DOI 10.3233/AAC-190480
IOS Press

Using argumentation schemes to find motives and intentions of a rational agent

Douglas Walton [†]
Centre for Research on Reasoning, Argumentation and Rhetoric, University of Windsor, Windsor, Canada

Abstract. Because motives and intentions are internal, and not directly observable by another agent, it has always been a problem to find a pathway of reasoning linking them to externally observable evidence. This paper proposes an argumentation-based method that one can use to support or attack hypotheses about the motives or intentions of an intelligent autonomous agent based on verifiable evidence. The method is based on a dialectical argumentation approach along with a commitment-based theory of mind. It is implemented using the Carneades Argumentation System, which already has 106 programmed schemes available to it and has an argument search tool. The method uses schemes, notably ones for abductive reasoning, argument from action to intention, argument from action to motive, and some new ones.

Keywords: Multiagent systems, evidential reasoning in law, finding intentions, artificial intelligence

1. Introduction

Research on intention recognition carried out in computer science for over thirty years has addressed the problem of identifying the intentions of an intelligent autonomous agent based on evidence from the agent's environment, including its known actions. As shown in Section 2, some researchers in this area have devised computational systems that use abductive reasoning to support evidence-based conjectures about an automated agent's intentions. A comparable problem in the study of intelligent autonomous agents arises for goals and motives. Motives and intentions are internal mental constructs attributed to an intelligent autonomous agent based on external evidence of its actions, its speech acts, and other events that can be empirically observed and documented. But because motives and intentions are internal, and not directly observable by another agent, it has always been a problem for cognitive science to track the line of evidence-based reasoning to them from these externally observable events. This paper proposes an argumentation-based method of solving the problem using argumentation schemes, argument graphs, and other widely used argument technologies.

The method requires special argumentation schemes shown in the paper to be applicable to problems about using the factual evidence in a case to search for and find explanatory hypotheses about an agent's motives and intentions. The primary scheme is the one for practical reasoning representing an intelligent agent's reasoning from its goal, along with an action that the agent takes to be a means for achieving its goal, to the conclusion that it should move ahead with the action being considered. But here one is confronted with the delicate and highly contestable problem of how to differentiate, in some

[†] This paper was in the proofreading phase when the author Douglas Walton sadly passed away. The proofreading has been completed by the journal editors.

clearly reproducible way, among goals, motives and intentions. The paper provides a solution based on the model of goal-based practical reasoning of Walton [62] supplemented by special multiagent argumentation schemes along with tools available from formal multiagent argumentation systems in artificial intelligence.

The paper extends the model of goal-based practical reasoning of (Walton [62]) along with other research on argumentation schemes (Walton, Reed and Macagno [66]; Gordon, Friedrich and Walton [19]) to show how it can be applied to the problem of deriving hypotheses about the goals, motives or intentions of an intelligent autonomous agent from evidence using an evidence-based pathway of reasoning. This is carried out by applying some existing argumentation schemes to realistic examples, most of which are cases from evidence law, in order to try to figure out how the motives and/or intentions of an agent can be inferred by means of a traceable line of reasoning from the given set of evidential facts in a case.

The inspiration for this paper came from four main sources. One was some earlier work with Giovanni Sartor on teleological argumentation (Walton and Sartor [67]). A second one was the current research on studying schemes of a kind that are especially important in legal argumentation. Particular instances in point are a paper on teleological argumentation to and from motives (Walton [60]) that examined some legal cases using argumentation tools such as argumentation schemes, argument diagrams and the hybrid theory of Bex [6] that applied his theory of scripts and stories to the explanation of motives in evidential legal argumentation (Bex and Walton [8]). A third one was the paper of Walton and Schafer [68] that studied characteristics of evidential reasoning about motives in law. A fourth one was the extensive study of legal reasoning based on motives in many interesting legal examples which Leonard schematized as motive-based "inferential chains" (Leonard [30]).

These earlier works argued that a motive of a rational agent can be found by using inference to the best explanation to extrapolate the chain of argumentation containing goal-based practical reasoning backwards from the given evidence of a legal case to a hypothesis postulating that the agent reasoned forward from a motive to a conclusion to take action. The present paper presents a way of using the same argumentation schemes backwards from the evidence in a given legal case to a hypothesis about the agent's intention. This is an even more difficult task because the concept of intention is highly contested in many disciplines, and has a different meaning in the social sciences, and other fields, and in everyday conversational language, than the accepted meaning it has in law. So a necessary part of the project is to propose tentative definitions based on argumentation theory for the terms motive, intention and goal, using argumentation technology of the kind that is currently being used in artificial intelligence and law.

Terms such as intention, goal and motive are central to theory of mind as used in diverse fields such as philosophy, computer science and psychology. This paper begins with preliminary tentative definitions of these terms based on a new dialectical theory of mind that is not based on the actual beliefs and intentions of a rational agent (BDI model), but instead on the agent's commitments in a multiagent dialogue setting.

The systematic method for finding the goals, motives and/or intentions of an agent built in this paper is meant to be applied to problematic cases about intentions and motives in evidence law, ethics, argumentation and artificial intelligence. The method is applied to some relatively simple (textbook type) examples of a kind familiar in evidence law. The examples of arguments analyzed and visually represented as argument graphs in the paper apply the Carneades Argumentation System (Gordon, Prakken and Walton [20]; Gordon and Walton [21]).

As will be indicated in Section 2, the approach taken here requires assuming that a goal-directed agent of the kind instantiated in legal reasoning about evidence is rational to some extent (bounded rationality), as determined by a set of twenty-six properties defining an intelligent autonomous agent of a kind currently familiar in multiagent systems (Wooldridge [74]). It is shown how goal-directed practical reasoning carried out by such an agent can be structured in a clear and precise way that makes it useful for understanding artificial multiagent systems and applying them to legal reasoning about motives and intentions. Section 3 offers provisional definitions of the terms motive, intention and goal, made more precise in Section 10.

Section 4 analyzes a legal example of argument from evidence to motive showing how the argumentation in the example can be modeled using argumentation schemes along with argument diagramming tools. Section 5 presents the basic argumentation scheme for practical reasoning and explains the argumentation scheme for inference to the best explanation. Carneades (version 4) has these schemes and over 100 (so far 106 in total) available to the system.[1] You can input these schemes in version 2 manually. In version 4 they are put in automatically by the system if one manually includes the schemes in one's YML file. In Section 6 two legal examples of evidence-based reasoning from a knowledge base to a conclusion drawn about a rational agent's intention are extensively analyzed using argument diagrams showing how the argumentation structure incorporates schemes.

Section 7 presents the schemes for argument from motive to action and argument from motive to intention. Based on the discussion in Section 7, a distinction is drawn in Section 12 between two subspecies of each of the generic schemes. One argues that the conclusion more likely than shown by the previous evidence. The other argues that this agent was more likely to have carried out the action than some other agent or agents also be being considered. Based on this distinction, two new sub-schemes are formulated in Section 12.

Section 8 also extends the standard list of argumentation schemes by introducing three new multiagent schemes that track the reasoning in which a second agent plausibly infers the hypothesis that a first agent has a particular motive by using evidence about the first agent's foreseeing of consequences of an action that the second agent knows the first agent is considering carrying out. Section 9 shows with two examples how version 4.3 of Carneades has the capability to recognize a scheme that applies to an argument being interpreted, once the premises have been put in, using two of the schemes concerning motives and intentions. This development is particularly significant because it indicates how an artificial intelligence argumentation system can be used to draw on a knowledge base representing the factual and legal evidence in a given case in order to find motives and intentions in that case. This can be done more effectively with Carneades, because it has 106 programmed schemes available to it, and it has a capability for finding arguments [64].

In Section 11 (conclusions section), the six-step method for finding an intention or motive from an evidential knowledge base is summarized. The method is a general one for argumentation theory because it can be used with different computational systems, or even (less effectively though) without using any. In Section 12, six ways of extending the findings of the paper, by investigating more complex examples, are proposed.

2. Goal-directed reasoning of intelligent rational autonomous agents

Walton and Schafer [68] used argumentation tools such as argument diagramming and argumentation schemes, especially the ones for practical reasoning and abductive reasoning, along with plan recognition systems developed in computer science to build a multiagent model of agent reasoning that can be used

[1] https://raw.githubusercontent.com/carneades/carneades-4/master/examples/AGs/YAML/walton.yml

to infer an agent's goals motives and intentions from its actions. In their plan recognition model, one agent can draw inferences from another agent's observed actions to construct a plausible account of its expected goals and beliefs (Walton and Schafer [68], 9). Their model is dialectical and commitment-based (Hamblin [25], 37–49). One agent ensures that another agent with whom it is engaged in a dialogue has a particular motive by collecting evidence based on what the other agent has said and is known to have done in a database called a commitment set (Hamblin [24,25]; Walton and Krabbe [65]). It is the system that built the foundation for the work in the present paper, along with the multiagent framework of (Wooldridge [72,73]).

The framework representing the reasoning of an autonomous rational agent set out below is based on the HGS (hierarchical goal-seeking system) of Walton and Schafer ([68], 19). In the HGS framework an intelligent agent was defined as an entity capable of practical reason having five capabilities: (1) it can store information in a knowledge base called a commitment set, (2) it can be aware of some of the consequences of its actions, (3) it can take actions in light of its observation of the consequences, (4) it can aid these calculations by having sequences of actions organized into routines stored in its memory and (5) it is capable of executing sequences of practical reasoning from a goal to an action as well as backward sequences of practical reasoning from an action to a goal. Here we extend this framework by defining an intelligent rational agent of this sort in a more detailed way.

An *intelligent rational autonomous agent* (IRAA) is an entity that has the capability to carry out actions, or refrain from doing so, and that might realize one or more of its goals, based on its defeasible knowledge of its circumstances at the time it acts. An IRAA reasons from its goals to its actions and can make plans based on its knowledge of its external circumstances, acting within a group of other IRAA's that can communicate with each other. An agent in this sense can be a machine or a human (or an animal). Minimally, an IRAA is an entity that has the capability of forming goals and the capability of carrying out actions that might realize one or more of its goals, based on the defeasible knowledge of its circumstances at a given point in time. But we will need to add other capabilities. Here are twenty-two properties that can be used to indicate what level of bounded rationality an IRAA is capable of, classified under six general categories.

Actions, Goals and Control.

(1) An IRAA has control over carrying out actions (or refraining from actions) of a kind that can change its circumstances.
(2) An IRAA has goals, can set goals for itself, and can direct its actions based on these goals.
(3) An IRAA can retract or modify its goals, as it might do if it sees that its goals conflict.
(4) An IRAA can grasp how actions to achieve a goal fall into an ordered sequence where some actions are required to carry out others.
(5) An IRAA can organize goals and actions into a hierarchy of levels of abstraction.
(6) An IRAA will generally keep trying to achieve a goal even if it has previously failed (plasticity), unless it has reasons to stop trying.
(7) An IRAA will not continue trying to carry out an action that it knows is impossible.

Knowledge.

(8) An IRAA has the capability for perception and for collecting information from other sources such as reports by witnesses or experts.
(9) By these means an IRAA can find out about its current circumstances.

(10) An IRAA has sufficient resources of memory to retain knowledge of its circumstances as they change over time.
(11) An IRAA needs to have common knowledge about the normal ways things are expected to work generally, and in social institutions.
(12) An IRAA needs to be aware of at least some consequences of its past actions and keep them in memory as knowledge for possible use in the future.

Consequences of Actions.

(13) An IRAA can perceive or find out about the consequences of its actions.
(14) An IRAA can correct its previous or planned actions if it sees that the consequences of those actions are likely to run contrary to its goals.
(15) An IRAA can form hypotheticals about possible future consequences of its actions.
(16) An IRAA often needs to be flexible in planning by quickly adapting to new knowledge about its circumstances as it becomes available (flexibility).

Communication in Group Actions.

(17) An IRAA can be part of a group of intelligent agents that is trying to achieve a common goal, solve a problem or collaborate on deciding what to do.
(18) An IRAA typically needs to communicate with other agents to acquire new knowledge, act together with them and collaboratively solve problems.
(19) As an IRAA communicates by putting forward speech acts, such as ones making assertions or asking questions, and it replies to the speech acts of other agents.

Commitments.

(20) As an IRAA makes speech acts, it incurs commitments that can be recorded.
(21) It can be inferred from some of its speech acts that an IRAA is committed to a proposition, an action or a goal.
(22) An IRAA has the capability to add new commitments to its previous store of commitments and retract commitments as needed.

The idea is that a given agent can exhibit different levels and kinds of capabilities for rational argumentation, depending on which of these twenty-two properties it has or does not have. An IRAA ideally should have all these capabilities. But all we may need to know, in many cases of trying to find the intention of an IRAA, is that it has some of them. Note that the section headed Communication in Group Actions specifically refers to multiagent cases whereas the other sections can refer either to single or multiagent cases.

Since some of the examples in this paper are criminal cases where a factfinder is trying to use evidence to decide whether a defendant had an intention or not, or acted on that intention, and since criminal behavior does not appear to be all that rational in many cases, one could question to what extent using an IRAA-based model is applicable to real people in such cases. However, the concept of a rational person has played a significant role in evidential reasoning in criminal law. Vitiello [57], based on a legal study of the evidence in four criminal cases, "explained the recent controversies surrounding the identity of the reasonable person, and explored the difficult interplay of objective and subjective characteristics of the reasonable person" (Vitiello [57], 1437). The reasonable person, called the 'reasonable man' in the traditional legal literature, is a legal test for responsibility for some consequence of an action that an actual person carried out by asking whether it is one a reasonable person should be expected to foresee.

There is already a substantial literature in psychology on the notions of motive and intention, but because psychology is centrally interested in scientifically modeling empirically how human agents actually think and carry out actions, much of this literature is tangential to the task undertaken here. The purpose of this paper is to aid artificial intelligence in its task of building and computationally implementing models of practical reasoning representing rational deliberation that do not necessarily try to copy how humans reason or act but to devise artificial agents that can carry out tasks that can help humans with practical tasks.

Practical reasoning is typically known in computer science as goal-based reasoning, or goal-directed reasoning (Russell and Norvig [46], 259). There are two theories on how practical reasoning should be modeled. The commitment-based argumentation dialectical approach (Hamblin [24,25]; Walton and Krabbe [65]) is different in certain key respects from the BDI (belief-desire-intention) theory (Bratman [9]; Bratman, Israel and Pollack [11]; Wooldridge [73]; Paglieri and Castelfranchi [41]). Sarkadi et al. [48] and Panisson et al. [42] used the agent communication language KQML, the communication language in the JASON Multiagent Platform, along with speech act theory, to build a computational theory of mind based on the BDI architecture. However, in this paper a dialectical theory is employed.

In the dialectical theory, practical reasoning takes place in a setting where a group of rational agents are collectively taking part in a dialogue in which each contributes speech acts by making dialogue moves. Each move contains a speech act and there can be speech acts of different types. The dialogues also can be of different types (Walton and Krabbe [65]). In all of these types of dialogue when one agent makes a move in the form of a speech act, such as asking a question, the agent to whom the question was directed has to reply to it by making other kinds of allowable moves, for example answering the question, or arguing that asking this question is not relevant at this point. As each of the agents takes its turn making these argumentative moves, according to the rules governing when each of them can speak, or needs to reply, commitments (propositions) are inserted into or retracted from that agent's commitment store. A commitment is incurred when an agent has gone on record as putting forward a speech act that, according to the protocol of the dialogue, carries with it a commitment to whatever has been agreed to by performing that speech act. There can also be retractions of commitments in some instances.

One type of speech act that is often the central focus of attention in the dialectical theory is the putting forward of an argument. When the one agent puts forward an argument, according to the rules (protocol) for this type of dialogue she has to accept the conclusion of the argument if she accepts all the premises and the argument is accepted as a valid one in that type of dialogue. However there are usually alternatives. For example, the other party may reply, by asking an appropriate critical question or by putting forward a relevant counter-argument.

According to the BDI theory, an agent has a set of beliefs that are constantly being updated by sensory input from its environment. From these beliefs, the agent builds up desires (wants) that are then evaluated by desirability and achievability to form intentions. An intention is a persistent goal that is not easily given up. The two models are different because a commitment is not necessarily a belief. Belief implies commitment but not vice versa. Belief is a psychological notion whereas commitment is a dialectical notion based on dialogue rules (Engel [17]).

Practical reasoning is fundamental to artificial intelligence in a search for a solution to a problem carried out by finding a sequence of actions either from the available means or by seeking new means. In this approach, an intelligent goal-seeking agent receives information about its external circumstances and stores it in memory (Reed and Norman [45]). So far we have only considered instrumental practical reasoning of a kind that does not involve values that can be attributed to the agents. However, in addition to the simpler systems of practical goal-based reasoning there are also value-based systems of practical reasoning for multiagent deliberation (Atkinson, Bench-Capon and McBurney [2]; Rahwan and Amgoud [44]; Atkinson and Bench-Capon [1]). In these systems there is an ordering of the values for each agent so that when value conflicts arise they can be resolved by a value ordering.

In Bratman's [9] version of the BDI model, to form an intention to do something, the agent adopts a plan composed of its intentions, desires (wants) and beliefs. This approach makes it clear that practical reasoning in his BDI model is closely tied in with planning. Although Bratman's theory of practical reasoning, since it is based on desires and beliefs as well as intentions, is clearly committed to the BDI model, it often shifts to the language of commitment. (Bratman [9], chapter 7). Searle [51] also claims to be an exponent of the BDI model of practical reasoning, but he also often shifts to the language of the commitment model.

Planning, also called automated planning and scheduling, is a technology used in artificial intelligence based on an intelligent agent having a set of goals and being able to generate a sequence of actions that leads to the fulfillment of one or more of these goals. Solutions, which often resort to trial and error strategies, are found and evaluated prior to the execution point at which the agent carries out the action. Problem solving is another area of computing, but problem solving is similar to planning in that both technologies are based on a practical reasoning framework in which an agent concludes to carrying out an action based on its goals and what it takes to be the means that lead to fulfillment of a particular goal. Although planning and problem solving are considered different subjects, and represent goals and actions in somewhat different ways, the underlying structure of reasoning from goals to actions shared by both technologies is structurally similar (Russell and Norvig [46], 338). In both technologies, an agent has a goal, and uses a search algorithm to try to find a solution in the form of an action sequence that leads to realization of the goal (Russell and Norvig [46], 56).

Bratman's theory of shared agency, he claims (Bratman [10], Preface, ix), "departs in important ways" from the BDI model, but the fact that he continues to use the words 'belief' and 'desire' to describe his theory suggests that (maybe) it is still a BDI model. Perhaps it can be described as a commitment-based model that contains important elements of the BDI model.

Figure 1 displays the procedure whereby an IRAA uses value-based practical reasoning.

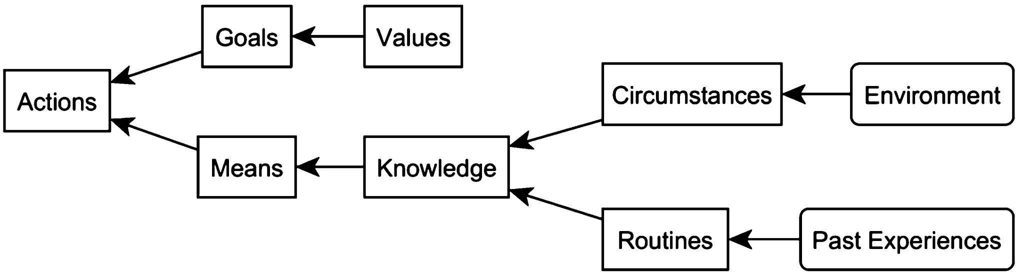

Fig. 1. Goal-directed reasoning of an IRAA.

Figure 1 shows how an IRAA uses practical goal-directed reasoning in the commitment model to conclude to a decision to take action. The means that could be used to carry out the goal is a sequence of actions that the agent knows about from two sources. First, it must know about the circumstances in which it finds itself and in which it needs to act. These can change rapidly as new evidence comes in from its environment. Second, it must be familiar with how things normally go in such a situation, based on what it has learned from past experience. For example, if the IRAA is driving a car, it must know how to slow the speed of the car by applying the brake. The term knowledge is used here in a sense different from its usage in traditional analytical philosophy where knowledge must imply truth. In connection with an IRAA, knowledge refers to information about a domain that is used for solving problems. This kind of knowledge used for real problem-solving must be fallible, because it is based on defeasible generalizations about the way things can generally be expected to go in a situation confronting the agent. In this sense, a knowledge-based system of the kind employed in artificial intelligence is a system that uses this kind of knowledge about a domain to act or solve problems (Poole and Macworth [43], 60).

Norman and Long [40] investigated problems in building automated agents that are able to act in real world circumstances where it is necessary to manage multiple goals as the intelligent agent moves forward to take action based on these goals and the changing circumstances of a given situation. For this purpose they distinguished between the concept of a goal and the concept of a motive as follows (Norman and Long [40], 279). A motive is a driving force that depends on the internal state of an agent and the state of the external world and that arouses and directs action to the achievement of goals. In contrast, a goal denotes a state that the system can achieve or prevent through action. A goal can be set arbitrarily, whereas a motive is not something that the agent selects as a choice, but is just there, representing an interest that the agent has. Norman and Long ([40], 283–287) presented an extended example of an automated warehouse agent that chooses between offers from suppliers of products differing in price, quality, availability, and reliability, and then takes the appropriate actions for ordering, receiving and sending out the selected product.

Research on intention recognition of the kind that has been going on in computer science for the past 30 years or so is described by Sadri [47] as the task of recognizing the intentions of an agent by analyzing its actions and the changes resulting from its actions in the observable environment. Among the applications of this work Sadri includes care of the elderly and assistance for cognitively impaired individuals, terrorism detection by recognizing the intentions of would-be attackers, and military applications recognizing the intentions of enemy maneuvers in order to plan countermeasures. His system, called WIREK (weighted intention recognition based on event calculus) is used to reason about an agent's intentions and actions based on information available in the environment that can make such calculations using evidence-based reasoning. Among the evidence sources are plan libraries containing what is known about the agent's goals and actions providing evidence that can be inputted into a graph search through state changes in a sequence of actions and events.

Meadows, Langley and Emery [37] describe their approach as one of plan understanding, as contrasted with plan recognition, because it involves the explanation of an agent's behavior in terms of an agent's mental states. Plan understanding is taken to be analogous to language understanding in that the analysis provides a connected account of the input. However instead of being a sequence of words, the observation that is taken as input is a sequence of relational actions and events as connected to mental states of an agent planning to carry out actions. They see plan understanding as an abductive procedure that involves the generation of plausible explanations based on an agent's beliefs and goals and on relevant external evidence. Their implemented system, called UMBRA, addresses the task of generating reasonable abductive inferences that explain an agent's observed actions by drawing conclusions about the agents presumed internal states, such as its goals, motives and intentions.

3. Defining motives, intentions and goals

In the BDI (belief-desire-intention) approach, an intention is generally taken to be a goal that the intelligent agent is particularly motivated to try to bring about (Bratman, Israel and Pollack [11]; Paglieri and Castelfranchi [41]; Tuomela [55]). Following this approach, an *intention* could be defined as a species of goal that the agent is strongly enough committed to on a continuing basis so that it acts as a force biasing the agent's practical reasoning in a particular direction. In this paper we will take a different approach that is based on the commitment model but considers how this model could be extended to accommodate some BDI features by examining how intention and motive are defined in law and artificial intelligence.

Criminal intent is an element of most crimes, according to an American Bar Association document outlining best practices in proving intent in criminal and civil law (Stover [54], 1). For example, assault requires intent to commit a battery, burglary requires intent to commit a felony in the dwelling, forgery requires intent to defraud, and tax evasion requires intent to violate the tax laws (Stover [54], 4). Proving intent in law is often difficult because an agent's intention is a mental element and by its nature private. Very often direct evidence of intent, for example by a confession from the accused, is not available. Therefore proving intent is usually a matter of piecing together different kinds of circumstantial evidence such as emails, public statements, recollections of participants who attend meetings, telephone call logs, and testimony from undercover officers (Stover [54], 5). Typically in legal argumentation, this kind of evidence is in turn supported or undermined by other evidence, creating a mass of evidence supporting or refuting some ultimate claim at issue in a trial or other legal setting.

One clue to drawing a provisional distinction between motive and intention is a remark in (Kramer [29], 2) stating that it is generally sufficient to define intent in law as a voluntary act knowingly done. This can be contrasted with the figurative definition of motive given in the case of United States v. Benton, 637 F.2d 1052, 1056–57 (5th Cir. 1981) as "the reason that nudges the will and prods the mind to indulge the criminal intent."

Interestingly, both definitions are fairly consistent with definitions of the same terms that can be found in artificial intelligence research on autonomous agents. The legal definition of motive at least partly fits the definition of motive of Norman and Long ([40], 279) as a driving force that depends on the internal state of an agent and the state of the external world and that arouses and directs action to the achievement of goals. Moreover, the legal definition of intent given above fits very well with the definition of intention given in a research monograph on multiagent systems citing writings on philosophy (Dunin-Keplicz and Verbrugge [15], 30):

> The key concept in the theory of practical reasoning is the one of *intention*, studied in depth in Bratman [9]. Intentions form a rather special consistent subset of goals that the agent wants to focus on for the time being. According to Cohen and Levesque [12], intention consists of choice together with commitment (in a non-technical sense). In our approach these two ingredients are separated: an intention is viewed as a chosen goal, providing inspiration for a more concrete social (pairwise) commitment in the individual case, and a plan-based collective commitment in the group case.

The theory of Cohen and Levesque [12] has proved to be useful for reasoning about autonomous agents and has been widely adopted in computational research on multiagent systems. They identified seven properties that must be satisfied by a reasonable theory to define the notion of an intention of a rational autonomous agent:

1. Intentions pose problems for agents, who need to determine ways of achieving them.

2. Intentions provide a ‘filter’ for adopting other intentions, which must not conflict.
3. Agents track the success of their intentions, and are inclined to try again if their attempts fail.
4. Agents believe their intentions are possible.
5. Under certain circumstances, agents believe they will bring about their intentions.
6. Agents do not believe they will not bring about their intentions.
7. Agents need not intend all the expected side effects of their intentions.

The properties of intentions cited from Cohen and Levesque [12] misrepresent intentions by talking about achieving them or bringing them about. People and other agents to do not achieve or bring about intentions, but rather carry them out, as the definition of intention proposed in Section 10 of this paper indicates. An intention is defined in Section 10 as a kind of internal commitment to do something, which may or may not be accompanied by steps to carry out the intention. An intention is thus better described as a future action that an agent has decided to perform and may already be taking steps to carry out. Their approach suggests the following definition of the notion of an intention that will be adopted in Section 10. An *intention* is a commitment that a rational agent is not only committed to, but has autonomously selected to move ahead with or carry out by some means that is available to it. Autonomy in this sense is defined here by characteristics 1, 2, 3, 4, 13, and 14 of an IRAA.

In contrast, *motive* was defined by Norman and Long [40] as a driving force from agent to action that depends on the internal goals and interests of an agent and the external circumstances known to the agent, which directs it towards actions that appear to the agent to lead to the achievement of its goals or fulfillment of its interests. So defined, motives are linked to goals by depending on them. A *goal* is defined in this paper as commitment that is part of an agent’s plan that the agent can achieve or prevent through action (Walton [62]). Typically, a rational agent’s goal is a state of affairs in the external world, like eating at a restaurant or winning a race. Using a combination of practical reasoning and abductive reasoning (defined in Section 5), a rational agent’s goal can often be inferred (if there is sufficient evidence) by reasoning backwards from its actions and statements that it has gone on record as being committed to. This procedure is called finding a goal (see Section 11).

A goal can be set arbitrarily, whereas a motive is not something that the agent selects as a choice, but is just there, representing an interest that the agent has. One key difference, in Norman and Long’s account, is that a goal may be satisfied by the agent’s achieving the state in question, whereas a motive is a force biasing the agent’s practical reasoning in a particular direction.

Leonard ([30], 445) has provided a useful historical survey of the various ways that the concept of motive has been defined by legal scholars. Some describe it as an emotion or state of mind that prompts a person in a particular way, or as a moving power which impels an agent from an action to a result. Another source notes that emotions such as hostility and jealousy are motives since they lead agents to act in a particular way. Others distinguish between having a motive and possessing a plan. What is especially interesting about Leonard’s survey of these legal usages is that there is an overlap between having a motive and having an intention to act in a certain way, and also an overlap between these two concepts and that of developing a plan to carry out a particular action or achieve a particular goal. Leonard holds that all three types of legal locutions have the element in common that there is an inference that flows from an initial reason to some conclusion such as an action, motive or intention. He offers the example of a person charged with arson who is alleged to have burned a building in order to collect insurance benefits. Such a person could be said to have a motive to burn the building, and therefore it can be inferred that he had a plan to burn the building. But he adds that this kind of evidence would also demonstrate that the person acted with intent rather than by accident. Moreover, he adds that the same kind of evidence would be legally admissible on any of these three theories.

It should also be noted that the term 'motive' is ambiguous (Walton [60], 219). Wigmore ([70], 146) illustrated this ambiguity with the example of a case where a defendant is accused of burning down the plaintiff's house. The plaintiff argued that the defendant's motive for burning down the plaintiff's house was his prior prosecution of the defendant in a lawsuit. But others might describe the plaintiff's motive as his hostile and vindictive emotion arising from the lawsuit. Thus motives can be defined as potentially action-initiating emotions, such as hunger or thirst, lust or revenge. However in other instances, motives are defined in a different way by using infinitive phrases such as 'to keep peace in the family' or 'to prevent him from revealing to the police that I stole a car'. Defined in this second way, a motive is taken to be a goal attributed to an agent that explains why the agent carried out an action. The first way of defining the notion of motive appears to make it fit in better with the BDI model, because in that model, desires, which could be emotions, are held to be the mainsprings of an action. The second way of defining the notion of motive appears to make it fit in better with the argumentation scheme for practical reasoning in a framework of goal-based action.

These provisional definitions give us a starting point for moving forward by using argumentation tools, and in particular certain argumentation schemes, to provide the fundamentals of an evidence-based method of abductive reasoning that can be used to prove or disprove claims about the goals, motives and intentions of an intelligent autonomous agent. In Section 10, deeper dialectical definitions of these two key terms will be proposed to make them more useful for computational argumentation.

4. Leonard's hypothetical example

In his substantial contribution to the subject, Leonard did not use the word 'argumentation' or show any evidence of referring to argumentation theory. Nevertheless, his inferential chains connecting motives, intentions and actions look (as shown by the analyses of these examples in this paper) are in outline similar to what argumentation schemes generally look like. They use variables, and each scheme is made up of a set or sequence of propositions linking premises to a conclusion. For example, in (Leonard [30], 460), the conclusion that D's (the defendant's) killing of V (the victim) was intentional is taken to be based on what Leonard called a motive inference (D had a motive to kill V), and a generalization (the proposition that normal people with a motive to kill a person are more likely to have killed that person intentionally than people who would kill that person without a motive to do so). According to Leonard (460) this kind of reasoning is "plausible" meaning that the premises make the conclusion somewhat more likely than it would be without the evidence provided by the premises. Leonard did not call such an inference an argument but he used variables in the forms of inference to apply them to particular legal examples relating to evidence. So they are not schemes as we know them. But it seems many of these inferences could be refashioned or adapted into fitting what is now known as an argumentation scheme (of some sort, not identified by Leonard).

Walton and Schafer ([68], 26–29) used several legal examples for their analysis, one of which was Leonard's car theft example (Leonard [30], 449). In this example, the defendant admitted killing the victim, but claimed the killing was an accident. In the past, the defendant had stolen a car. The victim was aware of the theft, and had threatened to inform the police. In this example, the evidence of the threat to reveal the defendant's car theft suggests a conclusion about the defendant's motive, to avoid getting a prison sentence.

> Suppose D is charged with the murder of V. D claims not to have been involved. If the prosecution possesses evidence that prior to the killing, D had been involved in a car theft, that V had learned

> about the theft, and that *V* had threatened to reveal the theft to the police, evidence of the theft could be admissible to prove *D*'s motive, and from that, *D*'s possible behavior.

In his analysis of the reasoning used in the car theft example, Leonard ([30], 448) identified three inferences forming parts of the chain of reasoning leading from the circumstantial evidence to different aspects of *D*'s action. Here is the first inference.

> EVIDENCE: *D* stole a car, *V* was aware of the fact, and *V* threatened to inform the police.
> INFERENCE: *D* had a motive to prevent *V* from revealing the theft to the police.
> CONCLUSION: *D* murdered *V* to prevent *V* from revealing the theft to the police.

The conclusion of this inference represents a kind of goal-based reasoning typically called practical reasoning in argumentation theory. On this construal, the conclusion can be analyzed as stating that *D*'s goal was to prevent *V* from revealing the theft to the police, and to realize that goal he took the action of murdering *V*. But if the crime of murder requires intent as one of its elements, the conclusion could also be interpreted as a statement about his intentions as well as his actions. Leonard commented that the same step of reasoning could also be used to tend to show that a killing took place. With a different conclusion in place, the first step of reasoning could be reconfigured as follows. Here is the variant on the first inference.

> EVIDENCE: *D* stole a car, *V* was aware of the fact, and *V* threatened to inform the police.
> INFERENCE: *D* had a motive to prevent *V* from revealing the theft to police.
> CONCLUSION: *D* killed *V*.

If the first step of reasoning is interpreted in this way, it could be described as an inference from the agent's motive, along with attendant circumstances, to the agent's action. This type of inference appears quite interesting from the point of view of argumentation theory, because it suggests a general form of reasoning from a motive to an action.

Leonard's second inference is fairly close in its structure to the variant on the first inference just above, and brings in some other complications which are not necessary to this paper, so let us pass on to the third inference which is more interesting from a point of view of argumentation theory. Leonard ([30], 449) structured this third kind of inference as follows.

> EVIDENCE: *D* stole a car, *V* was aware of the fact, and *V* threatened to inform the police.
> INFERENCE: *D* had a motive to prevent *V* from revealing the theft to police.
> CONCLUSION: *D* purposely killed *V* to prevent *V* from revealing the theft to police.

Assuming that the word 'purposely' can be taken to be equivalent to the word 'intentionally', for the purpose of this paper, the above inference can be taken to represent a species of argumentation that goes from a motive premise to a conclusion about the agent's intention. Leonard's further remarks (449) about the car theft example suggest that the word 'intentionally' would be appropriate.

> In the hypothetical case, if *D* admitted killing *V* but claimed that the killing was an accident, the theft evidence could be admissible against *D* to prove that *D* intended to kill *V*. The evidence of *V*'s threat to reveal *D*'s auto theft would give rise to an inference of a motive to prevent the revelation, and the existence of a motive would suggest that in killing *V*, *D* acted intentionally rather than by accident.

So here Leonard has identified three types of inference concerning motives and intentions that will be of special interest when we come to consider argumentation schemes in this paper.

Leonard [30] noted that his descriptions of the reasoning in his hypothetical case are based on an actual case, that of United States v Clark 9080 8F 2-D 1459, 14656 circuit 1993, indicating that the defendant's involvement with the theft activity is probative of defendant's motive and intent. He also described numerous other legal cases where inferences of the kind he applied in this case can also be applied to these other cases where evidence-based reasoning that is probative of a defendant's motive and intent are involved. One of these examples (Leonard [30], 484) is the case of Gibbs v. State (300 S.W.2d890 Tenn. 1957) where the defendant was charged with the murder of a woman, and the prosecution offered evidence of two other killings, the killing of the woman's husband and one of the couple's daughters. The prosecution argued that the killing of the husband provided the motive to conceal the defendant's identity as the killer of the wife. These cases are very interesting with respect to argumentation theory and it would be a good research project to model them with argument diagrams, argumentation schemes and the other kinds of argument technology explained in this paper.

Although the car theft example is a hypothetical case, Leonard also described and commented on several "real" legal cases. One of these cases (Commonwealth v. McCarthy: 119Mass. 354, 1876) from (Leonard [30], 475) can also be used to illustrate the scheme for argument from motive to intention. In this case the defendant was charged with arson, after it was alleged that he set a fire in a shed close to a building owned by a man named Gleason. The defendant had kept property in Gleason's building that he had insured for more than its value. The prosecution argued that the defendant had set the fire in order to collect the insurance money. On this basis, the prosecution alleged that Gleason intentionally set the fire.

Leonard ([30], 460) also described the structure of the inferential chain from motive to intention in a different but also interesting way when he used it to illustrate the kind of case where the defendant admits the act at issue but denies the essential mental element.

> MOTIVE INFERENCE: *D* had a motive to kill *V*.
> GENERALIZATION: Normal people with a motive to kill a person are more likely to have killed that person intentionally than people who kill that person without a motive to do so.
> CONCLUSION: *D*'s killing of *V* was intentional.

Leonard described this sequence of reasoning as representing a plausible form of inference (460), and it is easy to see why it fits that category because the generalization is about how normal people are more or less likely to have acted in a certain way depending on whether they have a particular motive or not.

The whole chain of argumentation in Leonard's hypothetical example depends on implicit premises which can be put in place, and need to be put in place by a jury or other audience based on their common knowledge shared with the victim and the defendant. One way of approaching the example would be to use the standard technique of enthymematic argument reconstruction (EAR) of the kind familiar in argumentation theory. For example, one might try to list the needed implicit premises along with the explicitly given premises in a key list and then to make an argument diagram of the entire sequence of argumentation. Each implicit premise is marked with an asterisk (*).

> Premise 1: *V* knew that *D* had stolen a car.
> Premise 2: *V* had threatened to inform the police about the theft.
> *Premise 3: If *V* informed the police there would be negative consequences for *D*.
> *Premise 4: If *V* did not inform the police, the negative consequences for *D* could be avoided.
> *Premise 5: Killing *V* would prevent *V* from informing the police.
> *Premise 6: Avoiding a negative consequence is a motive for doing (or not doing) something.

Conclusion 1: D had a motive for killing V.
Conclusion 2: D intentionally killed V to prevent V from informing the police of the theft.

Following the analysis above, it could be said that D's motive was to prevent negative consequences that he believed would otherwise likely happen. D's motive was to prevent V from revealing the theft to the police. We are not told why the police being informed is a negative consequence for D, but based on common knowledge on how these things normally work, it can be presumed that he thought that something seen as negative from his point of view would happen if the police were to learn about the theft. In a real case there would likely be factual evidence giving more information about why D was worried about this outcome.

At this point we will present an argument diagram to show how the current argumentation technology analyzes the argumentation in such cases. It is at this point where the devices of the key list and the argument diagram are introduced. A key list gives short forms for each of the propositions in the given argument.

Key List for the Car Theft Example.

VKnew: V knew that the defendant had stolen a car.
VThreat: V had threatened to inform the police about the theft.
NegCon: If V informed the police there would be negative consequences for D.
NotInform: If V did not inform the police, the negative consequences for D could be avoided.
KillPrevent: Killing V would prevent V from informing the police.
AvoidIsMotive: Avoiding a negative consequence is a motive for doing (or not doing) something.
Motive: D had a motive for killing V.
Intent: D intentionally killed V to prevent V from informing the police of the theft.

The problem now is to use the current argumentation technology to model the sequence of reasoning in this example leading to a conclusion about the defendant's motive. There are many different argumentation diagramming tools available, and also several formal argumentation systems in AI that incorporate the use of graphs that are essentially argument diagrams. In this paper we chose the graphical user interface of the Carneades Argumentation System. Carneades is a mathematical model consisting of mathematical structures and functions on these structures (Gordon [18]; Gordon and Walton [21]). It is also a computational model with a graphical user interface.[2]

The code for the first implementation of Carneades was written in 2006–2008. Carneades has been implemented in four main versions. Figures 2–9 in the paper are fairly typical argument diagrams of the kind produced by version 2 (2011). Version 2 is a desktop application called the Carneades Editor. This version is easy to use and is available free but the argument diagrams are a little messy and do not have a resolution that makes them easy to read when printed. Therefore these figures have been redrawn manually using the Graphml editor yEd. The last two figures in the paper were produced by version 4.3 using the editor yEd. Carneades produced graphml directly, even though it was edited in yEd afterward.

The third version was a multi-user web-application developed as a prototype application for analyzing licensing properties of open source software. The fourth version is based on a different formal model of argument (Gordon and Walton [21]) that provides improved support for practical reasoning and multi-criteria decision analysis. The source code of all four versions can be accessed on the Internet.[3] For consistency and ease of exposition mainly argument diagrams in the style of version 2 were used in this

[2] https://carneades.github.io/
[3] https://github.com/carneades

paper, but two of the schemes were diagrammed (Figs 9 and 10) using version 4.3 in a way that makes clear to the reader how all the schemes work (in Section 9). Both versions have the capability for argument invention (Walton and Gordon [64]) and both versions share the same underlying formal structure. The most popular 25 of Walton's argumentation schemes are already available in yml in Carneades 4.[4] However, in the compendium of schemes in (Walton, Macagno and Reed [66], chapter 9), there is a total of 106 argumentation schemes. Başak Kurtuldu has programmed the remaining 81 schemes of (Walton, Reed and Macagno [66]) for Carneades 4 and published them at GitHub.[5] Some of the schemes that Başak translated for Carneades were Hastings, Lorentzen, or Perelman schemes that were also described in the standard source (Walton, Reed and Macagno [66]).

The example graph in Fig. 2 was drawn in the style of version 2. Rectangular nodes contain propositions that are premises or conclusions in a connected sequence of argumentation. The round nodes indicate arguments.

It would be interesting to use Carneades to model the argumentation in a real case that has been tried, for example one or more of the real cases described by Leonard, even though it would require using a very large Wigmore-style diagram to represent all the relevant evidence in the case. But actually Leonard's hypothetical example can be used to show in a general way how the procedure of modelling any such case with Carneades would work, as shown in Fig. 2. Note that in Fig. 2 the explicit premises are shown as propositions contained in rectangles with a solid border. Implicit premises are shown in rectangles with a dashed (broken) border. Making any argument diagram of a real case in this manner requires choosing an interpretation of the argumentation in the case based on common knowledge, as will be explained in Section 6.

To carry this reconstruction out, we will add two items to the key list above.

> *DmurV*: D murdered V.
> *Ddenies*: D denies being the person who murdered V (see Leonard [30], 459).

In general, in any real case decided at trial, there will be a mass of evidence on both sides and to model the argumentation in such a case it will need to be shown how arguments based on this evidence leads to other arguments that support or attack the ultimate *probandum*. For this reason we have put eight rectangular nodes in the argument graph representing points where factual evidence will generally be needed to attack or support the other rectangular nodes in the diagram. These nodes represent premises or conclusions that will need to be backed up by this factual evidence in order to represent the whole network of evidence and argumentation that needs to be weighed in order to arrive at a conclusion on the issue of whether D murdered V or not.

Leonard's example is a hypothetical one, and for that reason we are free to extend it by adding some other assumptions if they will help us to illustrate the point being made. As shown in Fig. 2, argument a7 is backed up by some evidence that is not specified in the hypothetical original example. However, for purposes of illustration, we could specify that proposition by inserting the proposition 'D knew that if the police found out about his theft it would be a second offense for him, which would lead to a lengthy prison sentence'. By briefly extending the example in this way, an idea is given to the reader of what kind of evidence is in the unspecified evidence rectangles shown in Fig. 2.

An issue will be decided in a criminal case on the basis of the standard of proof of beyond reasonable doubt. To win, the prosecution has to show beyond reasonable doubt that D murdered V. To win, the

[4] walton.yml

[5] https://github.com/basakkurtuldu/walton-argumentation-schemes-for-carneades-4

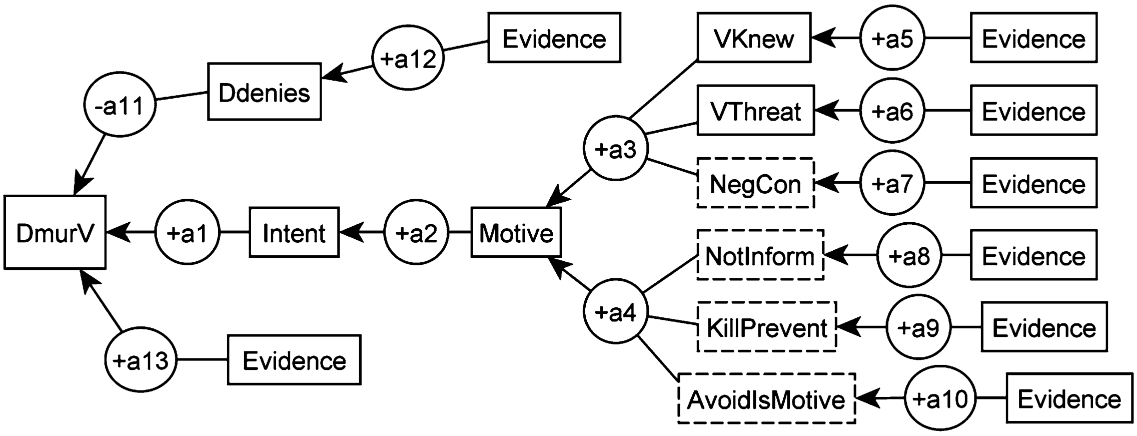

Fig. 2. The structure of argumentation from motive to intent in a typical case.

defense has to show that the prosecution's case that has been made is too weak to prove the conclusion that *D* murdered *V* beyond reasonable doubt. Carneades uses burdens and standards of proof to weigh the strength of the argumentation on one side against the argumentation of the other side using argument graphs generally fitting the kind of structure illustrated in Fig. 2. The two main arguments supporting the conclusion that *D* had a motive for killing *V* are a3 and a4. Each of these arguments is displayed as a linked argument where the premises go together to support the conclusion. In this type of argument, as treated by Carneades, each of the premises needs to be supported by enough evidence to prove the conclusion in order for the conclusion to be accepted.

The reader is here reminded to observe the convention of putting the explicitly stated propositions that function as premises or conclusions in the argumentation by placing them in rectangular nodes with solid borders, in contrast with implicit propositions marked with dashed (broken borders). This convention will be explained and justified at greater length in conjunction with the argument map shown in Fig. 4.

Argument a3 takes the form of a very common kind of reasoning called *argumentum ad consequentiam*, or argument from consequences (Walton, Reed and Macagno [66], 332). Argument from consequences has both a positive and negative form. In argument from positive consequences, a policy or course of action is supported by citing good consequences of carrying it out. In argument from negative consequences, a policy or course of action is rejected by citing bad consequences of carrying it out. In the format stated below, the scheme for argument from negative consequences is stated first, followed by the scheme for argument from positive consequences.

Major Premise: If *A* is brought about, then consequences *C* will occur.
Minor Premise: Consequences *C* are bad.
Conclusion: Therefore *A* should not be brought about.

Major Premise: If *A* is brought about, then consequences *C* will occur.
Minor Premise: Consequences *C* are good.
Conclusion: Therefore *A* should be brought about.

The terms 'good' and 'bad', or what are taken to be their equivalents, 'positive' and 'negative', indicate that both of these schemes are based on a prior scheme called argument from values (Walton et al. [66], 321).

Now let's take a closer look at the question of whether argument a2 fits the scheme for argument from negative consequences. At first it looks like it does, because it is implicit in the argument that if V were to inform the police about D's theft, some negative consequences of the police receiving this information might normally be anticipated. But if you consider the argument more carefully, it goes from the premises that if the victim informed the police about his crime, the defendant would get a prison sentence, and the other premise in the linked argument, the proposition that getting a prison sentence is a negative consequence for the defendant, to the conclusion that the defendant would look for a means to prevent the victim from informing the police. This conclusion does not match the conclusion specified in the scheme for argument from negative consequences. On the contrary, it seems to be a prediction or presumption about what the defendant would or might plausibly do. It is telling us about something that might naturally be in the defendant's mind. It is a kind of conjecture related to the defendant's motive.

For other reasons as well, this argument is not straightforward to analyze using the current resources of argumentation theory because it has several features that are problematic to model. First of all, some of the premises are statements concerning what the defendant and the victim could be taken to know. For this reason, to a considerable extent, the argumentation is about the internal mental states of the victim and especially of the defendant. It is like we need to get a picture of what was going on in the defendant's mind as he reasoned his way through how to act, in relation to his goals and his perceived circumstances of the case with respect to what is possible and what is not. There appear to be many different ways the argumentation in the example could be analyzed, depending on how we reconstruct the state of mind of the defendant, as far as we can surmise it from the factual evidence of the case.

Although the argumentation is clearly about goal-directed reasoning or means-end reasoning, traditionally called practical reasoning, it is not one of the more straightforward kind in which an intelligent automated agent moves forward to conclude to carrying out an action based on its goals and its perceptions of what are the best means to achieve these goals. The Carneades Argumentation System uses argumentation schemes of two particularly central types to model the argumentation in this kind of case. So next we need to review these two schemes.

5. Practical and abductive reasoning

According to the BDI model of practical reasoning (Bratman [9]; Bratman, Israel, & Pollack [11]; Wooldridge [73]; Paglieri & Castelfranchi [41]), an agent possesses a set of beliefs that are continually being updated by sensory input, and a set of desires that are evaluated to form intentions. In the commitment model of practical reasoning, each agent has a commitment set containing the propositions about its circumstances that it has accepted. As each move is made, commitments are inserted into or retracted from this set (Hamblin [24], 1971). In the commitment-based approach, practical reasoning is modeled in a dialogue format in which IRAAs communicate with each other using orderly sequences of speech acts and argumentation schemes such as the one for practical reasoning (Walton [62]).

Stripped to its basics, the simplest form of the commitment-based scheme has two premises and a single conclusion (Walton, Reed, & Macagno [66], 323). The "I" in this scheme stands for an IRAA that looks forward to the future and tries to carry out the goals in its current plan.

Goal Premise: I have a goal, G.
Means Premise: Carrying out this action A is a means to realize G.
Conclusion: I ought (practically speaking) to carry out this action A.

Using this stripped-down heuristic version of the scheme for practical reasoning, the IRAA jumps quickly from its goal (assuming simplistically that it has only one at the moment), and finding some available means to achieve this goal, it immediately concludes to carrying out the action and does so. This way of proceeding could be rational in some instances, but in others is could be classified as jumping to a conclusion, a species of fallacious reasoning.

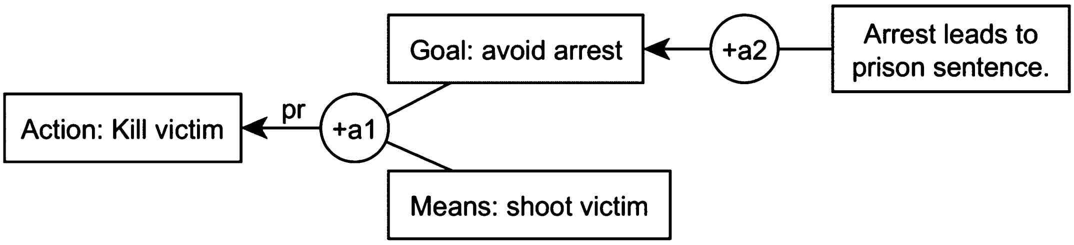

Fig. 3. Defendant's practical reasoning in the car theft example.

Another complication is that the selected means can be either a necessary condition or a sufficient condition. Von Wright [58] used the following example to illustrate necessary condition practical reasoning: X wants to reach the train on time; unless X runs he will not reach the train on time; therefore, X must run. Audi ([3], 87) offered the following example of a sufficient condition scheme: "I really need a peaceful visit in the country", the unstated assumption being that a peaceful visit to the country will be sufficient for what I really need. Yet another complication is that there can be a chain of practical reasoning combining instances of the necessary condition scheme with instances of the sufficient condition scheme.

Groups of agents need to act together using practical reasoning to carry out their shared goals (Tuomela [55]). For this purpose, multiagent groups of agents need to communicate with each other to form plans and agree on goals (Walton [62]). Groups of intelligent autonomous agents that form teams, or even swarms, such as an ant colony, need to have collective goals that all the agents in the group are committed to. It is important to emphasize therefore, that not only will each agent in the group have its individual goals, but groups of agents engaging in coordinated activities can be committed to a collective goal.

For such agents to communicate with each other, the speech acts by which the agents communicate with other agents in a group need to depend not just on individual goals but also on collective goals that the group is committed to. When the speaker performs the speech act of making an assertion, this needs to be seen as more than just an attempt to communicate its belief about some proposition to another agent. It also needs to take into account the communicative purpose of the utterance in relation to the type of conversational interaction in which it occurs (Macagno and Walton [35]). Therefore to understand goal-directed communication it is necessary to take into account not just the individual actions as required by traditional speech act theory, but also the social actions performed by the IRAAs.

When the agents move forward to deliberate on what is the best course of action for them to take using practical reasoning, each agent will often have to try to figure out what the goals and intentions, and other mental states of the other agents, can be reasonably presumed to be, based on the evidence available at the time. Even in order to communicate in such deliberations, it is necessary for the agents to engage in group dialogues in which they take turns putting forward speech acts, such as asking questions, where another agent has to make conjectures about what is the best way to interpret an agent's speech act in order to give a useful reply. To take into account how such utterances need to be understood in relation to communicative contexts in which there is common goal pursued by the IRAAs in a group, Walton and Krabbe [65] distinguished six types of dialogue representing common types of social goal-directed argumentation: persuasion dialogue, negotiation, inquiry, deliberation, information seeking and eristic (quarrelsome) dialogue. For the purposes of this paper, the most important of these is the formal model of deliberation dialogue presented in (Walton, Toniolo and Norman [69]).

Agents often communicate with each other using indirect speech acts (Grice [23]). For example, if one agent says to the other 'Can you pass the salt?' the responding agent needs to recognize that the other agent is not asking a yes-no question about his or her ability to pass the salt, but is making a polite request to pass the salt (McRoy and Hirst [36]; Macagno and Walton [34]). In current linguistic theory, the capability for effective communication necessary to avoid the negative consequences that can occur from miscommunication in such cases is about the respondent arriving at a reasonable conjecture about what the speaker's intention should be taken to be. Following the approach of Macagno and Walton [33] arriving at such a reasonable conjecture about how to interpret an agent's message based on its speech acts needs to be carried out by using what is called abductive reasoning or inference to the best explanation.

The theory of abductive reasoning in (Walton [59]) is built around a different model of explanation from the traditional Hempelian one in philosophy which models explanation as deduction (or induction) from general laws (Miller [38]). This new dialectical approach models an explanation as a dialogue between two agents where one IRAA is presumed by another to understand something and the other agent asks a question meant to enable it to come to also understand it. The model is based on the view of explanation articulated by Scriven ([50], p. 49): "Explanation is literally and logically the process of filling in gaps in understanding, and to do this we must start out with some understanding of something." The dialogue theory of explanation is built on the notion that one agent can come to understand the commitments of the other party by a question-reply dialogue. For these reasons, the theory of intentions of IRAAs in this paper is based on practical goal-based reasoning but is also dialectical (see Section 10).

It is well known that Peirce modeled abductive inference as inference to the best explanation and showed how it represents a form of reasoning used in science. But evidence law also offers examples of abductive reasoning where a conclusion is defeasibly inferred from an evidential fact to a plausible explanation of it. Wigmore ([71], 418) in discussing a legal case quoted a passage from (Sidgwick [52]): "By the best explanation is meant... that solitary one out of all possible hypotheses which, while explaining all the facts already in view, is narrowed, limited, hedged, or qualified, sufficiently to guard in the best possible way against undiscovered exceptions". Wigmore not only picked up this idea from Sidgwick but also used an argument diagram tool he called an "evidence chart" to balance the mass of evidence in a legal case by mapping out the argumentation on both sides.

Josephson and Josephson ([27], 14) studied the uses of abductive reasoning in science and medical decision-making using this scheme for abductive reasoning.

H is a hypothesis.
D is a collection of data.
H explains D.
No other hypothesis can explain D as well as H does.

Therefore H is probably true.

Elsewhere in the literature different schemes for abductive reasoning have been proposed (Walton, Reed and Macagno [66]). But the Josephsons' format strips the scheme down to its basics and works well here. In their account ([27], 14), the judgment of probability of the conclusion following from an abductive inference can be estimated by taking six factors into account.

1. how decisively H surpasses the alternatives
2. how good H is by itself, independently of considering the alternatives (we should be cautious about accepting a hypothesis, even if it is clearly the best one we have, if it is not sufficiently plausible in itself)
3. judgments of the reliability of the data
4. how much confidence there is that all plausible explanations have been considered (how thorough was the search for alternative explanations)

They add two additional considerations: (5) pragmatic considerations, including the costs of being wrong, and the benefits of being right, and (6) how strong the need is to come to a conclusion at all, especially considering the possibility of seeking further evidence before deciding. These six factors work very well as critical questions matching the scheme.

In their account, the conclusion is a tentative presumption that turns on which is the "best" explanation at some given point in a collection of evidence that should remain open to new evidence. But the process of collecting evidence may not be finished. New evidence may suggest a new explanation that may turn out to be better than the one inferred from the existing evidence. Hence the scheme for abductive reasoning can be evaluated using the six critical questions as parts of a dialogue in which understanding can be transferred from one rational agent to another.

6. Argument from evidence to an intention

Here we take what appears to be a very simple and ordinary example of a familiar kind of set of circumstances where I (a rational agent) am walking down the sidewalk and reach an intersection where I see a car approaching along the street I am about to cross. I see that the car's left signal light is flashing and I infer that the driver is intending to make a left turn. I see that the car is moving to the left lane and this is another piece of evidence supporting my hypothesis that the driver intends to make a left turn. I draw the conclusion that the driver's intention is to make a left turn. We all confront this kind of situation on a daily basis where we need to use reasoning to make a rational decision on how to proceed based on evidence that we see and know. The following sequence of reasoning in this kind of situation can be represented using a Carneades-style argument map. To derive an interpretation of the line of reasoning used to arrive at the conclusion about the driver's intention, the first step using Carneades is to identify the propositions making up the premises and conclusions in the sequence of argumentation in the case by building a key list.

Key list for the driver example.

Intends: The driver intends to signal a left turn.
MovingTo: The car is moving to the left turn lane.
SeeCar: I see the car moving to the left turn lane.
**Activated*: The driver activated the left turn signal indicator.
**NoBetterActivate*: There is no better explanation of his having activated the left turn indicator.
**ExplanActivate*: A satisfactory explanation of his having activated the left turn indicator is that he intends to signal a left turn.
LeftFlash: The left turn signal light is starting to flash.

**NoBetter Flashing*: There is no better explanation of the flashing signal light.
**ExplanFlashing*: A satisfactory explanation of the flashing signal light is that the driver activated the turn indicator.
See: I see that the signal light is starting to flash.
**Normal*: The normal way for a driver to signal a turn is to activate the turn indicator.

Before we can build an argument diagram from the key list, we have to draw a distinction between the propositions explicitly stated by sentences in the text of the example and some other propositions that need to be inserted as missing premises or conclusions for the sequence of reasoning to make sense. This distinction was applied in the key list above by prefixing an * mark to each proposition that is an implicit premise or conclusion. The remaining four propositions are the explicit premises or conclusions in the original argument.

But before we can proceed to making up the argument map, we need to see how the reader of the example can extract certain assumptions based on common knowledge of how things generally work in this kind of situation that need to be used as implicit premises or conclusions in the chain of reasoning. We need to do this in order to understand how the kind of abductive reasoning typically used in such cases makes sense to us as rational agents who have a common sense knowledge base about driving that can interact with goals and practical reasoning.

Common knowledge (Minsky [39]; Schank and Abelson [49]; Hosseini et al. [26]) is used as a device to reconstruct argumentation by finding implicit assumptions in a natural language argument of the kind illustrated by the driver example. For a more complex legal example that nevertheless illustrates this procedure more fully, the reader can look to the interpretation of the example of the weak and strong man (Walton [63]). In this example of an assault case, the weaker man argues that it is not plausible that he would assault the stronger man. The stronger man replies that it is not plausible that he would attack the weaker man because it would look bad for him if the case went to trial. In (Walton [63]), three interpretations of this classic example are presented using eight Carneades-style argument diagrams to mark implicit premises and conclusions. As the analysis of the example clearly illustrates, there can be different depths of interpretation of a given sequence of argumentation in a natural language texts. What the driver example (see Fig. 5 with its use of the scheme for abductive reasoning) also shows is that argument diagramming requires insertion of implicit material if it fits arguments into schemes with explicit generalized conditionals as major premises.

In AI systems the proposition ‘If you hold a knife by its blade then it may cut you’ is cited as an example (Singh et al. [53], 3) of the use of common knowledge to reconstruct everyday common sense reasoning based on contextual information that rational agents know about. Such common knowledge can be expected to vary in different settings of argument use (Hosseini et al. [26]). In a textbook on informal logic, Govier ([22], 120) has used examples such as ‘Human beings have hearts’ and ‘Many millions of civilians have been killed in twentieth-century wars’. The commonsense knowledge problem in AI is how to create a database that contains the general knowledge parties to a natural language conversation need to have in order to communicate and interact with each other as rational agents. In Carneades, the user of the system has to input a knowledge base that contains enough common knowledge in a particular domain of knowledge to enable missing premises and conclusions to be put in by the system.

To insert the implicit propositions into an argument graph of the kind used by Carneades, or any other comparable system, in order to reasonably interpret the line of reasoning in it, we need to follow the interpretation procedure outlined in Fig. 4 (adapted from Walton [61], 92).

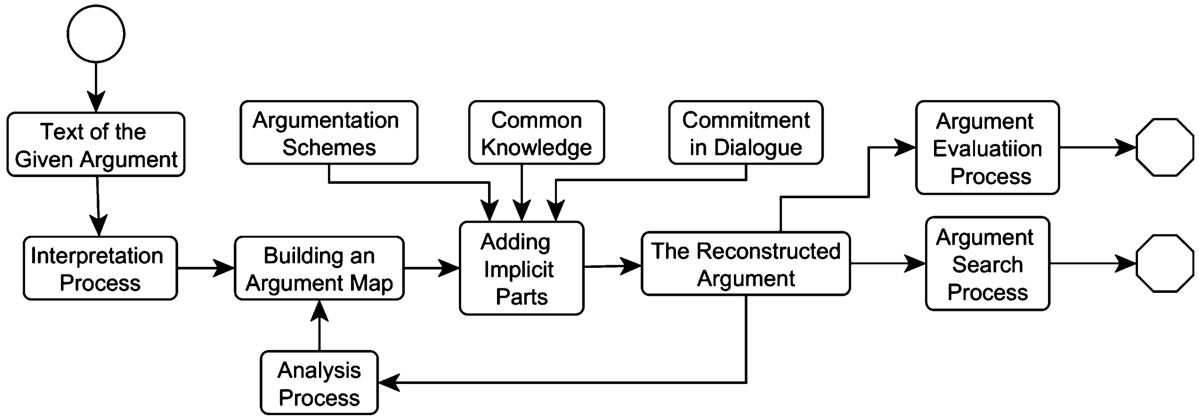

Fig. 4. The argument interpretation procedure.

Finally we can proceed to the construction of an argument diagram shown in Fig. 5 representing the interpretation of the sequence of reasoning in the driver example. In this diagram, the explicit premises or conclusions are represented as propositions contained in the rectangular nodes that have a solid border. The implicit propositions that have been put into the diagram based on common knowledge are contained in rectangular nodes that have a dashed (broken) border. Note that the scheme for abductive reasoning, of the kind called inference to the best explanation by the Josephsons, is indicated by the notation ib, which appears on two of the argument arrows, the one representing a1 and the one representing a3.

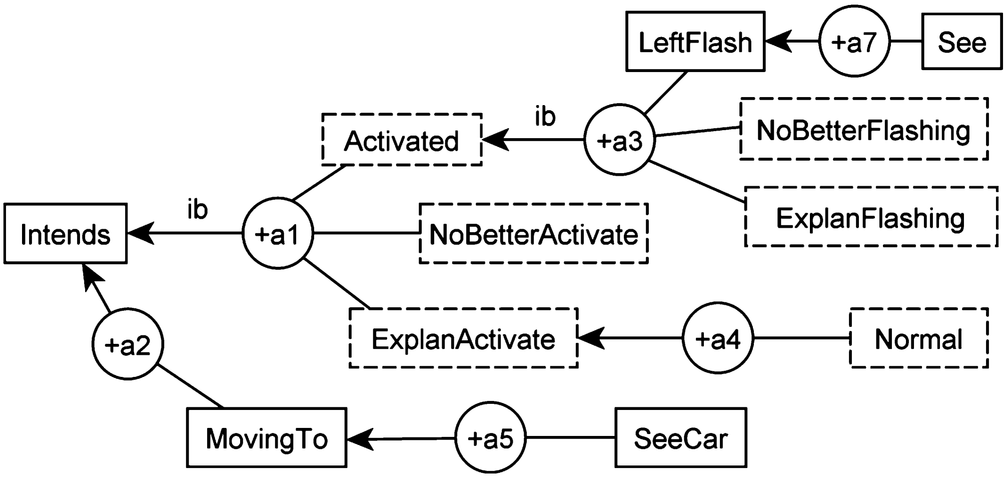

Fig. 5. Abductive reasoning in the driver example.

Note that two of the premises in the abductive argument a3 (labeled as ib for inference to the best explanation) are implicit, and that all three of the premises in the abductive argument in a1 are implicit. This shows the lesson that the insertion of implicit presumptions as unstated premises or conclusions is a necessary component of argument diagramming. In argument a1, the proposition that the driver activated the left turn signal indicator is one premise in an abductive argument that contains two other premises, the proposition that there is no better explanation of his having activated the left turn signal indicator and the proposition that a satisfactory explanation of his having activated the left turn signal indicator is that he intends to signal a left turn. These three premises form an abductive argument leading to the conclusion that the driver intends to signal a left turn. Argument a3 is also an abductive argument, because it is based on what is taken to be a satisfactory explanation and the premise that there is no alternative explanation of the flashing signal light.

This simple example gives the reader an idea of how arguments from evidence to an intention can be modeled in the Carneades Argumentation System. Note how the example shows that applying a formal argumentation system based on argument graphs to realistic examples involving intentions and practical reasoning needs to be based on the capability to fill in implicit premises and conclusions representing how this kind of reasoning works in normal situations of the kind we are familiar with based on common sense reasoning.

The evidence-based reasoning procedure for evaluating the evidence both for and against the claim that an intelligent rational agent had a particular intention in mind requires a weighing of the given evidence in a case. In the approach of this paper, this task requires analyzing the argumentation brought forward by both sides in a case where the claim is disputed. Bex and Walton [7] modeled the argumentation by deploying three legal cases to show how inference to the best explanation is fundamental to understanding how abductive reasoning is an important part of the evidential reasoning in such cases. One of them is the case of Jackson v. Virginia (443 U.S. 307) (Bex and Walton [7], 121–125). The essential details can be given as follows (Walton [62], 144).

> When C was a member of the staff at a county jail where J was incarcerated, she had befriended him. After J's release the Sheriff and another police officer saw C and J in a diner and also observed that they were drinking. As they left the diner, the Sheriff offered to keep J's revolver until he sobered up, but he declined the offer. The next day, C's body was found in a secluded parking lot. She had been shot twice. On the day of the crime, J had been seen by witnesses shooting at targets with his revolver while drinking. The same day he had seen the Sheriff at the diner, J had driven from Virginia to North Carolina. He was convicted of the first-degree murder of C by a Virginia court, but appealed the decision on the ground that the evidence was insufficient to support a finding that he had intended to kill her.

Intent (indicating the so-called "guilty mind") is an element of the crime of murder, and was central to this case. The issue of the trial was whether J intended to kill C or not. The evidence-based reasoning procedure for evaluating the evidence both for and against the claim that an intelligent rational agent had a particular intention in mind requires a weighing of the given evidence in a case. In the approach of this paper, this task requires analyzing the argumentation brought forward by both sides in a case where the claim is disputed.

Key list for the shoot to kill example.

DrinkDiner: C and J were drinking together in a diner.
PoliceSaw: A police officer saw C and J drinking together in a diner.
C'sBody: C's body was found in a parking lot.
ShotTwice: C was shot twice.
ShootingTargets: J was shooting at targets while drinking.
WitSaw: Witnesses saw J was shooting at targets while drinking.
Incapable: J was incapable of premeditation.
Intox: J was intoxicated.

IfIntoxIncap: If J was intoxicated he was incapable of premeditation.
NotSoIntox: J was not so intoxicated as to be incapable of premeditation.
Statement: In his statement, J said he had not been drunk, even though he had been "pretty high".
DroveHome: J later drove home without mishap from Virginia to North Carolina.
**Familiar*: J was familiar with firearms.
Intent: J intended to kill C.

Using the key list, the structure of the argumentation joining them together as a sequence of argumentation can be visualized using the argument diagram shown in Fig. 6. This diagram displays the relevant evidence by mapping the arguments on both sides.

Notice that in this instance, there was only one implicit proposition put in, the proposition that J was familiar with firearms. Other missing premises or conclusions could also be put in. For example, using the argumentation scheme for argument from witness testimony, the proposition that a police officer saw C and J drinking together in a diner could have been supplemented by the additional implicit premise that the police officer was a witness. This instance shows how an argumentation scheme can be used to find an implicit premise in an argument.

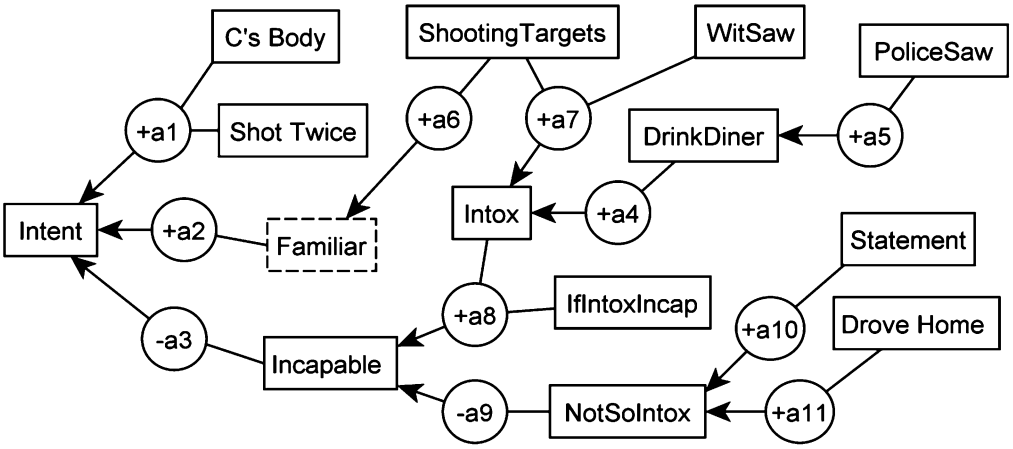

Fig. 6. A legal example of arguments pro and con an intention claim.

The mental element of the crime of murder is stated in the rectangle at the far left containing the proposition that J intended to kill C. The two pro arguments used to support this claim are a1 and a2. There is also a con argument, a3. The premise of the con argument is the proposition that J was incapable of premeditation. Attacking this premise, in turn, is the con argument a9. Supporting the premise that J was not so intoxicated as to be incapable of premeditation are the two pro arguments a10 and a11.

7. Arguments from motive to action and from motive to intention

One of the inferences used in Leonard's car theft example takes the form of an argument from an agent's action to the agent's motive (see Section 2). This inference ([30], 459) can be reconfigured into an argumentation scheme for argument from motive to action shown below.

Conditional Premise: If agent a had a motive to bring about A then a is somewhat more likely to have brought about A.

Motive Premise: a had a motive to bring about A.
Conclusion: a brought about A.

This scheme will be shown to be very useful as a fundamental argumentation scheme for representing certain kinds of inferences from a motive to an action. However, as structured by Leonard ([30], 459) the conditional premise of the inference is stated as a generalization of this kind: If agent a had a motive to bring about A then a is somewhat more likely to have brought about A than another agent who lacked a motive. But there is also another interpretation of the conditional premise that comes into play in some cases: If agent a had a motive to bring about A then a is somewhat more likely to have brought about A than the existing evidence indicated before this particular argument from motive to action was brought into play.

The difference between these two generalizations brings out a point that will be familiar to legal professionals in criminal cases. Sometimes the argument from motive to action is used to draw a conclusion when choosing between a set of suspects by picking out one agent who is more likely to have committed the crime than other suspects who do not appear to have a motive. In other cases, motive is used as evidence in a different way. For example, the prosecution may have collected a lot of evidence suggesting that one agent is more likely to have committed the crime because he or she had a strong motive, such as financial gain. But in this kind of case, all the evidence is being weighed together and it is not specifically a matter of choosing between one agent who had a motive and another agent who either had no motive or had a less compelling motive.

A premise that agent a had a motive to bring about A provides varying degrees of support for a conclusion that a brought about A, depending on other aspects of the situation. In Leonard's hypothetical example, it is a given that D killed V, and the question is not whether D did so but whether D did so intentionally, rather than accidentally as D claimed. In another type of situation, where it is not clear whether anybody did A, the premise that a had a motive to do A provides extremely weak support to the conclusion that a did A, because everybody at any time has all sorts of motives to do things that they would never dream of doing. In a third type of situation, where it is assumed that somebody did A but it is not known who, the existence of a motive for a to do A supports at best a conclusion that a may have done A – a conclusion that warrants further investigation of a in an investigative context and points to the need for further evidence in a probative context.[6] The schemes can be modified to take account of these and other variations using a classification tree, as shown in Section 12.

At this point, to prevent the basic schemes from becoming overly complex, we will proceed with the simple versions of the two schemes presented above, and not (yet) consider the generalization that if an agent had a motive to bring about an action, that agent is somewhat more likely to have brought about the action than another agent who lacked a motive. Also, it might be added parenthetically here that whether you include the generalization expressed in the conditional premise of the scheme or simply leave that premise implicit, and infer directly from the motive premise to the conclusion of a premise stating that an action took place, is partly a matter of what formal system you are using to model argumentation schemes. More light will be thrown on this factor in Section 9.

So far then, we have at least seen how an argument that goes from a motive to an action can be configured with a usable argumentation scheme. But we are still left with the problem of how to argue the other way around, from facts about actions and circumstances to a motive. Teleological reasoning can be used to establish the existence of a motive by drawing an inference from premises concerning facts of a case to a conclusion that a motive exists (Walton and Schafer [68]). A sequence of teleological

[6] I thank an anonymous referee for bringing this important point to my attention.

reasoning leads from a set of evidential circumstances in a case to a hypothesis that postulates the existence of a motive.

Here is an extension of the example to a form of argument that Leonard ([30], 449) identified in the car theft example as argument from motive to intention (excluding the potential ambiguity about the scheme for argument from motive to action discussed just above).

> *Premise*: The defendant had a motive to prevent the victim from revealing the theft to the police.
> *Conclusion*: The defendant intentionally killed the victim to prevent the victim from revealing the theft to the police.

At a more general level, an argumentation scheme representing evidence-based argumentation from motive to intention (mi) can be formulated as follows.

> *Conditional Premise*: If agent a had a motive to bring about A then a is somewhat more likely to have intentionally brought about A.
> *Motive Premise*: a had a motive to bring about A.
> *Conclusion*: a is somewhat more likely to have intentionally brought about A.

This argument can be framed as the argumentation scheme fitting argument +a2 in Fig. 6, where *Intent* in the expanded key list stands for 'The defendant intentionally killed the victim'.

Figure 2 gave a panoramic view of the evidential reasoning in the case, leading to the ultimate conclusion that the defendant killed the victim with intent. Figure 7 adds the schemes ia and mi.

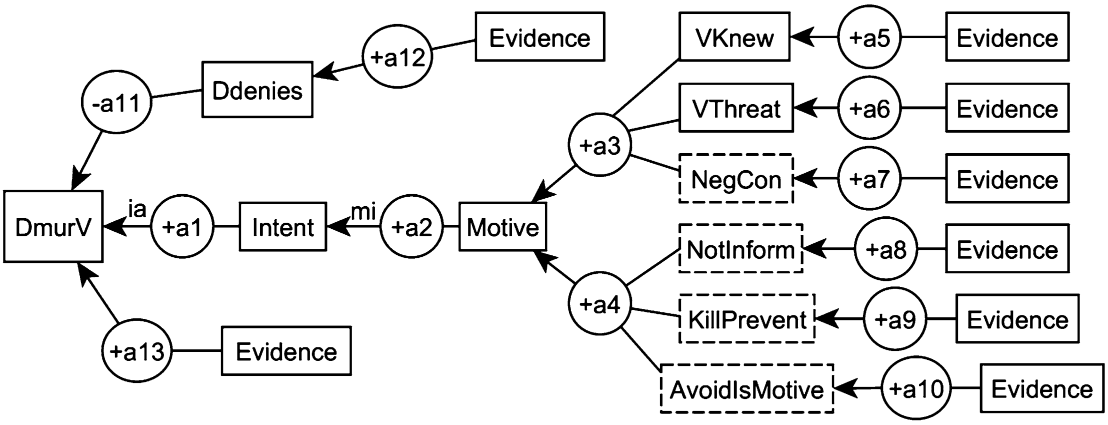

Fig. 7. Argument from evidence to motive to intention.

Now we are in a position to see how Carneades evaluates the arguments in the cases as a whole. To begin with we need to say (following the comments made on Fig. 2), that any such evaluation is subject to how murder is defined in different jurisdictions. The conclusion that the defendant murdered the victim needs to be proved in common law jurisdictions by the beyond reasonable doubt standard.

According to the interpretation of the car theft example visualized as a graph after the fashion of the Carneades Argumentation System, arguments a3 and a4 are individually both linked arguments. However, when they go together to lead to the motive conclusion, the two arguments go together to form a convergent argument supporting the motive conclusion. This means that, for example in the case of a3, all three premises have to be accepted for the argument to carry probative weight to support the motive conclusion. However when you examine the two arguments a3 and a4 from a point of view of how they work together to support the motive conclusion, they work as a convergent argument, meaning that while each of them might give some support to the conclusion in the absence of the other, when taken together they increase the probative weight of the motive conclusion. Each of the arguments a3 and a4, separately, may present some evidence of a weakened and incomplete kind, supporting the motive conclusion, but when the two are taken together the weight of support for the motive conclusion rises significantly. At any rate this is one plausible way of interpreting these arguments.

Another feature to observe is that the argumentation scheme for argument from motive to intent (mi) is shown to apply to argument a2. If this is correct, it shows that a2 is defeasibly valid. The same comment applies to argument a1 which is labeled with the scheme for argument from intention to action (ia). Given that this is one way of interpreting the argumentation in the car theft example, let us now see how Carneades could be used to evaluate this argumentation.

Let's start with the assumption that all six items of evidence shown in the rightmost column in Fig. 8 are accepted. These propositions are shown in a green background.

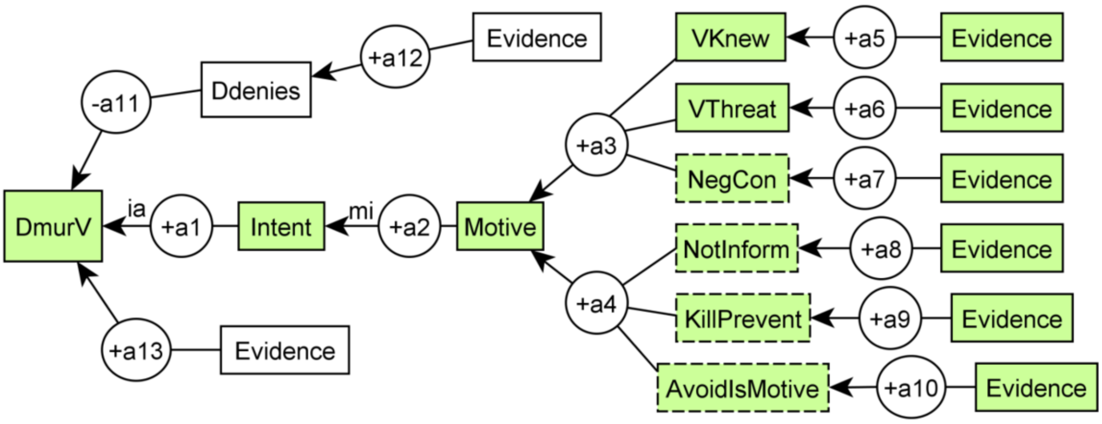

Fig. 8. First step of argument evaluation in the car theft example.

Let's suppose in addition that all of the arguments a5 through a10 are defeasibly valid. Since each of these arguments has only one premise, Carneades will automatically color all six of the conclusions of these arguments with a green background. Let's assume as well that arguments a3 and a4 are defeasibly valid. Since all three premises in both arguments are accepted, the conclusion of this convergent argument (Motive) is also shown with a green background. From this point the argumentation propagates forward through the next two steps, arguments a2 and a1 respectively, to the conclusion that D murdered V.

What has been shown by the evidence considered so far is that the prosecution has a strong line of argumentation supporting the ultimate conclusion to be proved by them, the allegation that D murdered V. This is as far as Leonard's hypothetical example takes us. But as indicated above, if this were a real example there would in all likelihood be a mass of other evidence not only supporting this allegation, but also generally there will be a mass of other evidence in the form of arguments put forward by the defense tending to cast doubt on the prosecution's claim that D murdered V. Some attempt is made to show how this additional evidence is generally taken into account in a real case of this sort by displaying argument a11, a con argument that attacks the proposition that D murdered V. For example the prosecution might bring forward evidence showing that D had an alibi. Moreover in all likelihood there would be additional evidence not only based on motive but on other factors. This is indicated by argument a13.

Let's extend the hypothetical example, simply for the purpose of showing how a mass of evidence in a real trial would generally be expected to work, by assuming that there is evidence supporting D's denial that he murdered V. Once this new part of the evidence is shown in Fig. 9, the ultimate proposition to be proved by the prosecution, the proposition that D murdered V, is now shown in a rectangle with a white background. So even though Leonard's example is a hypothetical one, it can still be used to show how argumentation schemes and argument diagrams can be used in tandem to both provide an interpretation of the argumentation in a given case and even evaluate this argumentation.

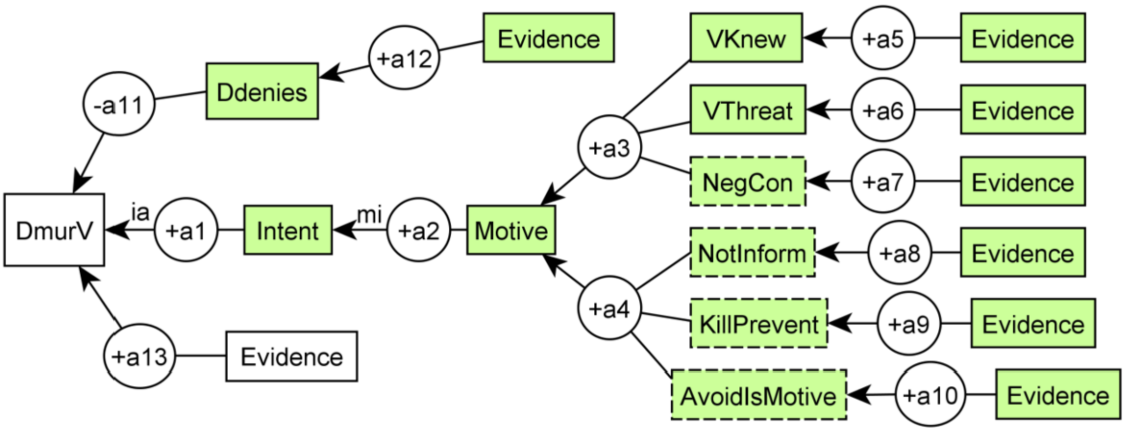

Fig. 9. Second step of argument evaluation in the car theft example.

Another thing that can be mentioned in passing is that in addition to being used to evaluate arguments, Carneades can also be used to invent arguments (Walton and Gordon [64]). A language for representing argumentation schemes presented in (Gordon, Friedrich and Walton [19]) has twenty-five schemes built into the system, including the scheme for practical reasoning and the scheme for inference to the best explanation, and these schemes can be used for the purpose of automatically constructing arguments. So, for example, it can be used to find new arguments that can be added to an argument diagram that will provide an evidential trail from the given evidence at any particular point in an example of an argument, pointing forward to paths that can be found for arguing from that evidence to support or attack the hypothesis that an agent had a particular, goal motive or intention.

8. Multiagent argumentation schemes

The typical argumentation schemes studied so far in the literature on practical reasoning in intelligent autonomous agents tend to involve a single agent reasoning from its goals to its actions. A good example that can be used to illustrate this point is the scheme for argument from negative consequences described in Section 4.

A common example of practical reasoning based on argument from negative consequences would be a case where someone is considering taking a drug that would have beneficial effects, but might also have side effects that are regarded as negative. Now there is an argument against continuing to take the drug. In such cases the agent needs to come to a conclusion on what to do using practical reasoning, by weighing the negative consequences against the positive benefits of the drug in helping with the problem. The general approach using any formal argumentation system of the kind currently available in artificial intelligence is to collect the evidence into a larger argument diagram displaying the evidence supporting the pro and con arguments, and weighing the arguments on the one side against those on the other.

What is interesting to observe in this paper is that in order to model argumentation concerning goals, motives, intentions and other internal mental states that an agent is presumed to have, based on the available evidence, it can be useful to expand the standard argumentation schemes to include new ones involving multiple agents. Below three such schemes especially applicable are introduced. In these three schemes, the variables x and y range over IRAAs and the variable A ranges over actions. The first one is called the multiagent scheme for argument from an agent's perceived negative consequences to the agent's motive.

Major Premise: If x knows that y foresees that there would be negative consequences for y if y does A, then that is evidence that y has a motive for not doing A.
Minor Premise: x knows that y foresees that there would be negative consequences for y if y does A.
Conclusion: There is evidence that y has a motive for not doing A.

The major premise is a common knowledge generalization. The minor premise is a common knowledge presumption where the evidence to support it comes from the circumstances of the case as known by x and y.

The second one is called the multiagent scheme for argument from an agent's perceived positive consequences to the agent's motive.

Major Premise: If x knows that y foresees that there would be positive consequences for y if y does A, that is evidence that y has a motive for doing A.
Minor Premise: x knows that y foresees that there would be positive consequences for y if y does A.
Conclusion: There is evidence that y has a motive for doing A.

The third one is called the multiagent scheme for argument from an agent's perceived elimination of consequences to the agent's motive (em).

Major Premise: If x knows that y foresees that negative consequences for y would be eliminated if y does A, then that is evidence that y has a motive for doing A.
Minor Premise: x knows that y foresees that negative consequences for y would be eliminated if y does A.
Conclusion: There is evidence that y has a motive for doing A.

What is distinctive about these three forms of reasoning is that one agent has to use abductive reasoning to draw a conclusion about the presumed internal mental states of another agent that is also involved in the sequence of multiagent practical reasoning in the example being considered. This conclusion can then be used as a presumption that acts as a premise in the extended sequence of practical reasoning leading to some other conclusion. In this way we can build an argument diagram displaying a practical reasoning structure in which different goals, motives and intentions are connected to each other in the sequence of argumentation as a whole.

It is also possible to add a fourth scheme for the case where an agent's performing an action would eliminate positive consequences for the agent, but since there is no example of that in this paper, we have not added to this type of scheme at this point.

The third multiagent scheme above (em), the one for argument from an agent's perceived elimination of consequences to the agent's motive, applies to the argumentation in the car theft example. In this application, let x be the audience (the jury) looking for a motive to attribute to y, the defendant. Their reasoning takes the following form. The intelligent agent could be a jury, a judge or any kind of audience capable of autonomous reasoning. Let's call it the jury.

> *Major Premise*: If the jury knows that the defendant foresaw that negative consequences for him would be eliminated if he killed the victim, then the jury has evidence that the defendant has a motive for killing the victim.
> *Minor Premise*: The jury knows that the defendant foresaw that negative consequences for himself would be eliminated if he killed the victim.
> *Conclusion*: Therefore, there is evidence that the defendant had a motive for killing the victim.

The problem now is to see how the argumentation scheme em (along with other schemes) can be applied to the car theft example. For this purpose we need to revise the original key list for the car theft example and from it build a new argument diagram. The original eight propositions key list in Section 4 for Leonard's hypothetical example appear above the dotted line, and the additional five propositions required to represent the extended version of the example are shown listed below the dotted line.

Key list for the extended car theft example.

> *VKnew*: V knew that the defendant had stolen a car,
> *VThreat*: V had threatened to inform the police about the theft.
> *NegCon*: If V informed the police there would be negative consequences for D.
> *NotInform*: If V did not inform the police, the negative consequences for D could be avoided.
> *KillPrevent*: Killing V would prevent V from informing the police.
> *AvoidIsMotive*: Avoiding a negative consequence is a motive for doing (or not doing) something.
> *Motive: D* had a motive for killing V.
> *Intent*: D intentionally killed V to prevent V from informing the police of the theft.

. . . .

> *2nd Time*: The defendant knew that this crime would be a second-time offense.
> *Prison*: The defendant knew that the penalty for a second-time offense of auto theft is a prison sentence.
> *Inform*: If the victim informed the police about his crime, the defendant would get a prison sentence.
> *PrisonNeg*: Getting a prison sentence is a negative consequence for the defendant.
> *Goal*: The defendant's goal was to prevent the victim from informing the police.

According to the dialectical theory of goal-based reasoning (Walton [62]), a goal is a proposition about some contemplated future event or action that a group of rational agents can commit themselves to as part of the plan of action that the group is deliberating about. On this approach, a goal is a species commitment that can be set, retracted or altered. There is also a considerable literature in the social sciences where the notion of a goal is defined in a different way, an empirical way (Locke and Latham [31]), but one that is nevertheless more or less compatible with the normative concept of a goal found in argumentation theory.

Practical reasoning is an argumentation scheme representing a sequence of reasoning from a goal to a sequence of actions that should ideally lead to the realization of the goal. But deliberation and practical reasoning do not necessarily model how people think or act in reality. Rather they are normative models about how such a group of people (or even a single agent in some instances) ought to reason or engage in argumentation together or with others so that they can arrive at a decision on what their best option to move forward is, by using pro and con arguments to comparatively evaluates the evidential weight of each option. Deliberation of this kind is very often undertaken with time pressure, absence of complete knowledge of the circumstances, and inconsistency in the database serving as the evidence to weigh the worth of the options (Walton, Toniolo and Norman [69]). Argumentation theory furnishes a useful model of deliberation because it is defined by all three of these characteristics, using defeasible reasoning to arrive at conclusions that may rationally need to be retracted as new knowledge of the circumstances are made available to the agents. This kind of reasoning is modelled using argumentation schemes of a kind that are generally defeasible.

The reader now needs to recall that the scheme for practical reasoning was introduced in Fig. 3. In Fig. 3 it was shown that the defendant's goal was to avoid arrest and the means he allegedly used was to kill the victim, thus preventing the victim from telling the police about the defendant's prior crime. This scheme has now been re-applied in the instance of practical reasoning (pr) displayed at the left of Fig. 10. An argument a2 in the middle, the multiagent scheme em, the one for argument from an agent's perceived elimination of consequences to the agent's motive, was used to derive the conclusion that the defendant's goal was to prevent the victim from informing the police. This goal is put together with the means premise stating that killing the victim would prevent the victim from informing the police.

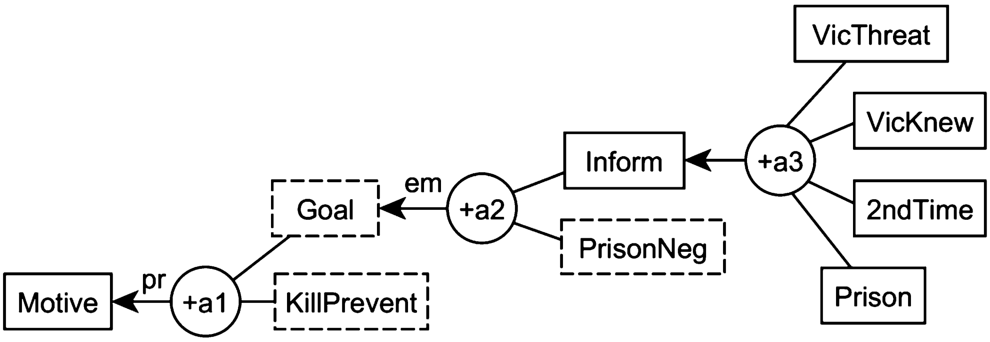

Fig. 10. Application of schemes to the car theft example.

An interesting thing about the argumentation shown in Fig. 10 is that it shows how arguments from a goal to a motive can be reconstructed using the scheme for practical reasoning. It shows how an agent's goal together with its knowledge of the available means can be used to infer the existence of the agent's motive.

Now it is shown how version 2 of Carneades can be used to evaluate the argumentation in an extended version of the car theft example. The evaluation of the extended version of Leonard's hypothetical example works along the same lines as the evaluation of his original example. If all the premises of an argument have been accepted, and the argument fits an argumentation scheme (or is otherwise accepted as defeasibly valid even though it does not fit a known scheme), then the conclusion is shown with a green background, indicating it to needs to be accepted.

Evaluating the inference depends on the standard of proof, such as preponderance of the evidence or beyond reasonable doubt. The user can select the appropriate standard of proof.

Let's say argument a3 meets all these requirements put in by the user. Once this has been done, Carneades will automatically put a green background in the Inform rectangle in Fig. 10.

Looking at Fig. 11, it is now easy to see how the rest of the evaluation procedure will work. Since both premises of argument a2 have now been accepted, assuming that the other requirements for argument a2 are met, Carneades will automatically color the background of the Goal premise green. But now,

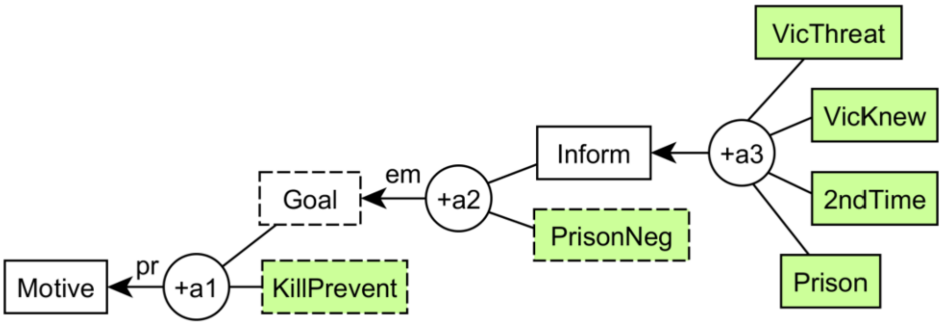

Fig. 11. First step in the evaluation procedure.

since both premises of argument a1 have been accepted, assuming the other requirements for a1 have been met, Carneades will automatically color the Motive box green. By this means it is proved by the evidence (according to the standard of proof put in by the user) that the defendant's motive for killing the victim was to avoid getting a prison sentence.

Version 4.3 has some capabilities that versions 2 and 3 did not have, and these capabilities are relevant to how Carneades can use argumentation schemes to model arguments such as the ones so far analyzed and evaluated in this paper.

9. Examples of formalizing schemes with Carneades 4.3

Version 4.3 has a set of 25 schemes available to the system as constraint handling rules that codify the 25 well known argumentation schemes from the longer list (Gordon, Friedrich and Walton [19]). When applying this system, the user selects a scheme and inserts its premises. When that is done, the argument fitting the scheme is immediately configured in the argument graph on the computer screen with the nodes representing the premises and conclusion of that scheme and the name of the scheme automatically appearing on or near the arrow representing the argument.[7] By this means version 4.3 validates arguments by matching them to schemes, and it can check for syntactic and semantic errors in a knowledge base using a constraint handling rules inference engine. To illustrate how this works, two of the schemes in this paper, the one for argument from motive to action and the one from motive to intention, will suffice to show how it is done.

The first scheme, as formulated in Section 7, is the one for argument from motive to action.

> *Conditional Premise*: If agent *a* had a motive to bring about *A* then *a* is somewhat more likely to have brought about *A* than another agent who lacked a motive.
> *Motive Premise*: *a* had a motive to bring about *A*.
> *Conclusion*: *a* brought about *A*.

The second scheme, as formulated in Section 7, is the one for argument from motive to intention.

> *Premise*: The defendant had a motive to prevent the victim from revealing the theft to the police.
> *Conclusion*: The defendant intentionally killed the victim to prevent the victim from revealing the theft to the police.

[7] Here is the link where you can carry out the procedure: http://carneades.fokus.fraunhofer.de/carneades/.

For reasons of length of the paper, no full legal cases where the circumstantial evidence supports the application of either of these schemes is analyzed using Carneades but the reader can get an idea of how this is done by looking at the analysis of the car theft example in Section 2. How the schemes are actually programmed in Carneades version 4.3 is indicated in the formulation of arguments a1 and a2 below. In Carneades version 2 the issue is defined as the ultimate proposition to be proved or disproved. Carneades 4.3 displays issues as choices to be made, and these issues can change over the argument graph as new evidence comes in.

However, in the example of the code below, note that a list of assumptions for the argument is included. With Carneades, the critical questions are modeled as implicit premises of two kinds. Assumptions are taken to hold in the way of the ordinary premises, meaning that if they are raised as critical questions, the burden of proof is on the proponent of the argument to back them up with the further evidence. Exceptions are taken not to hold, and if they are challenged the burden of proof is on the critical questioner to provide further evidence before the argument is no longer taken to hold. The first parts of the code below define the terms used in the program.

```
issues:
statements:
    intended_to_bring_about(p,a):
        text: p is somewhat more likely to have intended to
bring about a.
        label: in
    undercut(a2):
        text: a2 is undercut.
        label: out
    motive(p,a):
        text: p had a motive to bring about a.
        label: in
    action_conditional_on_motive(p,a):
        text: "If agent, p, had a motive to bring about
action, a, then that agent is somewhat more likely to have
brought about that action."
        label: in
    brought_about(p,a):
        text: p brought about a.
        label: in
    undercut(a1):
        text: a1 is undercut.
        label: out
    intention_conditional_on_motive(p,a):
        text: "If agent, p, had a motive to bring about
action, a, then that agent is somwhat more likely to have
intended to bring about that action."
        label: in
assumptions:
    - motive(p,a)
    - action_conditional_on_motive(p,a)
    - intention_conditional_on_motive(p,a)
arguments:
    a1:
        premises:
            - motive(p,a)
            - action_conditional_on_motive(p,a)
        conclusion: brought_about(p,a)
        weight: 1.00
        scheme: motive_to_action
        parameters: [p, a]
        undercutter: undercut(a1)
    a2:
        premises:
            - motive(p,a)
            - intention_conditional_on_motive(p,a)
        conclusion: intended_to_bring_about(p,a)
        weight: 1.00
        scheme: motive_to_intention
        parameters: [p, a]
        undercutter: undercut(a2)
```

An example argument diagram that is produced by inputting an argument that combines the scheme for argument from motive to action with the scheme for argument from motive to intention is displayed in Fig. 12.

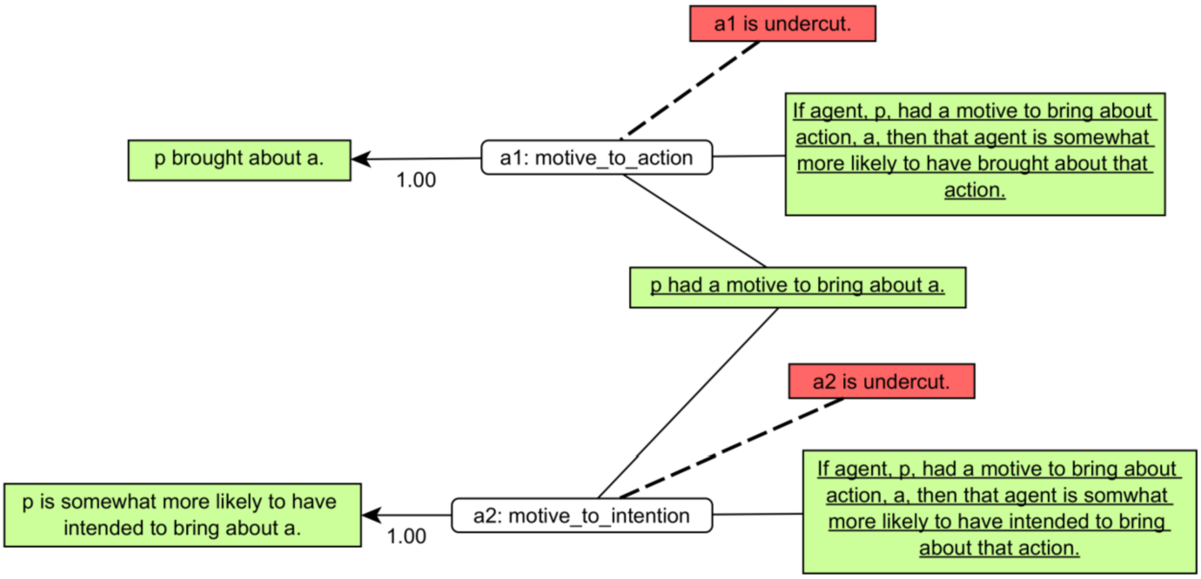
a1 is undercut.
If agent, p, had a motive to bring about action, a, then that agent is somewhat more likely to have brought about that action.
p brought about a.
1.00
a1: motive_to_action
p had a motive to bring about a.
a2 is undercut.
If agent, p, had a motive to bring about action, a, then that agent is somwhat more likely to have intended to bring about that action.
p is somewhat more likely to have intended to bring about a.
1.00
a2: motive_to_intention

Fig. 12. Argument graph of motive and intention schemes in Carneades 4.3.

Version 4.3 treats undercutters in a way that is similar to version 2, but as shown in Fig. 12, a defeasible rectangle stating that a given argument is undercut is generated automatically whenever a scheme is instantiated.

In Fig. 12 the two argument schemes are shown as parts of a combined diagram where the statement that p had a motive to bring about a is a common premise shared by both arguments a1 and a2. In both instances an undercutter is indicated by a broken line joining the undercutter to the argument node. In both instances, the undercutter is shown with a red background indicating that it is 'out', meaning that it does not apply in the given example. Because all the premises of both arguments a1 and a2 are accepted, as indicated by being shown in a rectangular box with a green background, and each of the two arguments fits a scheme, the conclusions of both arguments are also indicated as accepted.

10. Definitions in the dialectical theory of mind

In this section the two legal terms 'motive' and 'intention' initially (tentatively) defined in Section 3 are defined more precisely the way they should be within the dialectical theory. An *intention* as defined in this paper, following Cohen–Levesque [12] and others, is a commitment that an IRAA has committed to using its autonomy, based on its capabilities for free and open choice, as defined by characteristics 1, 2, 3, 14, 16, and 26 of an IRAA (in Section 2). The provisional operative definitions of the terms motive, goal and intention were applied to examples showing how the method provides a pathway from the commitment model to the BDI model linking to postulations about an agent's internal states of mind by an evidence-based sequence of rational argumentation. But it needs to be recalled from Section 2 that IRAA's need to communicate with each other in order to carry out practical goal-based reasoning together. Hence although the analysis of this paper has concentrated mostly on the aspect of argumentation schemes, there is much more work to be done on the dialectical aspect, especially with regard to extending the results here to the wider field of multiagent systems (MAS).

Practical goal-based reasoning is already in the mainstream of both AI and MAS. However, in MAS planning it has been considered to be a social problem to be collectively solved by a group of agents. During this process an intensive discussion needs to takes place with the use of different dialogue types, including persuasion dialogue and inquiry dialogue (Dunin-Keplicz and Strachocka [14]) following the general dialectical theory of Walton and Krabbe [65] that distinguishes between seven main types of dialogue. This seminal research inspired many followers to formalize relatively free dialogues, typically in multi-modal logics, but more recently also in paraconsistent frameworks, permitting both inconsistencies and gaps in knowledge (Dunin-Keplicz and Powala [13]).

In the current models of deliberation dialogue in the AI literature (Kok et al. [28]; Walton et al. [69]) each agent pursues its own goals, but the overall goal of the deliberation dialogue is for the group of agents to decide what course of action they should choose in a set of circumstances where alternative courses of action are available. This implies that these rational agents have two kinds of goals, the collective goal of the dialogue, which is fixed externally for the agents at the outset and the individual goals that each agent has over and above this collective goal. It is assumed that each agent will have its own interests and individual plans. This means that agents may disagree not only about plans, but also about knowledge of the circumstances, which in turn could change the collective goal along the way by changing the choice alternatives.

Walton ([60], 220) offered a dialectical definition of the concept of a motive in law as a 5-tuple $\{M, F, A, S, T\}$ where M is a motive, F is a set of circumstances representing the facts in a case, A

is a set of argumentation schemes, including the scheme for practical reasoning, *S* is a set of so-called story-schemes (Bex [6]) and *D* is a set of formal dialogues of different types. A motive is based around a dialogue exchange between two IRAA's in which the primary agent has carried out some actions and also communicated with the secondary agent through speech acts, and the secondary agent has the task of trying to reconstruct a hypothesis from this evidence that explains what the motive of the primary agent presumably was. The secondary agent uses inference to the best explanation to construct plausible hypotheses about what the motive supposedly was and for this purpose tries to reconstruct the plan of the primary agent by using the circumstantial evidence and asking the secondary agent further questions if possible.

In this paper it is argued that the same dialectical framework can be used to define the legal notion of intention based on the commitment model of argumentation. On this definition, the primary IRAA has carried out an action in a given set of circumstances known to both agents, and the secondary agent is trying to come up with a hypothesis that explains why the primary agent carried out this action. But here the why-question expresses a search for a goal that can explain the first agent's action as rational because it was based on some sort of goal that the first agent was committed to at the time of the action. The argumentation scheme for practical reasoning is especially important here because what the secondary agent is really looking for is some sort of goal that can be attributed to the agent.

However, as noted above, practical goal-directed reasoning is also characteristic of one of the meanings of the term motive that can be distinguished. So care needs to be taken in some cases to differentiate between a goal and a motive. And indeed, in some cases the intention and the motive may refer to the same goal. Nevertheless, as explained by Leonard ([30], 439) the concepts of motive and intention are used in different ways in legal reasoning. The concept of motive is not normally an element of the crime at issue in a given case, whereas the concept of intent is often an element of a charge, claim or defense.

In the commitment-based dialectical theory, both motive and intent are elicited by two kinds of why-questions in the context of a dialogue between two rational agents. When a secondary agent asks a primary agent to explain what his intention was when he carried out an action, what she is requesting is for the primary agent to produce a sequence of goal-directed actions culminating in the primary agent's action that was carried out. In other words, what she wants is a coherent action plan that can be attributed to the agent and that can offer a sequence of coherent autonomous actions contained in a network of goal-based reasoning.

In contrast, when the secondary agent asks the primary agent to explain what his motive was, the question is a different one. But here a distinction needs to be drawn between these two meanings of the term 'motive'. The questioner may be asking the answerer to try to articulate some emotion, by naming or describing it, as he thinks represents what he felt at the time and was the driving force of his action. Or, in the second sense of the word 'motive', the questioner is asking for something that is closer to what might be described as the answerer's intention. In this meaning of the term, the questioner is asking for some goal that the answerer takes to be a rational basis of his carrying out the action as a means to obtain the goal in the given circumstances. In this sense of the term, the distinction between intention and motive is that the why-question asking for a motive is not asking for an extensive plan or complex sequence of actions, but simply asking the answerer what he took his goal to be.

It might be added here that any such account of this type, whether it be in the case of a motive or an intention given by the answerer is subject to further questioning by extending the dialogue and by taking the factual circumstances of the case into account as evidence. For this purpose we have advocated the use of an argument diagramming tool that can take all the relevant evidence of a given case into account by building a large argument graph that takes advantage of using argumentation schemes. So the dialectical method is to map the dialogue derived from the text of the case onto a graph structure, such as an argument diagram, using argumentation schemes, and then evaluate the whole network of argumentation in a given case as representing a mass of evidence, using a graphic tool such as a Wigmore evidence chart (Wigmore [70]).

A rational agent's intention is what the agent is aiming to do, and is defined in this paper as a form of commitment to action using practical reasoning in a context of deliberation. On this account, if an action is intentional, there is a description of the action by linking it to the agents planning and practical reasoning by an argument graph which represents the sequence of actions possibly leading to the carrying out of the person's intention. An agent's goal in performing an action is what the person aimed toward achieving in its plan to accomplish some outcome by performing the sequence of actions in the right order.

Using the criteria of the dialectical theory, the goal and the intention can be distinguished in a given case. For example, a driver may intend to turn left, signal this intention in the usual way, and then carry out this intention by intentionally turning left. But the driver's goal may be to get to a particular destination that he has input into his GPS device, as he follows the sequence of actions displayed on the screen as leading to that goal.

11. Conclusions

This paper has provided a framework for an argumentation-based method to support, attack or find hypotheses about the goals, motives or intentions of intelligent agents using pieces of evidence. The presented method has been built on well-known argumentation schemes, such as those for practical reasoning and inference to the best explanation. An additional series of argumentation schemes and examples of growing complexity have provided a basis for argument technology to find explanations about a supposedly rational agent's goals, motives or intentions that are important for understanding and modeling legal reasoning. The findings are also of interest for contemporary reasoning and decision making systems used in many other fields.

In general outline the method used in the paper applies tools from argument technology to analyze and evaluate arguments based on or concluding to a motive or an intention through a sequence of six steps. The procedure is a general one for argumentation theory that can be used with different computational argumentation systems, or even by using pencil and paper.

1. Collect the evidence from the text of discourse recording the speech acts of the agents, their actions, and the other relevant external circumstances known to the agents at the time.
2. Use this evidence to apply the argumentation schemes to construct an interpretation of the argumentation, based on the evidence, in the format of an argument graph.
3. Fill in implicit premises and conclusion in the connected sequence of argumentation by applying the schemes and filling in gaps acceptable as common knowledge.
4. Weigh the pro arguments for the interpretation against the con arguments by judging how acceptable the premises are that describe the circumstances of the case.
5. By applying the argumentation schemes, derive the conclusions from the premises.
6. Apply standards of proof to determine whether the argument supports the claim of goal, motive, or intention strongly enough to justify acceptance (or refutation) of the conjecture.

There are various formal argumentation systems that can be used to implement this general method. Here we have chosen Carneades as an example of an implemented system that has these capabilities because it already has 106 argumentation schemes available to the system and has the capability for finding arguments using an automated argument assistant.

Carneades has this capability, called argument invention (Walton and Gordon [64]). Hence Carneades can find motives and intentions in legal argumentation by reasoning backwards from a knowledge base comprised of the known facts and applicable legal rules of a case. However, unfortunately, the knowledge base (every fact and rule in it) would have to have been constructed with the schemes in mind because it would have to use the same predicates as the scheme. Currently, this is a similar amount of work to constructing every argument manually. Otherwise, the schemes would not be instantiated. If one added an "ontology" and knowledge engineering, the system could be made easier for the argument-maker.

By abductive reasoning Carneades can automatically construct hypotheses about an agent's presumed motives or intentions based on the evidence in that case. It carries this task out by using argumentation schemes and implicit premises in arguments to build an argument graph connecting the agent's commitments. These commitments are based evidentially on actions the agent has known to have carried out in the past and what the agent has gone on record as stating in the past. So, for example, a judge or lawyer (or anyone using such an argument assistant tool) could use this capability to find an agent's motive or intention, once the evidential data of a case (relevant facts and rules) have been collected into an evidential knowledge base.

12. Six proposals for extending the method

In conclusion, six proposals for extending the dialectical method developed in this paper are suggested. The first proposal is to add the new schemes proposed in the paper. Now that it has been shown how to add two of the schemes to version 4.3 of Carneades in Section 9, what should be done next is to add the further schemes shown to be important for understanding arguments about motives, goals and intentions, such as the three new multiagent schemes. With these new schemes Carneades will have acquired an enhanced capability to find arguments. Unfortunately however, in order for this to be usable enhancement to finding arguments, either the predicates of the new schemes have to drawn from those used by statements in the knowledge base or the predicates in the knowledge base have to be used in constructing the new scheme. What this shows in general is that even though it is possible in principle for AI argumentation systems to find arguments supporting or attacking a claim that an agent has a particular motive or intention using argumentation schemes, there is much more work to do in order to make the system easily usable by the average user.

The second proposal is to take a step in the direction of current research for annotating argumentation schemes (Visser et al. [56]). This step is suggested by the need to revise the scheme for the argument from motive to action by introducing a distinction between two subspecies of this general type of argument. The reason for advocating this extension is the discussion in Section 7 concerning the two variants of the scheme for argument from motive to action arising from the analysis of the car theft example.

The generic form of the scheme for argument from motive to presumed action was structured as follows in Section 7.

Generic scheme for argument from motive to Action[8].

[8] It could be added that strictly speaking, the conclusion should be the proposition that *a* is somewhat more likely to have brought about *A*, but since it is understood that the inference is defeasible, this qualification can be omitted as a convenience where useful.

> *Conditional Premise*: If agent a had a motive to bring about A then a is somewhat more likely to have brought about A.
> *Motive Premise*: a had a motive to bring about A.
> *Conclusion*: a brought about A.

But also in Section 7 it was shown that there are two possible ways of interpreting this scheme depending on what kind of example it is meant to be applied to, insofar as that is known in the given case. The reader will recall that as structured by Leonard ([30], 59) the conditional premise generalization was understood to be interpreted as follows: If agent a had a motive to bring about A then a is somewhat more likely to have brought about A than another agent who lacked a motive. But it was also shown in the discussion in Section 7 that there is also another interpretation of the conditional premise that is applicable in some cases: If agent a had a motive to bring about A then a is somewhat more likely to have brought about A than the existing evidence indicated before this particular argument from motive to action was brought into play.

The proposal arising from these two schemes is to build a small classification tree as part of a general procedure of argument annotation meant to help an argument analyst find the right scheme to apply to an argument found in a given natural language text (Visser et al. [56]). In a decision tree of this kind, the user asks a question to search for the right type of argument appropriate for an analysis of the case.

In this instance, the node that appears at the top of the tree is the generic scheme for argument from motive to action stated in the previous paragraph as representing the general scheme for argument from motive to action. But beneath that node two subspecies can be distinguished, each one of which is represented as a separate branch of the tree. The generic scheme for arguing from a motive to an action needs a completion of the major premise that answers the question: more likely than what? This question can be answered in two ways depending on the evidence of the case and the details of the natural language text of the given case. Here are the two schemes proposed.

Scheme for argument from motive to action for comparing one agent to another.

> *Conditional Premise*: If agent a had a motive to bring about A then a is somewhat more likely to have brought about A than another agent who lacked a motive
> *Motive Premise*: a had a motive to bring about A.
> *Conclusion*: a brought about A.

Scheme for argument from motive to action for evidential increment.

> *Conditional Premise*: If agent a had a motive to bring about A then a is somewhat more likely to have brought about A than the previous evidence indicated.
> *Motive Premise*: a had a motive to bring about A.
> *Conclusion*: a is somewhat more likely to have brought about A than the previous evidence indicated.

The proposal made here is that the generic scheme can be applied to all instances of argument from motive to action, but if the details of the case indicate that one or the other of these more specific schemes is the one that should be applied to the given argument, according to the textual evidence of the case, then that is the interpretation that should be given to the argument.

Recent work with the argument technology group at the University of Dundee (to be published) is concentrated on building a general decision tree applicable to a long list of schemes so that a user can decide which scheme to use in a given natural language text. This procedure of building a decision tree for argument annotation rests on the capability of classifying so-called clusters of schemes, where several more specific schemes are classified under a more general or generic type of scheme. This type of work is becoming increasingly important, so it is useful to consider decision tree structures where a general scheme can be instantiated in specific cases in various ways using more specific schemes. This capability shows promise of being helpful for the current research on argument mining using machine learning and argumentation schemes.

The third proposal flows from Fig. 10, which is the first instance in the paper of an argument or argument scheme with a conclusion attributing a goal to an agent. If we want to extend the paper to be about reasoning from evidence to goals as well as about reasoning from evidence to motives and intentions, this kind of reasoning could be further investigated. For example, one could use an argument from a motive for carrying out an action, and the carrying out of the action, to the conclusion that the agent's goal in carrying out the action was to accomplish what the agent had a motive to do. For example, one might consider the following hypothetical argument that could be extended from Leonard's car theft example.

> Premise: The defendant killed the victim.
> Premise: The defendant had a motive to prevent the victim from informing the police that the defendant stole a car.
> Conclusion: There is good evidence that the defendant's goal in killing the victim was to prevent the victim from informing the police that the defendant had stolen a car.

Whether such an argument has ever actually been used in law is not the issue here. It is a new form of argument that could possibly be investigated in addition to the other ones considered in this paper. It is an argument from an action and a motive to a goal. It illustrates at least the possibility of combining two premises, one about an action, and the other about a motive, to argue to a conclusion about a goal. At any rate, such complex arguments of this sort might possibly lead researchers to search for new examples that might be of some interest.

The fourth research proposal is to extend this to the work of Bex et al. [5] who used a value-based version of the argumentation scheme for practical reasoning and an action-based alternating transition system (AATS) to build a formal argumentation framework that can be used to reason abductively to an agent's motives (Bex et al. [5]). They use the general term 'motivation' in such a way that it can apply not only to motives but also to other comparable everyday minimalistic notions that might be used by a juror or detective in a legal case. Their analysis uses different terminology and is based on different technology, that is, a different formal argumentation system than the one used in this paper. In the present paper an instrumental scheme for practical reasoning is used that requires only a goal and means as premises, and does not take the promotion of values into account during practical reasoning. But the method used in this paper can be extended to take cases of value-based practical reasoning into account.

A fifth useful direction for future research proposed here is to extend the work in this paper by using a related technology: the formal STIT model of the notion of an intelligent agent bringing about some outcome by carrying out an action (Belnap and Perloff [4]; Elgesem [16]). This notion of bringing about is meant to capture the idea that for a state of affairs to be brought about by an action it is not enough for the action to be a sufficient condition for the state of affairs, but it is also required that the action be a necessary condition of the state of affairs (Lorini and Sartor [32], 777). Incorporating a formal notion of action of this kind that could be attributable to an IRAA would enable us to work towards a deeper analysis of the logical structure of multiagent practical reasoning.

The sixth proposal is to test the argument invention tool of Carneades by applying it to some real (more complex) legal cases where there is a need to find arguments concluding that an agent has or does not have a particular motive or intention, using the evidential database[9] and the argumentation schemes. As noted in Section 3, a rational agent's goal can often be inferred (if there is sufficient evidence) by using a combination of practical reasoning and abductive reasoning (as illustrated by the examples in this paper). Using the procedure of finding a goal described in Section 3, another rational agent can reason retrospectively from the first agent's actions and statements that this first agent has gone on record as being committed to, and thereby derive a conclusion about the first agent's presumed goal. To use this procedure, it is assumed that the second agent has collected a knowledge base of propositions that are presumed to be commitments of the first agent, based on the evidence contained in this knowledge base. Using this set of propositions as the admissible evidence, the second agent reasons backwards defeasibly, based on argumentation schemes and an argument diagram displaying the network of connected reasoning in the case, to come up with an evidence-based conclusion about what are taken to be the first agent's goals, motives and intentions.

References

[1] K. Atkinson and T.J.M. Bench-Capon, Practical reasoning as presumptive argumentation using action based alternating transition systems, *Artificial Intelligence* **171** (2007), 855–874. doi:10.1016/j.artint.2007.04.009.

[2] K. Atkinson, T.J.M. Bench-Capon and P. McBurney, Computational representation of practical argument, *Synthese* **152**(2) (2006), 157–206.

[3] R. Audi, *Practical Reasoning*, Routledge, London, 1989.

[4] N. Belnap and M. Perloff, Seeing to it that: A canonical form for agentives, *Theoria* **54** (1988), 175–199. doi:10.1111/j.1755-2567.1988.tb00717.x.

[5] F. Bex, T. Bench-Capon and K. Atkinson, Did he jump or was he pushed?, *Artificial Intelligence and Law* **17** (2009), 79–99. doi:10.1007/s10506-009-9074-z.

[6] F.J. Bex, *Arguments, Stories and Criminal Evidence: A Formal Hybrid Theory*, Springer, Dordrecht, 2011.

[7] F.J. Bex and D. Walton, Burdens and standards of proof for inference to the best explanation: Three case studies, *Law, Probability and Risk* **11**(2–3) (2012), 113–133. doi:10.1093/lpr/mgs003.

[8] F.J. Bex and D. Walton, Taking the dialectical stance in reasoning with evidence and proof, *The International Journal of Evidence and Proof* **23**(1–2) (2019), 90–99. doi:10.1177/1365712718813795.

[9] M. Bratman, *Intentions, Plans, and Practical Reason*, Harvard University Press, Cambridge, MA, 1987.

[10] M. Bratman, *Shared Agency*, Oxford University Press, Oxford, 2014.

[11] M. Bratman, D. Israel and M. Pollack, Plans and resource-bounded practical reasoning, *Computational Intelligence* **4**(4) (1988), 349–355. doi:10.1111/j.1467-8640.1988.tb00284.x.

[12] P.R. Cohen and H.J. Levesque, Intention is choice with commitment, *Artificial Intelligence* **42**(2–3) (1990), 213–261. doi:10.1016/0004-3702(90)90055-5.

[13] B. Dunin-Kęplicz and A. Powała, Multi-party persuasion: A paraconsistent approach, *Fundamenta Informaticae* **158**(1–3) (2018), 1–39. doi:10.3233/FI-2018-1640.

[14] B. Dunin-Kęplicz and A. Strachocka, Tractable inquiry in information-rich environments, in: *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence, IJCAI 2015*, AAAI Press, 2015, pp. 53–60.

[15] V. Dunin-Keplicz and R. Verbrugge, *Teamwork in Multi-Agent Systems: A Formal Approach. Chichester:Wiley*, 2010.

[16] D. Elgesem, The modal logic of agency, *Nordic Journal of Philosophical Logic* **2** (1997), 1–46.

[17] P. Engel, *Believing and Accepting*, Kluwer, Dordrecht, 2000.

[18] T.F. Gordon, *The Carneades Argumentation Support System, Dialectics, Dialogue and Argumentation*, C. Reed and C.W. Tindale, eds, College Publications, London, 2010.

[19] T.F. Gordon, H. Friedrich and D. Walton, Representing argumentation schemes with constraint handling rules, *Argument & Computation* **9**(2) (2018), 91–119. doi:10.3233/AAC-180039.

[9] However, it must be added that to do it this way, the evidential database would have to be hand-tailored to the schemes or vice-versa, which may or may not turn out to be practical.

[20] T.F. Gordon, H. Prakken and D. Walton, The Carneades model of argument and burden of proof, *Artificial Intelligence* **171** (2007), 875–896. doi:10.1016/j.artint.2007.04.010.

[21] T.F. Gordon and D. Walton, Formalizing balancing arguments, in: *Proceedings of the 2016 Conference on Computational Models of Argument (COMMA 2016)*, IOS Press, 2016, pp. 327–338.

[22] T. Govier, *A Practical Study of Argument*, 3rd edn, Wadsworth, Belmont, 1992.

[23] H.P. Grice, *Logic and Conversation, Syntax and Semantics*, P. Cole and J.L. Morgan, eds, Vol. 3, Academic Press, New York, 1975, pp. 43–58.

[24] C.L. Hamblin, *Fallacies*, Methuen, London 1970.

[25] C.L. Hamblin, *Introduction to Linguistics and the Parts of the Mind*, Cambridge Scholars Publishing, Newcastle upon Tyne, 2017.

[26] S.A. Hosseini, S. Modgil and O.T. Rodrigues, Enthymeme construction in dialogues using shared knowledge, in: *Computational Models of Argument*, S. Parsons, N. Oren, C. Reed and F. Cerutti, eds, IOS Press, Amsterdam, 2014, pp. 325–332.

[27] J.R. Josephson and S.G. Josephson, *Abductive Inference: Computation, Philosophy, Technology*, Cambridge University Press, Cambridge, 1994.

[28] E.M. Kok, J.-J. Meyer, H. Prakken and G.A.W. Vreeswijk, A formal argumentation framework for deliberation dialogues, in: *Argumentation in Multi-Agent Systems*, P. McBurney, I. Rahwan and S. Parsons, eds, Lecture Notes in Computer Science, Vol. 6614, Springer, Berlin Heidelberg, 2011, pp. 31–48. doi:10.1007/978-3-642-21940-5_3.

[29] F. Kramer, Intent and Motive are Different: Except When They Aren't, 2015, https://www.americanbar.org/content/dam/aba/events/criminal_justice/2015/Knowledge_Willfulness_Intent_Motive.authcheckdam.pdf.

[30] D.P. Leonard, Character and motive in evidence law, *Loyola of Los Angeles Law Review* **34** (2001), 439–536.

[31] E.A. Locke and G.P. Latham, Building a practically useful theory of goal setting and task motivation: A 35-year odyssey, *American Psychologist* **57**(9) (2002), 705–717. doi:10.1037/0003-066X.57.9.705.

[32] E. Lorini and G. Sartor, A STIT logic for reasoning about social influence, *Studia Logica* **104** (2016), 773–812. doi:10.1007/s11225-015-9636-x.

[33] F. Macagno and D. Walton, *Interpreting Straw Man Argumentation*, Springer, Cham, 2017.

[34] F. Macagno and D. Walton, *Implicatures as Forms of Argument, Perspectives on Pragmatics and Philosophy*, A. Capone et al., eds, Springer, Berlin, 2013, pp. 203–224. doi:10.1007/978-3-319-01011-3_9.

[35] F. Macagno, D. Walton and C. Reed, Argumentation schemes, history, classifications and computational applications, *IFColog Journal of Logics and Their Applications* **4**(8) (2017), 2493–2556.

[36] S. McRoy and G. Hirst, The repair of speech act misunderstandings by abductive inference, *Computational Linguistics* **21**(4) (1995), 475–478.

[37] B. Meadows, P. Langley and M. Emery, Seeing Beyond Shadows: Incremental Abductive Reasoning for Plan Understanding, 2013, Association for the Advancement of Artificial Intelligence, https://www.aaai.org/ocs/index.php/WS/AAAIW13/paper/viewFile/7028/6591.

[38] T. Miller, Explanation in artificial intelligence: Insights from the social sciences, *Artificial Intelligence* **267** (2019), 1–38. doi:10.1016/j.artint.2018.07.007.

[39] M. Minsky, A Framework for Representing Knowledge, Reprinted in *The Psychology of Computer Vision*, P. Winston, ed., McGraw Hill, New York, 1975, available at: http://web.media.mit.edu/~minsky/papers/Frames/frames.html.

[40] T.J. Norman and D. Long, Goal creation in motivated agents, in: *Intelligent Agents: Theories, Architectures, and Languages*, M. Wooldridge and N.R. Jennings, eds, LNAI, Vol. 890, Springer, Heidelberg, 1995, pp. 277–290. doi:10.1007/3-540-58855-8_18.

[41] F. Paglieri and C. Castelfranchi, The Toulmin test: Framing argumentation within belief revision theories, in: *Arguing on the Toulmin Model*, D. Hitchcock and B. Verheij, eds, Springer, Dordrecht, 2006, pp. 359–377. doi:10.1007/978-1-4020-4938-5_24.

[42] A.R. Panisson, S. Sarkadi, P. McBurney, S. Parsons and R.H. Bordini, On the formal semantics of theory of mind in agent communication, in: *6th International Conference Agreement Technologies*, AT 2018, Bergen, Norway, December 6–7, 2018, 2018, Revised Selected Papers, 18–32, https://nms.kcl.ac.uk/simon.parsons/publications/conferences/at18a.pdf.

[43] D.L. Poole and A.K. Macworth, *Artificial Intelligence: Foundations of Computational Agents*, Cambridge University Press, Cambridge, 2011.

[44] I. Rahwan and L. Amgoud, An argumentation-based approach for practical reasoning, in: *Proceedings of the 5th International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, ACM Press, Hakodate, Japan, New York, 2006, pp. 347–354. doi:10.1145/1160633.1160696.

[45] C. Reed and T.J. Norman, *Argumentation Machines: New Frontiers in Argument and Computation*, Dordrecht, Kluwer, 2003.

[46] S.J. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*. Prentice Hall, Upper Saddle River, NJ, 1995.

[47] F. Sadri, Intention recognition with event calculus graphs, in: *Proceedings – 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology – Workshops, WI-IAT*, 2010, pp. 386–391, https://www.doc.ic.ac.uk/~fs/Papers/Intention%20Recognition/HAI%20Camera%20ready%2023%20June%202010.pdf.

[48] S. Sarkadi, A.R. Panisson, R.H. Bordini, P. McBurney and S. Parsons, Towards an approach for modelling uncertain theory of mind in multi-agent systems, in: *6th International Conference Agreement Technologies*, AT 2018, Bergen, Norway, December 6–7, 2018, 2018, pp. 3–17, Revised Selected Papers, https://nms.kcl.ac.uk/simon.parsons/publications/conferences/at18b.pdf.

[49] R.C. Schank and R.P. Abelson, *Scripts, Plans, Goals and Understanding*, Erlbaum, Hillsdale, New Jersey, 1977.

[50] M. Scriven, The limits of explication, *Argumentation* **16** (2002), 47–57. doi:10.1023/A:1014917625208.

[51] J.R. Searle, *Rationality in Action*, The MIT Press, Cambridge, MA, 2001.

[52] A. Sidgwick, *Fallacies*, D. Appleton and Company, New York, 1884.

[53] P. Singh, T. Lin, E.T. Mueller, G. Lim, T. Perkins and W.L. Zhu, Open Mind Common Sense: Knowledge Acquisition from the General Public, in: *Proceedings of On the Move to Meaningful Internet Systems, 2002 – DOA/CoopIS/ODBASE 2002 Confederated International Conferences DOA, CoopIS and ODBASE*, Irvine, California, USA, October 30–November 1, 2002, pp. 1223–1237.

[54] S.C. Stover, Best Practices in Proving Specific Intent and Malice, American Bar Association Report on Litigation Materials, 2014, https://www.americanbar.org/content/dam/aba/administrative/litigation/materials/2014_sac/2014_sac/best_practices.authcheckdam.pdf.

[55] R. Tuomela, *Social Ontology: Collective Intentionality and Group Agents*, Oxford University Press, Oxford, 2013.

[56] J. Visser, J. Lawrence, J. Wagemans and C. Reed, Revisiting computational models of argument schemes: Classification, annotation, comparison, in: *Proceedings of the Seventh Conference on Computational Models of Argument (COMMA 2018)*, IOS Press, Warsaw, 2018, pp. 313–324.

[57] M. Vitiello, Defining the reasonable person in the. Criminal law: Fighting the lernaean hydra, *Lewis and Clark Law Review* **14**(4) (2010), 1435–1454.

[58] G.H. Von Wright, Practical inference, *The Philosophical Review* **72** (1963), 159–179. doi:10.2307/2183102.

[59] D. Walton, *Abductive Reasoning*. University of Alabama Press, Tuscaloosa. 2005.

[60] D. Walton, Teleological argumentation to and from motives, *Law, Probability and Risk* **10**(3) (2011), 203–223. doi:10.1093/lpr/mgr012.

[61] D. Walton, *Methods of Argumentation*, Cambridge University Press, Cambridge, 2013.

[62] D. Walton, *Goal-Based Reasoning for Argumentation*, Cambridge University Press, Cambridge, 2015.

[63] D. Walton, Plausible argumentation in eikotic arguments: The ancient weak versus strong man example, *Argumentation* **33**(1) (2019), 45–74. doi:10.1007/s10503-018-9460-3.

[64] D. Walton and T.F. Gordon, Argument invention with the Carneades argumentation system, SCRIPTed: A, *Journal of Law, Technology & Society* **14**(2) (2016), 168–207.

[65] D. Walton and E. Krabbe, *Commitment in Dialogue*, State University of New York Press, Albany, 1995.

[66] D. Walton, C. Reed and F. Macagno, *Argumentation Schemes*, Cambridge University Press, Cambridge, 2008.

[67] D. Walton and G. Sartor, Teleological justification of argumentation schemes, *Argumentation* **27**(2) (2013), 111–142. doi:10.1007/s10503-012-9262-y.

[68] D. Walton and B. Schafer, Arthur, George and the mystery of the missing motive: Towards a theory of evidentiary reasoning about motives, *International Commentary on Evidence* **4**(2) (2006), 1–47.

[69] D. Walton, A. Toniolo and T.J. Norman, Towards a richer model of deliberation dialogue: Closure problem and change of circumstances, *Argument and Computation* **7**(2–3) (2016), 155–173. doi:10.3233/AAC-160009.

[70] J.H. Wigmore, *The Principles of Judicial Proof*, Little, Brown and Company, Boston, 1931.

[71] J.H. Wigmore, *Evidence in Trials at Common Law*, Little, Brown & Co., Boston, 1940.

[72] M. Wooldridge, *Reasoning About Rational Agents*, The MIT Press, Cambridge, MA, 2000.

[73] M. Wooldridge, *An Introduction to MultiAgent Systems*, 1st edn, Wiley, Chichester, 2002.

[74] M. Wooldridge, *An Introduction to MultiAgent Systems*, 2nd edn, Wiley, Chichester, 2009.