). *Informal Logic : Issues and Techniques*. Montreal : McGill-Queen's University Press.

Ha, V. and P. Haddawy (1997). "Problem-focused incremental elicitation of multi-attribute utility models." In: *Proceedings of the Thirteenth Conference on Uncertainty in Artificial Intelligence.* San Francisco, CA, USA: Morgan Kaufmann, 215—222.
Hage, J. (1997). *Reasoning with Rules. An Essay on Legal Reasoning and Its Underlying Logic.* Dordrecht, The Netherlands: Kluwer Academic Publishers.
Hage, J.C., Leenes, R. and Lodder, A.R. (1994). Hard Cases: A Procedural Approach. *Artificial Intelligence and Law*, 2: 113-167.
Hamblin, C.L. (1970). *Fallacies.* London, UK: Methuen.
Hansson, S.O. (1996). "Decision making under great uncertainty." *Philosophy of the Social Sciences*, 26 (3): 369—386.
Hansson, S.O. (1999). "Adjusting scientific practices to the Precautionary Principle." *Human and Ecological Risk Assessment,* 5 (5): 909—921.
Hastings, Arthur (1963). *A Reformulation of the Modes of Reasoning in Argumentation*. Doctoral dissertation, Northwestern University.
Hintikka, J., and Sandu, G. (1997). Game-Theoretical Semantics. In: J. van Benthem and A. ter Meulen (Editors)*: Handbook of Logic and Language*, 361-410. Amsterdam, The Netherlands: Elsevier Science.
Hitchcock, David (1991). 'Some principles of rational mutual inquiry.' In: Frans van Eemeren, Rob Grootendorst, J. Anthony Blair & Charles A. Willard, eds., *Proceedings of the Second International Conference on Argumentation* (Amsterdam: SICSAT), 236-43.
Hitchcock, David (1994). 'Validity in conductive arguments.' In: Ralph H. Johnson & J. Anthony Blair (eds.), *New Essays in Informal Logic* (Windsor, ON: Informal Logic), 58-66.
Hitchcock, David (forthcoming). 'Pollock's theory of rationality.' Forthcoming in the proceedings of the conference of the Ontario Society for the Study of Argumentation, 'Argumentation at the turn of the century,' held at Brock University, St. Catharines, Ontario, May 1999.
Huber, Peter J. (1981). *Robust Statistics.* New York, NY: John Wiley and Sons.
Hunter, Anthony and Simon Parsons (Editors) (1998). *Applications of Uncertainty Formalisms.* Berlin, Germany: Springer Verlag (LNAI 1455).
Jacobs, Scott (1999). 'Argumentation as normative pragmatics.' In: van Eemeren et al., 397-403.
Jamieson, D. (1996). "Scientific uncertainty and the political process." *Annals of the American Academy of Political and Social Science.* 545: 35—43.
Kahneman, D., P. Slovic and A. Tversky (1982) (Editors): *Judgment under Uncertainty: Heuristics and Biases.* Cambridge, UK: Cambridge University Press.
Kienpointner, Manfred (1992). *Alltagslogik: Struktur und Funktion von Argumentationsmustern*. Stuttgart-Bad Cannstatt: Frommannn-Holzboog.
Krause, P., S. Ambler, M. Elvang-Gorannson and J. Fox (1995). "A logic of argumentation for reasoning under uncertainty." *Computational Intelligence.* 11 (1): 113—131.
Krause, P., J. Fox, P. Judson and M. Patel (1998). "Qualitative risk assessment fulfils a need." In: A. Hunter and S. Parsons (Editors): *Applications of Uncertainty Formalisms*. Berlin, Germany: Springer Verlag (LNAI 1455), 138—156.
Lapintie, Kimmo (1999). 'The concept of resolving differences of opinion and its practical implications in planning theory.' In: van Eemeren et al., 500-504.
Lilien, G.L., P. Kotler and K. S. Moorthy (1992). *Marketing Models*. Englewood Cliffs, NJ, USA: Prentice-Hall.
Lindley, D.V. (1985). *Making Decisions.* 2nd Edition. London, UK: John Wiley and Sons.
Lorenzen, P., and Lorenz, K. (1978). *Dialogische Logik.* Darmstadt, Germany:

Wissenschaftliche Buchgesellschaft.
Loui, Ron P (1993). “Process and policy: resource-bounded, non-demonstrative reasoning.” *Computational Intelligence*, 2 (5).
Loui, Ron P., Norman, J., Altepeter, J., Pinkard, D., Craven, D., Lindsay, J., and Foltz, M. (1997). “Progress on Room 5. A Testbed for Public Interactive Semi-Formal Legal Argumentation.” *The Sixth International Conference on Artificial Intelligence and Law. Proceedings of the Conference*, pp. 207-214. New York, NY: ACM.
Makau, Josina M. (1999). ‘The case for cooperative argumentation.’ In: van Eemeren et al., 548-551.
McBurney, Peter, Simon Parsons, John Fox and David Glasspool (1999). *Consequence Logics for Cancer Risk Counselling Systems: A Specification Framework.* London, UK: Technical Report, Department of Electronic Engineering, Queen Mary & Westfield College, University of London.
McBurney, Peter and Simon Parsons (2000a). "Intelligent systems to support deliberative democracy in environmental regulation." In: D. Peterson (Editor): *Proceedings of the AISB Symposium on AI and Legal Reasoning.* Birmingham, UK: AISB.
McBurney, Peter and Simon Parsons (2000b): "Risk Agoras: Dialectical argumentation for scientific reasoning." In: C. Boutilier and M. Goldszmidt (Editors): *Proceedings of the Sixteenth Conference on Uncertainty in Artificial Intelligence (UAI-2000).* Stanford, CA.
McIntyre, Alisdair (1985). *After Virtue.* London: Duckworth.
Paul McNamara and Henry Prakken (Editors) (1998): *Norms, Logics and Information Systems: New Studies in Deontic Logic and Computer Science.* Amsterdam: IOS Press.
Mengel, Peter. (1999). ‘The concept of argumentation in Peter Singer’s *Practical Ethics*.’ In: van Eemeren et al., 569-572.
Michie, D. and Johnston, R. (1985). *The Creative Computer*, Harmondsworth, Penguin Books.
Milgram, Elijah. “Practical reasoning.” Entry in the online *Dictionary of Mind* at http://artsci.wustl.edu/~philos/MindDict/practicalreasoning.html. Downloaded on June 21, 2000.
Millstone, E., E. Brunner and S. Mayer (1999). "Beyond ”substantial equivalence”." *Nature,* 401: 525—526.
Minthorn, A. (1996). ‘Human remains should be reburied.’ http://www.ucinet.com/~umatribe/kennman (not personally visited)
Naess, Arne (1966/1947). *Communication and Argument. Elements of Applied Semantics.* London, UK: Allen and Unwin. Norwegian version published in 1947.
Nisbett, R. and L. Ross (1980). *Human Inference: Strategies and Shortcomings of Social Judgment.* Prentice-Hall: Englewood Cliffs, NJ, USA.
von Neumann, John and Oskar Morgenstern (1944): *Theory of Games and Economic Behaviour.* Princeton, NJ: Princeton University Press.
Neyman, Jerzy and Egon S. Pearson (1928). “On the use and interpretation of certain test criteria for purposes of statistical inference, Part I." *Biometrika,* 20A: 175—240.
Page, Talbot (1978). "A generic view of toxic chemicals and similar risks." *Ecology Law Quarterly,* 7 (2): 207—244.
Parsons, Simon (2000). *Qualitative Approaches to Reasoning under Uncertainty.* Cambridge, MA, USA: MIT Press (in press).
Parsons, Simon, Carles Sierra and Nick Jennings (1998): “Agents that reason and negotiate by arguing.” *Journal of Logic and Computation,* 8(3): 261-292.
Parsons, Simon and Michael Wooldridge (2000): *The Complexity of Argumentation.*

Liverpool, UK: Technical Report, Department of Computer Science, Liverpool University.
Perelman, Chaim, & Lucie Olbrechts-Tyteca (1971/1958). *The New Rhetoric: A Treatise on Argumentation*, translated by John Wilkinson and Purcell Weaver. Notre Dame/London: University of Notre Dame Press. French-language original first published in 1958.
Pollock, John L. (1970). 'The structure of epistemic justification.' *American Philosophical Quarterly*, monograph series 4: 62-78.
Pollock, John L. (1987). "Defeasible reasoning." *Cognitive Science*, 11: 481-518.
Pollock, John L. (1995). *Cognitive Carpentry: A Blueprint for How to Build a Person*. Cambridge, MA/London: MIT Press.
Pollock, John L. (1999). *Download OSCAR*. http://www.u.arizona.edu/~pollock/oscar.html. Visited on April 24.
Prakken, Henry, and Giovanni Sartor (1996). "A Dialectical Model of Assessing Conflicting Arguments in Legal Reasoning." *Artificial Intelligence and Law*, 4: 331-368.
Prakken, Henry, and Gerhard Vreeswijk (*to appear*). "Logics for Defeasible Argumentation." In: D. Gabbay (Editor): *Handbook of Philosophical Logic*. Dordrecht, The Netherlands: Kluwer Academic Publishers.
Raiffa, Howard (1968). *Decision Analysis: Introductory Lectures on Choices Under Uncertainty.* New York,USA: McGraw-Hill.
Rawls, John (1971). *A Theory of Justice.* Cambridge, MA: Belknap Press of Harvard University Press.
Raz, Joseph, ed. (1978). *Practical Reasoning*. Oxford: Oxford University Press.
Rehg, William (1997). "Reason and rhetoric in Habermas' s Theory of Argumentation." In: W. Jost and M. J. Hyde (Editors): *Rhetoric and Hermeneutics in Our Time: A Reader.* New Haven, CN, USA: Yale University Press, 358—377.
Reiter, R. (1980). "A Logic for Default Reasoning." *Artificial Intelligence*, 13: 81-132.
Rescher, Nicholas (1977). *Dialectics: A Controversy-Oriented Approach to the Theory of Knowledge.* Albany, NY, USA: SUNY Press.
Robert, Sarah Corbin, with the assistance of Henry M. Robert III, James W. Cleary & William J. Evans. (1970) *Robert's Rules of Order, Newly Revised*, a new and enlarged edition. Glenview, IL: Scott, Foresman.
Richardson, Henry S. (1995). *Practical Reasoning about Ends*. Cambridge: Cambridge University Press.
Routley, Richard & Val (1980). 'Human Chauvinism and Environmental Ethics.' In: D. S. Mannison, M. McRobbie and Richard Routley (eds.), *Environmental Philosophy* (Canberra: Research School of Social Sciences, ANU), 96-189.
Russell, Stuart J. and D. Subramanian (1995). "Provably bounded-optimal agents." *Journal of Artificial Intelligence Research*, 1: 1—36.
Russell, Stuart J. and Eric Wefald (1991). *Do the Right Thing: Studies in Limited Rationality.* Cambridge, MA, USA: MIT Press.
Sandin, P. (1999). "Dimensions of the Precautionary Principle." *Human and Ecological Risk Assessment*, 5 (5): 889—907.
Schneider, C.E. (1998). *The Practice of Autonomy: Patients, Doctors and Medical Decisions.* Oxford, UK: Oxford University Press.
Schut, Martijn C. and Michael J. Wooldridge (2000a). "Intention reconsideration in complex environments." In: M. Gini and J. Rosenschein (Editors): *Proccedings of the Fourth International Conference on Autonomous Agents*, Barcelona, Spain.
Schut, Martijn C. and Michael J. Wooldridge (2000b). "The control of reasoning in resource-bounded agents." *Knowledge Engineering Review* (in press).

Seidenfeld, T., J. B. Kadane and M. J. Schervish (1989). "On the shared preferences of two Bayesian decision makers." *Journal of Philosophy*, 86 (5): 225--244.

Simon, Herbert (1957/1982). "Rational Choice and the Structure of the Environment" reprinted in H. A. Simon (1982): "Models of Bounded Rationality", Cambridge, MA: MIT Press, pp. 259-268..

Singer, Peter (1979). *Practical Ethics*. Cambridge/New York: Cambridge University Press.

Slade, Christina (1999). 'Reasons to buy: Teaching reasoning through television.' In: van Eemeren et al., 746-750.

Stirling, A. and S. Mayer (1999). *Rethinking Risk: A Pilot Multi-Criteria Mapping of a Genetically Modified Crop in Agricultural Systems in the UK.* Brighton, UK: Report, Science and Technology Policy Research Unit (SPRU), University of Sussex.

Toulmin, Stephen (1958). *The Uses of Argument*. Cambridge: Cambridge University Press.

Verheij, Bart (1996). *Rules, Reasons, Arguments: Formal studies of argumentation and defeat*. Dissertation Universiteit Maastricht. Summary and table of contents available at http://www.metajur.unimaas.nl/~bart/proefschrift/

Verheij, B. (1999a). "Automated Argument Assistance for Lawyers." *The Seventh International Conference on Artificial Intelligence and Law. Proceedings of the Conference*, 43-52. New York, NY: ACM.

Verheij, Bart (1999b). 'Argument mediation for lawyers: the presentation of arguments.' In: van Eemeren et al., 820-828.

Verheij, B. (1999c). "Logic, context and valid inference. Or: Can there be a logic of law?" In: *Legal Knowledge Based Systems. JURIX 1999: The Twelfth Conference* (eds. H.J. van den Herik, M.-F. Moens, J. Bing, B. van Buggenhout, J. Zeleznikow, and C.A.F.M. Grütters), pp. 109-121. Gerard Noodt Instituut, Nijmegen. The proceedings of JURIX 1999 are available on-line at: http://www.jurix.nl.

Vermaak, Marius (1999). 'On conversational constraint.' In: van Eemeren et al., 829-832.

Volquardsen, Björg (1999). 'Do advertisers argue in their campaigns?' In: van Eemeren et al., 842-845.

Vreeswijk, Gerard A.W. (1997). "Abstract argumentation systems". *Artificial Intelligence*, Vol. 90, pp. 225-279.

Walton, Douglas N. (1990). *Practical Reasoning: Goal-driven, Knowledge-based Action-guiding Argumentation*. Savage, MD: Rowman & Littlefield.

Walton, Douglas N. (1996). *Argument Schemes for Presumptive Reasoning*. Mahwah, NJ: Lawrence Erlbaum Associates.

Walton, Douglas N. (1998). *The New Dialectic: Conversational Contexts of Arguments.* Toronto, Canada: University of Toronto Press.

Walton, Douglas (1999). *One-Sided Arguments: A Dialectical Analysis of Bias*. Albany, NY: State University of Albany Press.

Walton, Douglas N., & Erik C. W. Krabbe (1995). *Commitment in Dialogue*. Albany, NY: State University of New York Press.

Wellman, Carl (1971), *Challenge and Response: Justification in Ethics*. Carbondale, IL: Southern Illinois University Press.

Windes, R. R., & Arthur Hastings (1965). *Argumentation & Advocacy*. New York: Random House.

Wohlrapp, Harald (1995). 'Resolving the riddle of the non-deductive argumentation schemes.' In: Frans H. van Eemeren, Rob Grootendorst, J. Anthony Blair & Charles A. Willard (eds.), *Analysis and Evaluation: Proceedings of the Third ISSA Conference on Argumentation (University of Amsterdam, June 21-24, 1994), Volume II* (Amsterdam: International Centre for the Study of Argumentation, Sic Sat), 55-62.

Wohlrapp, Harald (1996). 'Heterogenität als argumentationstheoretisches Problem.' In: Harm

Paschen & Lothar Wigger (eds.), *Schulautonomie als Entscheidungsproblem* (Weinheim: Deutschen Studienverlag), 43-57.

Wohlrapp, Harald (1998). ‘A new light on non-deductive argumentation schemes,’ *Argumentation,* 12: 341-350. Rewritten version of Wohlrapp (1995).

Wooldridge, Michael (2000). *Reasoning about Rational Agents.* Cambridge, MA, USA: MIT Press (in press).

Wooldridge, Michael and Simon Parsons (1998). “Intention reconsideration reconsidered.” In: J. P. Muller, M. Singh and A. Rao (Editors): *Intelligent Agents V.* Berlin, Germany: Springer Verlag (LNAI 1365).