1

Value-based Argumentation

Katie Atkinson and Trevor Bench-Capon

ABSTRACT. Value-based argumentation is concerned with recognising, accounting for, and reasoning with, the social purposes promoted by agents' beliefs and actions. Value-based argumentation frameworks extend Dung's abstract argumentation frameworks by ascribing an additional property to arguments, representing the values they promote, and recognising audiences. Values are ordered according to the preferences of an audience (different audiences will have different preferences) and an attack is successful only if the value of the attacked argument is not preferred to its attacker by its audience. Arguments can be related to values through the use of an argumentation scheme, thus enabling us to structure value-based argumentation. We describe the motivation of value-based argumentation, its formal description and properties, the argumentation scheme and its associated critical questions and some of the applications to which value-based argumentation has been put.

1 Philosophical motivations for value-based argumentation

The formal models of value-based argumentation that are presented in this chapter are intended to capture various philosophical concepts that are reflected in everyday human reasoning. In this section we explain the key philosophical accounts that motivated the development of the computational models of value-based argument.

1.1 Values and audiences

The inspiration for value-based argumentation originally came from the *New Rhetoric* of Perelman and Olbrechts-Tyteca [1969]. The key insight of the *New Rhetoric* was that the acceptability of an argument depended not only on the argument itself, or on available counterarguments, but on the audience to which it was addressed. For an argument to be accepted, *its audience has to accept it*. In subsequent work on this topic Perelman says:

> If men [sic] oppose each other concerning a decision to be taken, it is not because they commit some error of logic or calculation. They discuss apropos the applicable rule, the ends to be considered, the meaning to be given to values, the interpretation and characterisation of facts. [Perelman, 1980], p150.

Perelman's academic roots were in jurisprudence and he drew on legal disputes to support his argument:

> " Each [party] refers in its argumentation to different values [...] the judge will allow himself to be guided in his reasoning by the spirit of the system: i.e. by the values which the legislative authority seeks to protect and advance." [Perelman, 1980], p152.

Consideration of this had also been noted in AI and Law. In their highly influential paper, Berman and Hafner [1993] discussed what should happen in factor-based reasoning with cases [Bench-Capon, 2017] when there were no precedents to allow the case to be decided. They argued that in such cases the decision should be made according to which social purposes would be promoted by deciding for the plaintiff and which would be promoted by deciding for the defendant, and the decision made according to which would better serve the prevalent social values. Note that this means that different arguments can be accepted in different jurisdictions (attitudes to the death penalty in Georgia and Minnesota were very different in the 1970s), and at different times ("*stare decisis* would bow to changing values"[1]).

Thus there seems something missing from a purely logical view: sometimes the logic will fail to compel, and we will need to make a choice on other grounds. Since the situation occurs in important arenas like law, we do not want the choice to be arbitrary: we want to provide rational grounds for such choice. As Perelman puts it:

> Logic underwent a brilliant development during the last century when, abandoning the old formulas, it set out to analyze the methods of proof used effectively by mathematicians. ... One result of this development is to limit its domain, since everything ignored by mathematicians is foreign to it. Logicians owe it to themselves to complete the theory of demonstration obtained in this way by a theory of argumentation. [Perelman and Olbrechts-Tyteca, 1969], p10.

The situation is reflected in Dung's abstract argumentation. Sometimes, the acceptability of an argument will not be unequivocally determined by the framework. Given a dilemma (cycles with even length in standard argumentation frameworks [Bench-Capon, 2014]) the restrictive grounded semantics will allow neither horn to be embraced, whereas the more permissive preferred semantics will allow either proposition to be believed, but offer no reason to opt for one rather than the other. Value-based argumentation attempts to offer reasons for this choice as part of a "theory of argumentation".

1.2 Direction of fit

The other key influence on value-based argumentation was the work of John Searle on practical reasoning and his notion of *direction of fit* [Searle, 2003]. Searle wrote

[1] Justice Marshall in Furman v Georgia 408 U.S. 238 (1972).

> Assume universally valid and accepted standards of rationality, assume perfectly rational agents operating with perfect information, and you will find that rational disagreement will still occur; because, for example, the rational agents are likely to have different and inconsistent values and interests, each of which may be rationally acceptable. [Searle, 2003], p. xv.

Searle's idea was that such rational disagreement was possible because of direction of fit. There is only a single actual world, and a single history of that world, and so our beliefs about the present and the past have to match that actual world. Because there is only one actual world, there is a right answer to questions of fact, and while there may be disagreement, this is something that should be capable of resolution, given complete information. Values, interests and aspirations can play no part in such *theoretical* reasoning: that would be to indulge in wishful thinking.

The future is, however, a different matter. There are many possible futures, and we can, through our actions, play a part in determining which will come to pass. In *practical* reasoning, reasoning about what we should do, we attempt to fit the world to our desires, so that our actions will bring out the future that we prefer. But here different values, interests, aspirations and even tastes, may be a legitimate source of rational disagreement. Some may find it strange if someone prefers vanilla ice cream to chocolate, but it is not irrational. Of course, these aspirations can affect deeper matters: in politics a desire for tax rises may exhibit a preference for equality over economic growth. Such a preference is not a matter of rationality, but of the values that one wishes to be expressed in a society.

Thus in practical reasoning, rational disagreement is to be expected [Bench-Capon, 2002c]. The notion of direction of fit, however, applies not only to actions, but to the law. Disagreement is at the heart of law, and even at the highest level judges differ as to the proper outcome of a case. Five-to-four decisions occur in almost a fifth of cases heard by the US Supreme Court[2]. Not only is disagreement common, it is expected: that is why appeal courts typically comprise an odd number of judges, and why the more important the court the more the judges, so that the US Supreme Court has nine[3]. Nor can judicial agreement be considered irrational: after all, the minority will produce an opinion stating their reasons for their views. To a certain extent the judges are trying to fit their view of the current cases to the existing law: the doctrine of *stare decisis* means that their decision should be consistent with past decisions. For a logical analysis of precedential constraint, see [Horty and Bench-Capon, 2012]. However, it is often the case that the precedents do not

[2]Between 2000 and 2018, according to the US Supreme Court as reported in *The Washington Post* https://www.washingtonpost.com/news/posteverything/wp/2018/06/28/those-5-4-decisions-on-the-supreme-court-9-0-is-far-more-common/. Last accessed February 2020.

[3]Nine is the traditional number. As we write, in the run up to the 2020 Presidential election, there is speculation that this may be increased.

fully constrain the decision: it may be that all of them can be distinguished according to some features of the current case. For such cases the judges are free to decide for either party. Here they try to fit the law as it will be after their decision (for the current case will serve as a precedent for future cases) to the way they desire the law should be. That is, they consider which decision will promote the purposes of the law better, as described in [Berman and Hafner, 1993]. Therefore, as in practical reasoning, the values and aspirations of the judges will determine their decision [Atkinson and Bench-Capon, 2005]. The justification is that the majority opinion of a properly appointed court should reflect the prevailing values of its society.

1.3 Value-based argumentation

To reflect the situation where the dispute is about how best to fit the world to our desires it is clear that a basic assumption of Dung's argumentation frameworks, that attacks always succeed in defeating the argument they attack, must be relaxed. As an example, while it is true that Sarah will not be able to go on holiday if she buys a new car, this attack can simply be ignored if Sarah prefers the holiday: she can continue to make do with her current car. For a different person, however, perhaps a petrol-head like Jeremy, the attack will be decisive and the holiday plans abandoned.

Thus to reflect debates where values, aspirations and tastes matter, not only in everyday practical reasoning, but in important areas such law, politics and ethics as well, a method of augmenting Dung's framework with a notion of values was needed. *Values* was used to cover these subjective preferences. It is a term widely understood in this sense in popular media, and the notion of a *value premise* is a key part of the Lincoln-Davis debate format used throughout the USA as the basis for competitive debating in a number of leagues[4]. Thus the general notion of values is felt to be widely understood. For example, the French Republic was based on the three values of liberty, equality and fraternity. In value-based reasoning there have been many different sets of values used for different problems. Generally it is held that the identification of the relevant set of values is part of the formulation of the problem to be discussed [Atkinson and Bench-Capon, 2007c]. Some attempts have been made to provide a basis for the identification of values: e.g. van der Weide *et al.* [2009] used Schwartz Value Theory [Schwartz, 1992] and Bench-Capon [2020] used Maslow's hierarchy of needs [Maslow, 1943]. Often, however, it seems that a very general account is not best suited a particular problem, and the use of problem specific value sets remains common.

[4]Including the National Speech and Debate Association, or NSDA (formerly known as the National Forensics League, or NFL) competitions, and related debate leagues such as the National Christian Forensics and Communication Association, the National Catholic Forensic League, the National Educational Debate Association, the Texas University Interscholastic League, Texas Forensic Association, Stoa USA and their affiliated regional organizations.

2 Values in abstract argumentation frameworks

Value-based argumentation first appeared in the context as an extension to Dung's abstract argumentation framework, first in [Bench-Capon, 2002b] and later in the journal version [Bench-Capon, 2003a]. The basic idea was to extend Dung's notion of an augmentation framework as pair of a set of arguments and a set of the attacks between them, $\langle Ar, att \rangle$ by adding a set of values, V, a function mapping the members of Ar onto V, val, and a set of audiences, expressed as orderings on V, P. Note that P might contain all the factorially many possible orderings on V, or only a selection of them. This might be to represent a particular set of agents with specific preferences, or some constraint on the ordering itself. For example, in order to represent facts, theoretical arguments are typically related to the value *truth*. Then to avoid wishful thinking, truth must be the most preferred value for every audience.

2.1 Extending Dung's argumentation frameworks with values

Accordingly, a value-based argumentation framework (VAF) is defined as an extension of a Dung-style argumentation framework (AF).

Definition 2.1 (Value-Based Argumentation Framework (VAF)) *A value-based argumentation framework is a 5-tuple $VAF = \langle Ar, att, V, val, P \rangle$ where Ar is a finite set of arguments, att is an irreflexive binary relation on Ar, V is a nonempty set of values, val is a function which maps from elements of Ar to elements of V and P is the set of possible audiences (represented as orderings on V). We say that an argument $a \in Ar$ relates to a value $v \in V$ if accepting a promotes or defends v: The value in question is given by $val(ar)$. For every $ar \in Ar$, $val(ar) \in V$.*

Note that if there is a single value, (perhaps *truth*), a VAF is equivalent to a standard Dung AF. If every argument maps to its own distinct value, we have a similar situation to the Preference Based Frameworks of Amgoud and Cayrol [Amgoud and Cayrol, 1998] and [Amgoud and Cayrol, 2002], except that Preference Based Argumentation uses only a single ordering so that P has only one member, and there is only a single audience.

In order to evaluate the status of arguments with respect to an audience we produce an audience specific value-based argumentation framework.

Definition 2.2 (Audience-Specific VAF (AVAF)) *An audience-specific VAF is a 5-tuple $AVAF = \langle Ar, att, V, val, Valpref_a \rangle$, where Ar, att, V and val are as for a VAF, a is an audience, $a \in P$, and $Valpref_a$ is a preference relation (transitive, irreflexive and asymmetric), $Valpref_a \subseteq V \times V$, reflecting the value preferences of audience a. The AVAF relates to the VAF from which it is derived in that Ar, att, V and val are identical, and $Valpref_a$ is the set of preferences derivable from the ordering $a \in P$ in the VAF.*

Our purpose in introducing VAFs is to allow us to distinguish between one argument attacking another, and that attack succeeding, so that the attacked argument may or may not be defeated. Whether the attack succeeds depends on the value order of the audience considering the VAF. We therefore define the notion of defeat for an audience:

Definition 2.3 (Defeat for an Audience) *An argument ar* $defeats_a$ *an argument br for audience a,* ($defeats_a(ar, br)$)*, in an AVAF* $\langle Ar, att, V, val, Valpref_a \rangle$ *if and only if both* $attacks(ar, br) \in att$ *and not* $valpref(br, ar) \in Valpref_a$*.*

We can now define audience specific versions of the notions standardly associated with AFs:

Definition 2.4 (Acceptable to an Audience) *An argument* $ar \in Ar$ *is acceptable to an audience a with respect to set of arguments S,* ($acceptable_a(ar, S)$) *if:* $\forall(x)(x \in Ar \wedge defeats_a(x, ar) \rightarrow \exists(y)(y \in S \wedge defeats_a(y, x))$

Definition 2.5 (Conflict Free for an Audience) *A set S of arguments is conflict free for an audience a if:*
$\forall(x)\forall(y)(x \in S \wedge y \in S) \rightarrow (\neg(attacks(x, y) \in att) \vee (valpref(val(y), val(x)) \in Valpref_a))$

Definition 2.6 (Admissability for an Audience) *A conflict free for an audience a set of arguments S is admissible for the audience a if:* $\forall(x)(x \in S \rightarrow acceptable_a(x, S)$

Definition 2.7 (Preferred Extension for an Audience) *A set of arguments S in a value-based argumentation framework* $\langle Ar, att, V, val, Valpref_a \rangle$ *is a preferred extension for audience-a,* ($preferred_a$)*, if it is a maximal (with respect to set inclusion) admissible for audience a subset of* Ar*.*

A practical way of evaluating the status of arguments in an AVAF is to remove from the VAF all the unsuccessful attacks, those for which $valpref(br, ar) \in Valpref_a$, whereupon it can be treated as a standard AF. Thus for any AVAF, $vaf_a = \langle Ar, att_{avaf}, V, val, valpref_a \rangle$ there is a corresponding AF, $af_a = \langle A, att_{af} \rangle$ such that for $(x, y) \in att_{avaf}$, $(x, y), \in att_{af}$ if and only if $defeats_a(x, y)$. The preferred extension of af_a will contain the same arguments as the preferred extension for audience a of the VAF. Note that if the original VAF does not contain any cycles in which all arguments pertain to the same value, af_a will contain no cycles, since every cycle will be broken at the point at which the attack is from an inferior value to a superior one for audience a. Hence both af_a and vaf_a will have a unique, non-empty, preferred extension for such cases.

Theorem 2.8 ***Every AVAF with no single-valued cycles has a unique nonempty preferred extension.***

PROOF. Let $avaf$ be an $AVAF$, and let af be the standard argumentation framework resulting from removing all failing attacks. If $avaf$ is cycle-free, then af is cycle free and hence by Theorem 2.6 of Bench-Capon [2003a] it has a unique, not-empty preferred extension. But suppose avaf has a cycle. We know that this contains at least two values. Let v be the least preferred value in the cycle, and arg be the final argument in a chain relating to this value. The attack from arg to the next argument in the cycle will fail. Therefore this attack will not appear in af and the cycle will be broken at this point. This applies to all cycles in avaf. Therefore af is cycle free, and so has a unique, non-empty, preferred extension. QED

Moreover, since the AF derived from an AVAF contains no cycles, the grounded extension coincides with the preferred extension for this audience, and so there is a straightforward polynomial-time algorithm to compute it, given in [Bench-Capon, 2003a].

For the moment we will restrict consideration to VAFs which do not contain any cycles in a single value. For such VAFs, the notions of sceptical and credulous acceptance are of no relevance, since any given audience will accept only a single preferred extension. These preferred extensions may, and typically will, however, differ from audience to audience. We therefore introduce two useful notions: *objective acceptance*, arguments which are acceptable to all audiences irrespective of their particular value order, and s*ubjective acceptance*, arguments which can be accepted by audiences with the appropriate value order.

Definition 2.9 (Objective Acceptance.) *Given a VAF* $\langle Ar, att, V, val, Valpref \rangle$, *an argument* $a \in A$ *is objectively acceptable if and only if for all* $valpref \in Valpref$, *a is in every* $valpref$.

Definition 2.10 (Subjective Acceptance.) *Given a VAF* $\langle Ar, att, V, val, Valpref \rangle$, *an argument* $a \in A$ *is subjectively acceptable if and only if for some* $valpref \in Valpref$, *a is in that* $valpref$.

An argument which is neither objectively nor subjectively acceptable (such as one attacked by an objectively acceptable argument with the same value) is said to be *indefensible*.

All arguments which are not attacked will, of course, be objectively acceptable. Otherwise, objective acceptance typically arises from cycles in two or more values. For example, consider a three-cycle in two values, say two arguments with V_1 and one with V_2. The argument with V_2 will either resist the attack on it when it is preferred to V_1, or, when V_1 is preferred, fail to defeat the argument it attacks which will, in consequence, be available to defeat its attacker. Thus the argument with V_2 will be objectively acceptable, and both the arguments with V_1 will be subjectively acceptable. A more elaborate example is shown in Figure 1.

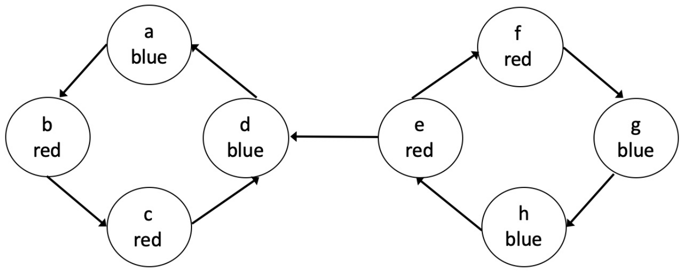

Figure 1. VAF with values red and blue

There will be two preferred extensions, according to whether $red > blue$, or $blue > red$. If $red > blue$, the preferred extension will be $\{e, g, a, b\}$, and if $blue > red$, $\{e, g, d, b\}$. Now e, g and b are objectively acceptable, but d, which would have been objectively acceptable if e had not attacked d, is only subjectively acceptable (when $blue > red$), and a, which is indefensible if d is not attacked, is also subjectively acceptable (when $red > blue$). Arguments c, f and h are indefensible. Results characterising the structures which give rise to objective acceptance are given in [6].

The question now arises as to whether it is possible to determine to which audiences an argument is acceptable. This question is fully explored in [Bench-Capon *et al.*, 2007].

2.2 Computational complexity results of value-based argumentation frameworks

Not long after VAFs were first proposed in the literature, a study was conducted on a number of decision problems in VAFs [Dunne and Bench-Capon, 2004]. In that paper it was shown that, for a given audience, those decision questions which are typically computationally hard in the standard Dungian AF setting, actually admit efficient solution methods in the value-based setting. The paper also highlighted a number of questions that arise solely in value-based frameworks that lack efficient decision processes.

The two key questions addressed in the paper concern the decision problems in VAFs of subjective and objective acceptance, as set out in Definitions 2.9 and 2.10 above. Concerning the decision problem of subjective acceptance, it is shown in [Dunne and Bench-Capon, 2004] that the complexity of this problem is NP-complete, and for objective acceptance, the decision problem is shown to be CoNP-complete. The paper also considers decision problems related to determining subjective acceptance by attempting to identify which pair-wise orderings are "critical" in that a given ordering will admit an audience for which an argument is subjectively accepted, whereas reversing this order will yield a situation in which the argument of interest is never accepted. Full results and their proofs are given in [Dunne and Bench-Capon, 2004]. Extrapolating from the results, they demonstrate that the identification of an argument as subjectively or objectively acceptable is just as hard as the corresponding problems of determining credulous and sceptical acceptance in standard argumentation frameworks; see [Dunne and Bench-Capon, 2002] for a full discussion of this point. Further complexity results, especially those concerning which audience can accept a given argument, can be found in [Bench-Capon *et al.*, 2007].

Further studies on computational complexity problems were later reported in [Dunne, 2010]. By considering properties of the directed graph structure defined by taking those values involved in conflicting arguments, Dunne identified an extensive class of argumentation systems for which the subjective and objective decision problems admit polynomial time solutions.

More recently, Nofal *et al.* [2014] examined specific questions in abstract argumentation frameworks under preferred semantics. They looked at the acceptance problem in standard argumentation frameworks, deciding whether a specific argument is in at least one preferred extension (i.e. it is credulously accepted) or in all such extensions (i.e. it is skeptically accepted). The paper presents an algorithm that enumerates all preferred extensions and builds algorithms that decide the acceptance problem without requiring explicit enumeration of all extensions. The improvements in efficiency brought about by the algorithms are achieved through a number of mechanisms: introduction of new labels for arguments' status, introduction of a new mechanism for pruning the search space so that transitions leading to dead ends are avoided at an early stage, and introduction of a cost-effective heuristic rule that yields earlier identification of arguments for transitions that might reach a goal state designating a preferred extension. The techniques developed for the acceptance problem in AFs are then used analogously to solve decision problems in VAFs, specifically deciding subjective and objective acceptance. Algorithms to solve these problems are defined and full proofs of the soundness and completeness of these algorithms is given in [Nofal *et al.*, 2014].

The studies referenced above set out properties of VAFs with a view to demonstrating their viability for use in domain applications. We now turn to considering how values are captured in accounts of structured argumentation.

3 Values in structured argumentation

In the previous section we showed how abstract value-based argumentation could be used to account for the subjective preferences which come into play when we are reasoning about how to make the world fit our desires. But the question arises: *how do values become attached to arguments?* The discussion in section 1 suggested that arguments for which value preferences are relevant are likely to arise in practical reasoning, reasoning about what to do. We will therefore begin our search for the link between arguments and values by looking at practical reasoning.

3.1 Practical syllogism

Practical reasoning was identified as different from theoretical reasoning by Aristotle in his *Nicomachean Ethics*, The discussion was revived by Anscombe [1978] and Kenny [1978]. Kenny's example of a practical syllogism is

K1: I'm to be in London at 4.15.
If I catch the 2.30 train, I'll be in London at 4.15.
So, I'll catch the 2.30 train.

Although Aristotle attempted to present the practical syllogism as a deduction, this position proved difficult to maintain, and Kenny's abductive presentation is now more common. It still has, however, a number of peculiarities.

- The conclusion is not really a prediction. Whether or not I actually catch the train is contingent on a number of things beyond my control. Rather it is a resolution, a *decision* to try to catch the train. The result of practical reasoning should not be a belief, but an action or a plan of action which will realise the desires one has decided to pursue.
- The truth of the premises is not enough to determine the decision. There may be earlier trains, and I may decide to catch one of those to be on the safe side. There may be many other ways of achieving the goal. Like any abduction, its soundness depends on it being the best (for me, in my current circumstance) way to achieve the goal.
- If I do catch the train, there will be many things that I cannot do. If I in fact prefer to do one of these things to being in London, then I may choose one of these other activities.
- There may be a number of other consequences of catching the train which are not desirable. These may be sufficiently undesirable that I decide not to catch the train.

These aspects are somewhat reflected in Searle's formulation in [Searle, 2003]:

S1: I want, all things considered, to achieve E.
The best way, all things considered, to achieve E is to do M.
So, I will do M.

In order to act on the basis of an argument such as K1, therefore, we need to consider alternative actions, alternative goals and any additional consequences, and then choose the best of these alternative goals and actions. Note the element of choice here: we can choose which of our goals we will seek to realise, and which actions to undertake to realise these goals. In order to decide which is best, I need to go beyond the goals themselves, and consider why these states of affairs are wanted. This is where values come in. It is our values that make certain states of affairs goals, because these states of affairs promote our values. In [Atkinson and Bench-Capon, 2014] there was a detailed discussion of how values give rise to a number of types of goal such as maintenance goals, achievement goals, avoidance goals and removal goals.

It is the values associated with these goals that determines which of them should be considered best by a particular person. Which is best will be determined by the preference ordering on values, and so may vary from person to person. Whether I decide to catch the train in K1 depends on the value served by being in London, and the values served by possible alternatives.

In order to assist with the formulation of a computational version of practical reasoning, we decided to propose an argumentation scheme, in the manner of [Walton, 1996].

3.2 Argumentation schemes

Walton's notion of an argumentation scheme is that it is a means of presumptive reasoning: if the premises are true, then we may presumptively draw the conclusion, subject to satisfactorily dealing with critical questions characteristic of the scheme.

Walton [1996] proposes two schemes relating to practical reasoning. The first is the *necessary condition scheme*

W1: G is a goal for agent *a*.
Doing action A is necessary for agent *a* to carry out goal G.
Therefore agent *a* ought to do action A.

The other was quite similar: the *sufficient condition scheme*.

W2: G is a goal for agent *a*.
Doing action A is sufficient for agent *a* to carry out goal G.
Therefore agent *a* ought to do action A.

Walton associates four critical questions with each of these schemes:

- WCQ1: Are there alternative ways of realising goal G?
- WCQ2: Is it possible to do action A?
- WCQ3: Does agent a have goals other than G which should be taken into account?
- WCQ4: Are there other consequences of doing action A which should be taken into account?

Although these arguments are fair reflections of the practical syllogisms K1 and S1, they have no link to values. As we saw above, values are essential for evaluation. Thus if critical question WCQ1 is posed, and it proves that there is an alternative action, say A2, without values we have no reason to say that this is a *better* alternative, and so choose to realise G with A2 rather than A.

For this reason we introduced an argumentation scheme which did have the required link to values. This scheme was first presented in [Atkinson *et al.*, 2004] and was more fully reported in [Atkinson *et al.*, 2006a]. The scheme was stated in [Atkinson, 2005] as:

AS1: In the circumstances R
we should perform action A
to achieve new circumstances S
which will realise some goal G
which will promote some value V.

In [Atkinson, 2005] and [Atkinson *et al.*, 2006a] sixteen critical questions were identified:

- CQ1: Are the believed circumstances true?
- CQ2: Assuming the circumstances, does the action have the stated consequences?
- CQ3: Assuming the circumstances and that the action has the stated consequences, will the action bring about the desired goal?
- CQ4: Does the goal realise the value stated?
- CQ5: Are there alternative ways of realising the same consequences?
- CQ6: Are there alternative ways of realising the same goal?
- CQ7: Are there alternative ways of promoting the same value?
- CQ8: Does doing the action have a side effect which demotes the value?
- CQ9: Does doing the action have a side effect which demotes some other value?
- CQ10: Does doing the action promote some other value?
- CQ11: Does doing the action preclude some other action which would promote some other value?
- CQ12: Are the circumstances as described possible?
- CQ13: Is the action possible?
- CQ14: Are the consequences as described possible?
- CQ15: Can the desired goal be realised?

- CQ16: Is the value indeed a legitimate value?

In [Atkinson and Bench-Capon, 2007c] a seventeenth CQ was added:

- CQ17: Can others act so as to take us to a state other than S?

This scheme allowed arguments for actions to be related to values: instantiating the scheme would give such an argument. Instantiating the critical questions would provide a means of attacking such arguments. This process of reasoning is illustrated in [Bench-Capon *et al.*, 2005] and [Atkinson and Bench-Capon, 2007c].

3.3 Semantics for structured value-based argumentation

In order to provide a semantic underpinning for this argument scheme and critical questions, use was made of the notion of Action Based Alternating Transition Systems (AATS) with values (AATS+V). These were introduced in [Atkinson and Bench-Capon, 2007a] and more fully reported in [Atkinson and Bench-Capon, 2007c].

An AATS is a type of state transition diagram, introduced in [Wooldridge and van der Hoek, 2005], formally based on Alternating-time Temporal Logic [Alur *et al.*, 2002]. In an AATS the states and transitions can be used to represent the current and future situations and the actions which will lead between them. In fact these transitions represent *joint actions*[5], that is the cumulative effect every agent relevant to the situation performing one action each. This means that a given action of a particular agent will appear in several transitions, depending on what the other agents do, and an agent may consequently not be in full control of the state that will be reached by using that action.

The definition of an AATS is:

Definition 3.1 (AATS ([Wooldridge and van der Hoek, 2005])) .
An Action-based Alternating Transition System *(AATS) is an* $(n+7)$*-tuple* $S = \langle Q, q_0, Ag, Ac_1, \ldots, Ac_n, \rho, \tau, \Phi, \pi \rangle$*, where:*

- Q *is a finite, non-empty set of* states*;*
- $q_0 \in Q$ *is the* initial state*;*
- $Ag = \{1,\ldots,n\}$ *is a finite, non-empty set of* agents*;*
- Ac_i *is a finite, non-empty set of actions, for each* $ag_i \in Ag$ *where* $Ac_i \cap Ac_j = \emptyset$ *for all* $ag_i \neq ag_j \in Ag$*;*
- $\rho : Ac_{ag} \rightarrow 2^Q$ *is an* action pre-condition function*, which for each action* $\alpha \in Ac_{ag}$ *defines the set of states* $\rho(\alpha)$ *from which* α *may be executed;*

[5]No suggestion of cooperation is intended here: the actions are joint solely in the sense that they are performed simultaneously.

- τ : Q × J_{Ag} → Q *is a partial* system transition function*, which defines the state* τ(q, j) *that would result by the performance of* j *from state* q. *This function is partial as not all joint actions are possible in all states;*

- Φ *is a finite, non-empty set of* atomic propositions*; and*

- $\pi : Q \to 2^{\Phi}$ *is an interpretation function, which gives the set of primitive propositions satisfied in each state: if* $p \in \pi(q)$*, then this means that the propositional variable* p *is satisfied (equivalently, true) in state* q.

As presented in [Wooldridge and van der Hoek, 2005], AATS have no values. Therefore they were extended in [Atkinson and Bench-Capon, 2007c] to include values, giving an AATS+V in which the transitions are additionally labelled with the values promoted or demoted by that transition. The additional definitions are:

Definition 3.2 (AATS+V ([Atkinson and Bench-Capon, 2007c])) *Given an AATS, an AATS+V is defined by adding two additional elements as follows:*

- V *is a finite, non-empty set of values.*

- $\delta : Q \times Q \times V \to \{+, -, =\}$ *is a* valuation function *which defines the status (promoted (+), demoted (–) or neutral (=)) of a value* $v_u \in V$ *ascribed to the transition between two states:* $\delta(q_x, q_y, v_u)$ *labels the transition between* q_x *and* q_y *with one of* $\{+, -, =\}$ *with respect to the value* $v_u \in V$.

With this definition it is possible to describe the practical reasoning argumentation scheme and critical questions in terms of the extended AATS+V. This gives us:

AS2 In the initial state $q_0 = q_x \in Q$,
Agent $i \in Ag$ should participate in joint action $j_n \in J_{Ag}$ where ${j_n}^i = \alpha_i$,
Such that $\tau(q_x, j_n)$ is q_y,
Such that $p_a \in \pi(q_y)$ and $p_a \notin \pi(q_x)$, or $p_a \notin \pi(q_y)$ and $p_a \in \pi(q_x)$,
Such that for some $v_u \in Av_i$, $\delta(q_x, q_y, v_u)$ is +.

We may now state the critical questions in these terms also.

- CQ1: $q_0 \neq q_x$ and $q_0 \notin \rho(\alpha_i)$.

- CQ2: $\tau(q_x, j_n)$ is not q_y.

- CQ3: $p_a \notin \pi(q_y)$.

- CQ4: $\delta(q_x, q_y, v_u)$ is not +.

- CQ5: Agent $i \in Ag$ can participate in joint action $j_m \in J_{Ag}$, where $j_n \neq j_m$, such that $\tau(q_x, j_m)$ is q_y.
- CQ6: Agent $i \in Ag$ can participate in joint action $j_m \in J_{Ag}$, where $j_n \neq j_m$, such that $\tau(q_x, j_m)$ is q_y, such that $p_a \in \pi(q_y)$ and $p_a \notin \pi(q_x)$ or $p_a \notin \pi(q_y)$ and $p_a \in \pi(q_x)$.
- CQ7: Agent $i \in Ag$ can participate in joint action $j_m \in J_{Ag}$, where $j_n \neq j_m$, such that $\tau(q_x, j_m)$ is q_z, such that $\delta(q_x, q_z, v_u)$ is $+$.
- CQ8: In the initial state $q_x \in Q$, if agent $i \in Ag$ participates in joint action $j_n \in J_{Ag}$, then $\tau(q_x, j_n)$ is q_y, such that $p_b \in \pi(q_y)$, where $p_a \neq p_b$, such that $\delta(q_x, q_y, v_u)$ is –.
- CQ9: In the initial state $q_x \in Q$, if agent $i \in Ag$ participates in joint action $j_n \in J_{Ag}$, then $\tau(q_x, j_n)$ is q_y, such that $\delta(q_x, q_y, v_w)$ is –, where $v_u \neq v_w$.
- CQ10: In the initial state $q_x \in Q$, if agent $i \in Ag$ participates in joint action $j_n \in J_{Ag}$, then $\tau(q_x, j_n)$ is q_y, such that $\delta(q_x, q_y, v_w)$ is $+$, where $v_u \neq v_w$.
- CQ11: In the initial state $q_x \in Q$, if agent $i \in Ag$ participates in joint action $j_n \in J_{Ag}$, then $\tau(q_x, j_n)$ is q_y and $\delta(q_x, q_y, v_u)$ is $+$. There is some other joint action $j_m \in J_{Ag}$, where $j_n \neq j_m$, such that $\tau(q_x, j_m)$ is q_z, such that $\delta(q_x, q_z, v_w)$ is $+$, where $v_u \neq v_w$.
- CQ12: $q_x \notin Q$.
- CQ13: $j_n \notin J_{Ag}$.
- CQ14: $\tau(q_x, j_n) \notin Q$.
- CQ15: $p_a \notin \pi(q)$ for any $q \in Q$.
- CQ16: $v_u \notin V$.
- CQ17: ${j_n}^i = {j_m}^i$, $j_n \neq j_m$ and $\tau(q_x, j_n) \neq \tau(q_x, j_m)$.

This formal account of the practical reasoning argumentation scheme and critical questions enable them to be used in agent systems designed to model practical reasoning scenarios; examples of these are provided in Section 4.

3.4 Dialogue interactions: values in persuasion and deliberation

In the previous sections we have considered reasoning with a specific audience which can determine the value order and evaluate the arguments accordingly. Often, however, values are crucial in dialogues where we have two or more audiences each with their own value order. Two distinct types of such dialogue are *persuasion* and *deliberation* [Walton and Krabbe, 1995].

In persuasion it is the person being persuaded who determines the value order [Bench-Capon, 2002c], since people will accept an argument only if it is acceptable on their own value ordering. Thus the proponents may not be able to use the arguments which convinced them because these will be acceptable on their value order, but perhaps not on the value order of the person they wish to persuade. Thus in a persuasion dialogue it is often necessary to elicit the value order of the other person, so that arguments acceptable to them can be found. Sometimes, however, it will not be possible to find arguments acceptable to the other, in which case the persuader must first try to get them to accept a value ordering and then to accept the argument which is the topic of the dialogue. Such dialogues are modelled in [Bench-Capon and Modgil, 2009]. A strategy for efficient persuasion in dialogues is given in [Atkinson *et al.*, 2012].

Deliberation is different in that while the value orders may well differ, neither party can determine what it should be. Therefore one purpose of a deliberation dialogue is to find a value ordering which will be acceptable to all concerned, so that a solution corresponding to this order can be found, which should be acceptable to all the parties. A set of speech acts to support dialogues based on this view of deliberation is given in [Atkinson *et al.*, 2013] and a tool showing how these speech acts can be used to generate persuasion and deliberation dialogues in agent systems is described in [Kirchev *et al.*, 2019].

4 Key applications of value-based argumentation

In this section we will illustrate the use of value-based argumentation in a number of domains.

4.1 General practical reasoning

We will begin by looking at the use of value-based argumentation in general practical reasoning. Our example will be that used in [Atkinson and Bench-Capon, 2007c], which adapts a well known brain teaser. AI students may be familiar with it as it is often used to illustrate search problems.

The situation is that a farmer is returning from market with a chicken (C), a bag of seeds (S) and his faithful dog (D). He needs to cross a river, and there is a boat (B) but it can only carry the farmer and one of his possessions. He cannot leave the chicken and seeds together because the chicken will eat the seeds. Similarly, he cannot leave the dog and the chicken unattended together because the dog will eat the chicken. His problem is how to organise his crossing.

We will represent the states by two lists, one for the items on the right bank, and one for items on the left. Thus [BCDS, _] will be selected from Q as the initial state, q_0. The complete set of states is shown in Figure 2

The transitions will be formed by various joint actions. We will assume that the animals will eat if they can, and so ignore the possibility of, for example, leaving the dog and chicken alone, and the dog doing nothing. This gives us the following six joint actions.

j_1: All do nothing

j_2: Farmer rows alone, chicken eats seeds if possible, dog eats chicken if possible

j_3: Farmer rows seeds, dog eats chicken if possible

j_4: Farmer rows dog, chicken eats seeds if possible

j_5: Farmer rows chicken, animals do nothing

j_6: All continue their journey home.

We can also identify a number of possible values[6]:

P: Progress - Promoted when farmer moves one of his possessions to the right side of the river. It is demoted when a state is revisited (through the always undesirable "goal" of repetition), and, to a lesser extent, when a possession is rowed from the right bank to the left (Pr). Rowing an item back is preferred to repetition, since repeating a state cannot be progress, whereas reaching a new state by returning an item to the left bank might be on a solution path, even though a *prima facie* backwards step.

S: Farmer has seeds - demoted when farmer loses seeds.

S: Farmer has chicken - demoted when farmer loses chicken.

F: Friendship - promoted when farmer travels with dog (it was for this companionship that he brought the dog with him).

We assume that the farmer values his possessions most, then wishes to make progress, and then have the joys of companionship. His value order is thus

$$C, S > P > Pr > F$$

We can now apply the joint actions to q_0 and label the transitions according to how they promote or demote the values. Initially five of the six actions are available, since the preconditions for j_6 are not satisfied. The result is shown as the first layer of Figure 2.

We can see that the only action which promotes a value without demoting a preferred value is j_5, and so the farmer will row the chicken, using the following argument:

- Farmer should row the chicken to promote Progress.

From q_2 three actions are possible. But two of them demote progress by reaching previous states, so the argument is

[6]Some labels for values are the same as the propositions used in the state description. The context makes it clear which is intended.

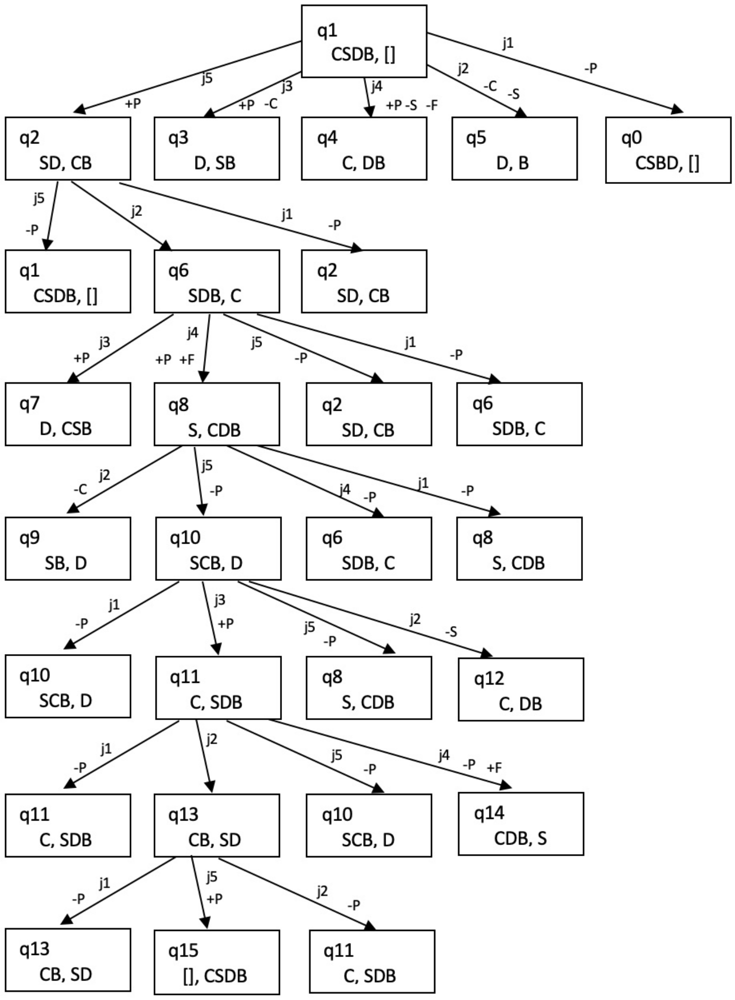

Figure 2. AATS+V for the farmer's river crossing problem. Note that when seeds and chicken are eaten, they no longer appear in the state descriptions.

- Farmer should not row the seeds, or do nothing, as that would demote progress. So Farmer should row alone.

Having reached q_6, there are two actions which promote progress, rowing the seeds, and rowing the dog. But rowing the dog additionally promotes friendship, and so that will be chosen. From $q8$ the only harmless action is to row the chicken to reach q_{10}. From q_{10} progress can be promoted by rowing the seeds, while all other actions demote a value. From here the only neutral action is to row alone to reach q_{13}. From here the farmer can promote progress by rowing the chicken. Now at last everything is on the right bank, and progress can be made by them all proceeding home.

This example shows how the puzzle can be solved by simply considering how to best promote values at every move. No look ahead is needed. In the standard search approach there are two variants. In the practical reasoning version this is resolved because in q_6 the farmer chooses to row the dog, to promote friendship as well as progress. For another example of practical reasoning, deciding whether to travel by aeroplane or train, see [Bench-Capon and Atkinson, 2009].

4.2 Domain-specific application: law

A domain in which value-based argumentation has been widely used is law, and in that domain arguing with values precedes abstract value-based argumentation and the formal modelling of structured argument with values by over a decade. The notion of values was in introduced to AI and Law by [Berman and Hafner, 1993][7]. In that paper Berman and Hafner noted that when using factor-based reasoning [Bench-Capon, 2017], often there were factor based arguments for both sides which needed to be chosen between. Factor-based reasoning as proposed in HYPO [Rissland and Ashley, 1987] and CATO [Aleven and Ashley, 1995], however, offered no rationale for choosing between them. The answer given in [Berman and Hafner, 1993] was that the arguments which better served the purposes of the law should be accepted. This idea was elaborated into a more formal theory of reasoning with cases as theory construction, in which value preferences were derived from precedents which were then applied to new cases, in [Bench-Capon and Sartor, 2001] and [Bench-Capon and Sartor, 2003], which was was the basis of the CATE [Chorley and Bench-Capon, 2005b] and AGATHA [Chorley and Bench-Capon, 2005a] systems. In [Greenwood *et al.*, 2003] it was proposed that the argumentation scheme for practical reasoning, latterly described in [Atkinson *et al.*, 2006a] and discussed above, could be used to generate the case based arguments required by factor-based reasoning and link them to values. The wild animals cases of [Berman and Hafner, 1993] had been modelled as a Dung style argumentation framework in [Bench-Capon, 2002a]. These various strands were brought together in [Bench-Capon *et al.*,

[7]Berman and Hafner used *purposes* rather than *values*, but they functioned in the same way. We will use *purpose* and *value promoted* as synonymous.

2005], which added values to the AF of [Bench-Capon, 2002a], and evaluated the arguments according to the resulting VAF.

In [Berman and Hafner, 1993] the example cases were some well known property law cases (often used as an introduction to property law in US law schools) concerning wild animals. That paper discussed three cases:

- *Keeble v Hickergill (1707).* This was an English case in which Keeble rented a duck pond, to which he lured ducks, which he shot and sold for consumption. Hickergill, out of malice, scared the ducks away by firing guns. The court found for Keeble. Two arguments for Keeble are possible: that he was engaged in an economically valuable activity, and that he was operating on his own land. The former reading is adopted in [Berman and Hafner, 1993], but others, e.g. [Bench-Capon and Rissland, 2001], prefer the latter.

- *Pierson v Post (1805).* In this New York case, Post was hunting a fox with hounds. Pierson intercepted the fox, killed it with a handy fence rail, and carried it off. The court found for Pierson. The argument was that Post had never had possession of the fox. The argument that hunting vermin is a useful activity which needs protection and encouragement formed the basis of the minority opinion. In this case, because of its legal setting, the original complainant, Post, whose role corresponds to the plaintiff in the other cases, is named second. We shall, however, refer to Post as the plaintiff and Pierson as the defendant to maintain consistency of role with the other cases.

- *Young v Hitchens (1844).* In this English case, Young was a commercial fisherman who spread a net of 140 fathoms in open water. When the net was almost closed, Hitchens went through the gap, spread his net and caught the trapped fish. The case was decided for Hitchens. The basis for this was that Young had never had possession of the fish, and that it was not part of the court's remit to rule as to what constituted unfair competition.

Later work [Bench-Capon and Rissland, 2001] also included four other cases in the discussion:

- *Ghen v Rich (1881).* In this Massachusetts case, Ghen was a whale hunter who harpooned a whale which subsequently was not reeled in, but was washed ashore. It was found by a man called Ellis, who sold it to Rich. According to local custom, Ellis should have reported his find, whereupon Ghen would have identified his lance and paid Ellis a fee. The court found for Ghen.

- *Conti v ASPCA*[8] *(1974).* In this New York case, Chester, a parrot owned by the ASPCA, escaped and was recaptured by Conti. The ASPCA found

[8] The American Society for the Prevention of Cruelty to Animals.

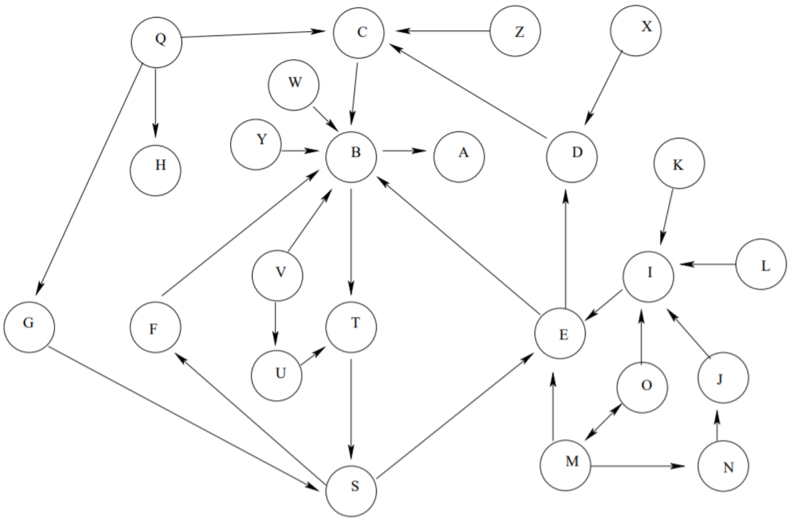

Figure 3. Dung style AF for wild animals cases, as given in [Bench-Capon, 2002a]

this out and reclaimed Chester from Conti. The court found that they were within their rights to do so.

- *New Mexico vs Morton (1975)* and *Kleepe vs New Mexico (1976)*. These two cases concerned the ownership of unbranded burros normally present on public lands, which had temporarily strayed off them. Both were won by the state.

These seven cases were formalised as a Dung style AF in [Bench-Capon, 2002a] and this was also used in [Bench-Capon *et al.*, 2005]. It is shown in Figure 3.

The twenty six arguments, the arguments they attack and the values associated with them in [Bench-Capon *et al.*, 2005] are shown in Table 1.

The basic approach in [Bench-Capon, 2002a] was to remove the arguments not applicable to a particular case and then consider preferred extensions. Then if argument A was sceptically acceptable, the plaintiff would win, but otherwise the defendant would win (the burden of proof is on the plaintiff). This, however, is not straightforward in the Dung style AF since there are even-length cycles in the AF, and so there will be multiple preferred extensions, some with A and some without.

The cycles in question are:

- the two-cycle M-O, which arises in *Pierson*

Table 1. **Arguments in the Wild Animal Cases.**
CL = Clear law, UA = Useful activity, PR = Protect property rights, EA = Economic activity, CR = The court should not make law

ID	Argument	Attacks	Values
A	Pursuer had right to animal		claim
B	Pursuer not in possession	A, T	CL
C	Owns the land so possesses animals	C	PR
D	Animals not confined by owner	C	
E	Effort promising success made to secure animal made by pursuer	B, D	CL
F	Pursuer has right to pursue livelihood	B	EA
G	Interferer was trespassing	S	PR
H	Pursuer was trespassing	F	PR
I	Pursuit not enough (Justinian)	E	CL
J	Animal was taken (Justinian)	I	CL
K	Animal was mortally wounded (Puffendorf)	I	CL
L	Bodily seizure is not necessary (Barbeyrac), interpreted as animal was brought within certain control (Tompkins)	I	UA
M	Mere pursuit is not enough(Tompkins)	E, O	CL
N	Justinian is too old an authority (Livingston)	J	
O	Bodily seizure is not necessary (Barbeyrac), interpreted as reasonable prospect of capture is enough (Livingston)	I, M	UA
Q	The land was open	G, H, C	PR
S	Defendant in competition with the plaintiff	E, F	EA
T	Competition was unfair	S	EA
U	Not for courts to regulate competition	T	CR
V	The iron holds the whale is an established convention of whaling	B, U	CR
W	Owners of domesticated animals have a right to regain possession	B	PR
X	Unbranded animals living on land belong to owner of land	D	PR
Y	Branding establishes title	B	
Z	Physical presence (straying) insufficient to confer title on owner	C	CL

- the four-cycle B-T-S-E, which arises in *Young*
- the four-cycle B-T-S-F, which arises in *Young*

This is precisely the situation for which Berman and Hafner commended the use of values: we need to choose between M, which promotes *clarity*, and O which promotes *useful activity*. In the actual case of *Pierson*, clarity was chosen, so that M was able to resist the attack of O, and so A was not in the preferred extension.

In the case of the two four-cycles that appeared in *Young*, the case was in fact resolved by the acceptance of U, which claimed that deciding what constituted unfair competition was outside the remit of the court. With T defeated, S defeats F, and so defends B. Similarly, S also defeats E and so B is acceptable. Now B defeats A and so the defendant won. Note that V was absent from *Young*. It was, however, present in *Ghen*, which concerned whaling, an industry long governed by clear conventions. Here the courts felt that just as it was not in their remit to determine what was unfair competition, neither could they overturn established conventions on the matter. Thus V was able to defeat U and B and so enable the plaintiff to win. This was forced in the standard AF, but in a VAF the attack from U to T can be resisted by preferring the value of *economic activity* to that promoted by restricting the court's remit, which would enable Young to win, even in the absence of an applicable convention. Such a shift in attitude may well occur (attitudes to regulation of competition swing back and forth), and so *Young* may at some future time be overturned. This illustrates a feature of value-based argumentation in law: because value preferences can change, the outcome of a case may be different at different times and in different jurisdictions. This captures the essence of the role of values in this kind of legal reasoning. A more elaborate discussion in [Bench-Capon *et al.*, 2005] also investigates the role of intermediate concepts [Lindahl and Odelstad, 2004] in moving from facts to legal conclusions.

Further discussions of value-baed reasoning in the wild animals cases can be found in [Bench-Capon, 2003b] and [Atkinson and Bench-Capon, 2007b]. In [Wyner *et al.*, 2007] an additional case, *Popov v Hayashi* [Atkinson, 2012] was included. This celebrated case[9], concerned a record breaking home run baseball hit by Barry Bonds of the San Francisco Giants. There was a struggle amongst crowd members over its possession. Popov first laid his glove on the ball, but Hayashi emerged from the ensuing melee with the ball. The wild animals cases were cited in the case, analogies being drawn between the hunted animals and the "fugitive baseball" [Finkelman, 2001]. This case and the wild animals cases were further discussed in [Bench-Capon, 2012].

4.3 Domain-specific application: e-participation

Another domain in which value-based argumentation has proved effective is e-participation. Political disputes often turn on disagreement as to values, and so this is a natural way to model such disputes. In PARMENIDES [Atkinson *et al.*, 2006b], a policy was presented for critique by members of the public through a software tool. The policy was presented as an instantiation of the practical reasoning scheme AS1 given above. Thus the policy was presented

[9]It was the subject of the 2004 comedy documentary film *Up for Grabs* https://www.imdb.com/title/tt0420356/

in terms of an understanding of the current situation and what it was meant to achieve in terms of facts, goals and values. The user was then given the opportunity to critique the policy in terms of relevant critical questions characteristic of the scheme[10]. In this way disagreement with the policy could be made precise, and different motives for disagreement identified. For example, different people might doubt whether the current situation was indeed as suggested, others might doubt that the policy would achieve its ends, and yet others might oppose these ends because rejecting the values they promote. PARMENIDES was further developed in [Cartwright and Atkinson, 2009] and later PARMENIDES formed the basis for the development of the Structured Consultation Tool (SCT) [Bench-Capon *et al.*, 2015], produced as part of the IMPACT project[11]. The SCT enabled not only a policy proposal to be critiqued, but also for the users to make proposals of their own, which could be automatically critiqued by instantiating critical questions [Wardeh *et al.*, 2013].

We will base our example in this chapter on that of [Atkinson *et al.*, 2011], which was also used in [Wyner *et al.*, 2012]. The example is an issue in UK Road Traffic policy. The number of fatal road accidents is an obvious cause for concern, and in the UK there are speed restrictions on various types of road, in the belief that excessive speed causes accidents. The policy issue which we will consider is how to reduce road deaths. One option is to introduce speed cameras to discourage speeding.

Following [Atkinson and Bench-Capon, 2007c] the first step is to build a model. In [Atkinson *et al.*, 2011] there was an extensive discussion of how to construct the model on the basis of responses received to a Green Paper[12]. Like [Wyner *et al.*, 2012] we will focus on the final refinement of the model presented in [Atkinson *et al.*, 2011], which includes responses from road safety organisations, motoring lobby groups, representations about financial constraints and civil liberties groups.

We now present the AATS+V. States are composed from the propositions considered relevant. In the model of [Atkinson *et al.*, 2011] the propositions that were considered are (given as pairs of positive and negative propositions):

R: The number of road deaths is acceptable; There are more road deaths than there should be.

S: Many motorists break the speed limits; Speed limits are generally obeyed.

P: Privacy is respected; There are additional intrusions on privacy.

[10]Not all critical questions were used: for example, those relating to the components of the model were not appropriate.

[11]Integrated Method for Policy Making Using Argument Modelling and Computer Assisted Text Analysis, in the European Framework 7 project (Grant Agreement No 247228) in the ICT for Governance and Policy Modeling theme (ICT-2009.7.3).

[12]A Green Paper is a Government publication issued as part of a consultation process that details specific issues, and then points out possible courses of action in terms of policy and legislation in order to receive feedback from interested parties.

These three propositions give rise to, potentially, eight states. We may, if we wish, exclude one or more of these as impossible. For example, if we believe that it is impossible that the number of road deaths is acceptable and yet many motorists break the speed limits, we may introduce constraints on states to filter it out. In [Wyner *et al.*, 2012], we specify all the possible states available for consideration. One state is designated as the current state:

- Many motorists break the speed limits $\wedge$ There are more road deaths than there should be $\wedge$ Privacy is respected.

We consider the impact of changes of state in terms of three values:

L: human life (Life);

B: the financial cost to the Government (Budget); and

F: the impact on civil liberties (Freedom). Here the principal concern is the impact on privacy.

The main agents involved are the *Government* (G), and *Motorists* (M), each considered as a body. In some cases the consequences of action are indeterminate (or at least cannot be determined using the elements we are modelling). To account for this we introduce a third agent, termed *Nature* (N). The action ascribed to Nature determines the outcomes of the actions of the other agents, where these outcomes are uncertain or probabilistic. We take the Government to be the independent agent, the one attempting to fulfill its values, while the actions of the Motorists and Nature are relative to its choices.

The Government has three actions: introducing speed cameras (G_1), educating motorists (G_2), or doing nothing (G_3). Motorists may reduce their speed or do nothing. Nature has two actions according to which fatal accidents are or are not reduced as a result of the Government and motorist actions. Actions are assumed to be always carried out together with other agents, represented as joint actions. The joint actions available are:

j_0: Government does nothing, motorists do nothing and nature does nothing.

j_1: Government introduces cameras, motorists do nothing and nature does nothing.

j_2: Government introduces cameras, motorists reduce speed and nature reduces accidents.

j_3: Government introduces cameras, motorists reduce speed and nature does nothing.

j_4: Government educates motorists, motorists reduce speed and nature reduces accidents.

j_5: Government educates motorists, motorists do nothing and nature reduces accidents. (Being more skilled, the drivers can cope with their excessive speed).

Finally, we have transitions, which relate a source state, a destination state, a joint action, a list of values promoted, and a list of values demoted. The joint action can only be carried out where, in some sense, the conditions for doing the action are met (e.g. where motorists are not speeding, then they cannot reduce speed) and result in a state that also makes sense (e.g. where motorists reduce speed and nature reduces accidents, then motorists are not speeding and accidents are reduced). We can presume that accidents are always reduced when motorists are educated since either they do not speed or can control their vehicles better. The transitions from q_0 are shown in Table 2. We are not interested in what happens in subsequent states. The whole AATS+V is shown as Figure 4.

Table 2. **Final Transition matrix.**

	j0	*j1*	*j2*
q0	⟨q0,-,-⟩	⟨q0,+B,-F⟩	⟨q5,+L+C,-F⟩
	j3	*j4*	*j5*
q0	⟨q6,+C,-F⟩	⟨q2,+L+C,-B⟩	⟨q3,+L,-B⟩

On the basis of this model, it seems that introducing speed cameras is a reasonable proposal. The hope is that this would induce motorists to cut their speed, and that the number of accidents would fall, so that j_2 is performed and q_5 is reached. This can be expressed in the form of AS1:

The current state is: Many motorists break the speed limits ∧ There are more road deaths than there should be ∧ Privacy is respected.

The action is: The government should introduce speed cameras.

The destination state is: Speed limits are generally obeyed ∧ The number of road deaths is acceptable ∧ There are additional intrusions on privacy.

The values promoted are: Life, Compliance

Note that the Government is expressing a preference for Life and Compliance over Freedom, which is demoted by the action.

This proposal can now be the subject of criticisms. For example,

CQI There might be disagreement as to the current situation: it would be possible to deny that many motorists broke the speed limits, or to claim that the number of road deaths was, if fact, acceptable.

CQ2 It might be argued that the action would not have the stated effects. Introducing speed cameras could reach q_4 or q_6 which would fail to prommote one or both of our values.

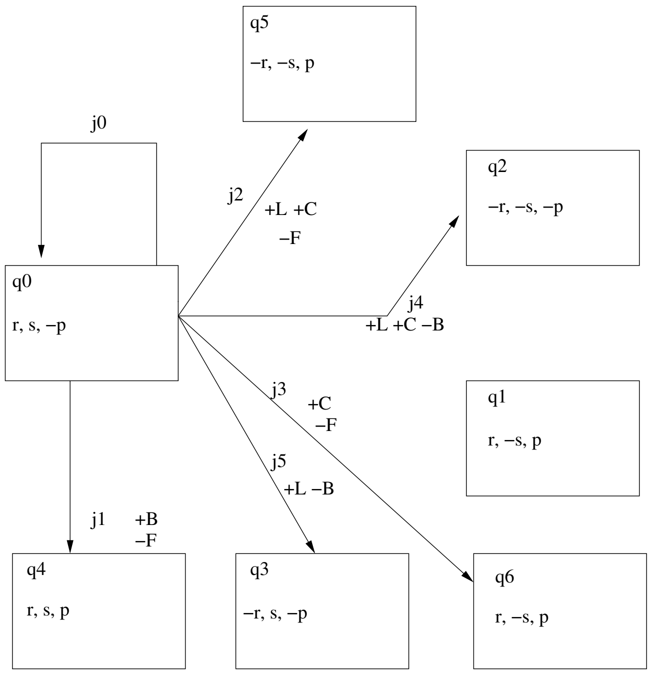

Figure 4. AATS+V for speed camera debate, as given in [Atkinson *et al.*, 2011]

CQ9 The action may demote a value. For example, freedom is demoted by the proposal.

CQ11 Other values can be promoted. There is no ground for the criticism in our example.

CQ13 It might be argued that the model is flawed and the proposed action is not possible. For example, it might be argued that the installation of speed cameras on the scale proposed was simply infeasible.

CQ17 Perhaps one or other of the agents will not perform the hoped for outcome. For example, it might be argued that reducing speed will not in fact reduce accidents and so the joint action will be j_3 leading to q_5 and so failing to promote life.

Using these methods to generate arguments, we can perform a full analysis. There are five arguments to perform an action from instantiating AS1.

PR1 We should perform G_1 to reach q_5 to promote L

PR2 We should perform G_1 to reach q_5 or q_6 to promote C

PR3 We should perform G_1 to reach q_4 to promote B

PR4 We should perform G_2 to reach q_2 or q_3 to promote L

PR5 We should perform G_2 to reach q_2 to promote C

Two arguments to refrain from an action are generated by a contrapositive form of AS1:

NPR1 We should not perform G_1 to avoid q_5 and q_6 since this would demote F

NPR2 We should not perform G_2 to avoid q_2 and q_3 since that would demote B

We accept that q_0 is the current state, and that other features of the model are correct. But we still have CQ17, which gives rise to three objections:

Ob1 Motorists may choose M_0 not M_1: attacks PR1, PR2 and PR5.

Ob2 Reducing speed may not reduce accidents and deaths. Attacks PR1.

Ob3 Motorists may choose M_1 not M_0: attacks PR3.

We now reach the final stage, when we weigh the merits and demerits of competing options, which requires us to identify the attacks between arguments. One source of attack is that a value is demoted: thus NPR1 attacks PR1, PR2 and PR3, and NPR2 attacks PR4 and PR5. Another source of attack, giving rise to symmetric attacks, is an alternative way of promoting the same value:

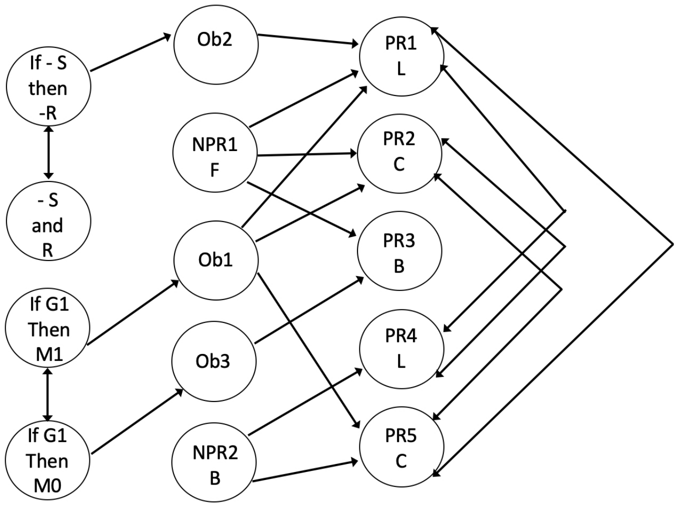

Figure 5. VAF for speed camera debate

thus PR1 and PR4 mutually attack, and PR2 and PR5 mutually attack. Finally we have different actions promoting different values: PR1 and PR5 and PR2 and PR4 mutually attack in this way. Finally we can have attacks which question the motive put forward: if PR1 is advanced to justify speed cameras, some may argue that the real expectation is that q_4 will be reached and that the real motive is to save money, rather than lives. This, however, does not challenge the action, but the justification, and we will not include these attacks here. Finally we have arguments representing the actual responses of motorists and nature to the introduction of speed cameras. These will form two two-cycles. We can now evaluate the arguments using a VAF. The VAF is shown in Figure 5.

On the left of the diagram are the two epistemic questions that need to be resolved. In default of anything better let us assume that, on the best evidence available, it is reasonable to expect that motorists will in fact reduce their speed, and that reducing speed will indeed lessen accidents and deaths. Having resolved these two cycles, we have answered the attacks from Ob1 and Ob2, while Ob3 is no longer attacked and will defeat PR3. When arguments are defeated, we can remove them and their attacks (and attacks on them) from the VAF to obtain the simpler VAF, as shown in Figure 6. Note that if we had made different assumptions about the epistemic questions a different VAF, and ultimately a different position, would result from this simplification. When an argument is not defeated, but its attack is resisted by a preferred argument, we mark it as *ineffective*. We cannot ignore it, since we have no argument to defeat it, but we will not act upon it. There are no such arguments as yet, since we have not yet exercised preferences, but only chosen between different

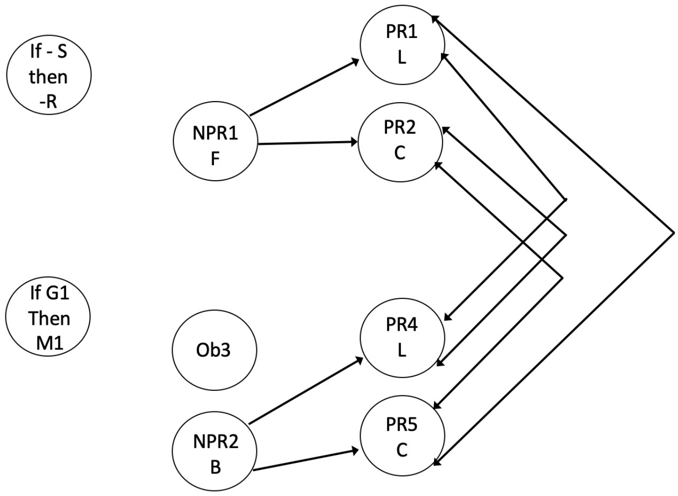

Figure 6. VAF for speed camera debate after epistemic choices

factual assumptions.

We next consider the two negative arguments based on PRAS2; once we have reached Figure 5 by resolving the epistemic questions, these are unattacked. These arguments will therefore succeed in defeating the arguments they attack unless the value of the attacked argument is preferred to that of the attacker. For NPR1 we must therefore consider Privacy/Freedom against Life to resolve the attack on PR1, and against Compliance to resolve the attack on PR2. A reasonable order would seem to be $L > F > C$: saying that intrusion on privacy is a necessary evil to save lives, but would not be acceptable simply to ensure compliance with speed limits without other gains. NPR1 thus becomes ineffective, which we show in the diagram by shading the argument node. This yields the VAF in Figure 7.

The final question to resolve is whether PR4 can be accepted given NPR2: that is, can we prefer L to B? Unfortunately we are regarding budget as a hard constraint and so we must answer that $B > L$. This means that PR4 falls, leaving a preferred extension for an audience with of $B > L > F > C$ comprising: the two factual assumptions, that motorists will reduce their speed, and that less speed means fewer accidents and deaths; the accepted course of action to install cameras to save lives; and two other considerations, that privacy must unfortunately be lessened (represented by the undefeated but ineffective argument), and that budgetary constraints preclude education as an alternative (represented by Obj3). Of course similar reasoning with different assumptions and different value orders would produce different results. If we assumed that motorists would continue to speed with the same value order, we would still install the cameras, but this time on the basis of PR3. If we made

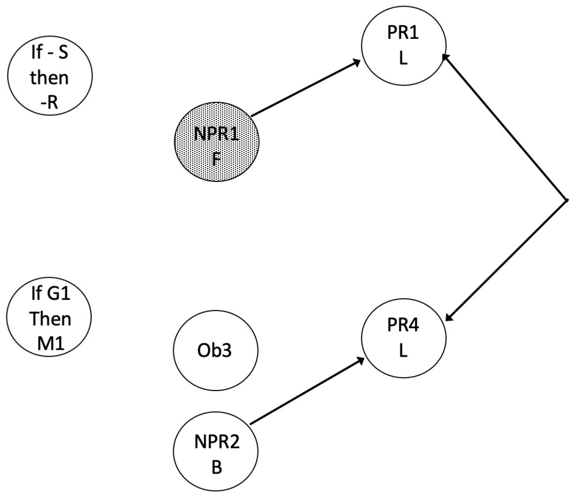

Figure 7. VAF with preferences $L > F > C$

the original assumptions but used the value order $B > F > L > C$, we could do nothing, since we would have no way of saving lives without infringing privacy that we could afford, and if we had the value order $F > B > L > C$, we would educate motorists rather than install cameras.

Finally, if we prefer life to freedom, but money is available so that it was possible to prefer L to B, we would have two equally valued arguments, PR1 and PR4, neither attacked except by each other. In this case we should be inclined to choose PR4, since this would mean that the undefeated NPR1 would no longer have to coexist with an argument it attacks, so that it no longer need be regarded as ineffective[13]. In this way we are able to respect the value of privacy, even though F is not preferred to L.

Considerations of these varied alternatives allow us to see how the policy positions favoured depend very critically on how we rank values: the acceptability of a proposal will often depend on whether the public mood has been correctly judged in this respect.

4.4 Domain-specific application: behavioural economics

Value-based reasoning has also been used to explore two "games" used in behavioural economics, the *Dictator Game* [Engel, 2011] and the *Ultimatum Game* [Oosterbeek *et al.*, 2004]. Classical economic theory assumes that people will behave in the manner of "economic man" described as follows by John Stuart Mill [Mill, 1844]:

[13]One disadvantage of the approach [Amgoud and Vesic, 2011] in which arguments which resist their attacks also defeat them is that it fails to distinguish between defeated arguments and those which must be acknowledged even though not followed.

> [Economics] is concerned with him solely as a being who desires to possess wealth, and who is capable of judging the comparative efficacy of means for obtaining that end.

However experiments performed in behavioural economics cast doubt on this key assumption. In the Dictator Game one player is given a sum of money and is then asked to give the second player as much or as little of it as he wishes. Classical economics would suggest that the player will give nothing, so maximising his own return. Experimentally, however, the results suggest otherwise: most players will give something to the other, sometimes as much as half. No studies report that the canonical model was observed. In one typical study [Forsythe *et al.*, 1994], given $10 to distribute, 79% of participants gave away a positive amount, with 20% giving away half. The mode sum given away in that study was $3. The explanation is that other values come into play here: suggestions include concern for the other, simple generosity, concern for image (no one likes to be thought selfish). This game was thoroughly explored using value-based reasoning in [Bench-Capon *et al.*, 2012]: here we will discuss the more interesting Ultimatum Game.

In the Ultimatum Game the first player is also given a sum of money and asked to decide how much he wishes to offer to the other player. But this time the second player can refuse, in which case both get nothing. Now classical economics suggests that the first player will offer the smallest amount possible and the second player will accept it because, for economic man, anything is better than nothing. As with the Dictator Game, these expectations are not borne out in practice. For example, Nowak and colleagues report that the majority of proposers offer 40–50% and about half of responders reject offers below 30% [Nowak *et al.*, 2000]. These results are robust, and, with some variations, are replicated in all the many studies. Oosterbeek *et al.* [2004] report a meta-analysis of 37 papers with 75 results from Ultimatum Game experiments, which have an average of 40% offered to the responder. The experiments of Henrich *et al.* [2001], carried out over 15 small-scale societies in 12 countries over five continents, report mean offers between 26% and 58%, and note that in some societies there is considerable variation in which offers are rejected: however, again, none suggests that the canonical model is followed by those making and responding to offers. The Ultimatum Game was modelled in [Bench-Capon *et al.*, 2012] and [Atkinson and Bench-Capon, 2018].

First we must model the game as an AATS+V. Obviously the states must include the money held by the two agents. We also wish to represent the reactions of the two players. When the offer is made, it is important whether the second player perceives it as fair, or as insulting. We therefore use a proposition which is true when the second player is annoyed by the offer made. At the end of the game we can consider the reaction of the first player. In particular, if the offer is rejected, a first player who made an ungenerous offer is likely to feel regret that he did not offer more. We therefore use a fourth proposition to record whether the first player feels regret.

Next we turn to actions. Obviously we need that the first player can offer n% of the available sum to the second player and that the second player can accept or reject it. The reception the offer receives will, however, depend critically on the size of n. We will therefore distinguish four cases: where $n > 50$, where $n = 50$, where $n > 0$ but < 50 and where $n = 0$. We should also recognise that the two actions are not chosen simultaneously, and that the choice to accept or reject will depend on how the second player reacts to the offer of the proposer. We therefore introduce a third action, in which the second player chooses a threshold, t, above which he will regard the offer as just, and below which he will feel insulted. We will assume that $t > 0$ and $t < 50$, discounting players who will not be satisfied with even an equal share. While the second player accepts and rejects, the first player can do nothing. This gives the set of joint actions shown in Table 3.

Table 3. Joint Actions in the Ultimatum Game

Joint Action	Player 1	Player 2
j1	A1:Offer > 50	B1:Set t < 50
j2	A2:Offer 50	B1:Set t < 50
j3	A3:Offer n < 50 and > 0	B1:Set t < n
j4	A3:Offer n < 50 and > 0	B1:Set t > n
j5	A5:Offer n = 0	B1:Set t > 0
j6	A4:Do nothing	B2:accept
j7	A4:Do nothing	B3:reject

Now we must identify some values and the transitions which promote and demote them. First there is economic value, the money, which we shall call M. This can be promoted in respect both of player 1 (M1) and in respect of player 2 (M2). These values are promoted to different degrees according to the size of the player's share. From the literature it appears that some people seem to value fairness, which we shall call E for equality. This is either promoted or not. Third we have the value of generosity (G), which again has been identified as a motivation by various experimenters. Whereas M will be promoted to varying degrees according to the amount of money, E is either promoted or not. What of G? Experimental evidence suggests that the impact of G does not increase as the amount given increases: we will therefore consider that G, like E, is either satisfied or not, and that any effect of the size of the gift is reflected in M2. Finally either player may be content with the outcome, and we represent this as C1 and C2. Again we will not model degrees of contentment. The AATS+V is shown in Figure 8.

Here will focus on the decision of the second player: the first player needs to think about this since the main aim of an offer is to have it accepted. The VAF for the second player is shown in Figure 9.

What the second player will do will depend on how it orders its values. Thus an offer above 50, or below 50 but above the second player's threshold

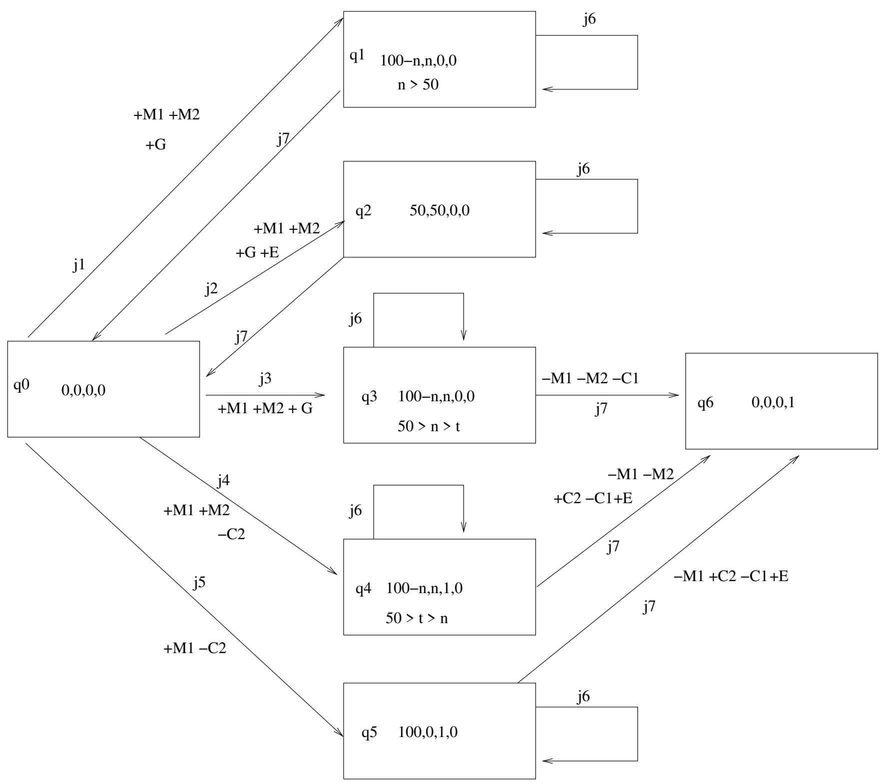

Figure 8. AATS+V for Ultimatum Game, as given in [Bench-Capon *et al.*, 2012]

Table 4. Value promotion and demotion in the Ultimatum Game

Joint Action	Proposal	Response	Promoted	Demoted
j1	vho	accept	M1,M2,G, C2	E
j2	vho	reject	G	M1,C1
j3	eo	accept	M1,M2,G,C2	
j4	eo	reject	G	M1,C1
j5	fo	accept	M1,M2	E
j6	fo	reject		M1
j7	lo	accept	M1,M2,C1	E,C2
j8	lo	reject		M1,C1

of acceptability (states q_1 and q_3), will only be rejected if the player prefers equality to both its own and the other player's, money: $E > M1, M2$. Given the set of values we have used, we would expect any player to accept an offer of half the sum, since rejecting in q_2 promotes nothing and demotes money for both players. If the second player is insulted by a non-zero offer and so is in q_4, however, he has a choice of whether to punish the first player and so restore its own pride, or to take the money. Normally we would expect that the player will prefer its own money and its own contentment to the money and contentment of the other agent, and so require $M2 > C2 > M1, C1$ for acceptance, or $C2 > M2 > M1, C1$ for rejection. If E is preferred to both $M2$ and $C2$ the second player will also reject the offer, but here motivated by a desire for equality, rather than the insult. Finally if a zero offer is made we would expect rejection, either because of the insult, or because equality is desired. Indeed a zero offer will only be accepted if the second player prefers the others player's money or contentment to its own contentment: $C1, M1 > C2$. This would be an extreme example of altruism, and we would expect it to be rare. This ordering would also lead to acceptance in q_4.

What the second player will do is crucial. In [Atkinson and Bench-Capon, 2018] the Ultimatum Game was used to explore how an agent can take account of the expected actions of others. There the three actions of our above model were compressed into a set of joint actions as shown in Table 4.

There we say that player one can make a very high offer (vho) of more than half, an equal offer (eo) of half, a fair offer (fo) at the threshold for the second player, or a low offer (lo), below that threshold. All of these may be accepted or rejected by the second player, giving eight joint actions, promoting and demoting values as shown. Note that equality cannot be promoted, since the initial state is one of equality. From this table we can see why most players will make a least a fair offer: only if the first player is desperate to "get one over" the other will a low offer be made, since only a low offer promotes $C1$ but carries with it a high probability of demoting $M1$ and $C1$. How high the offer will go depends on how much the player values the wealth and happiness

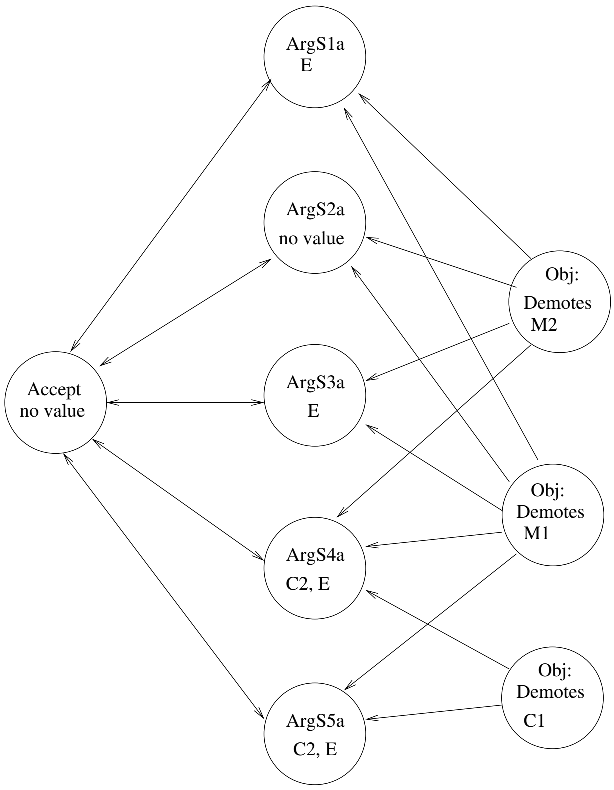

Figure 9. VAF for second player in Ultimatum Game, as given in [Bench-Capon *et al.*, 2012]

of the other, and whether it values a feeling of generosity.

In [Henrich *et al.*, 2001] 15 small scale societies from various parts of the world were studied, and it emerged that different groups behave differently. It was suggested that the different societies' actions in the Ultimatum Game could be accounted for in terms of the degree of cooperation and degree of commercial exchange found in daily life. We can relate these characteristics to a value profile. Suppose we associate the value of generosity with the cooperative groups such as the whale hunting Lamelara, and the recognition of C2 (the need to maintain good relations with the other) with commercial exchange. Those who do not engage in cooperative or exchange activities, we term solitary. In [Atkinson and Bench-Capon, 2018] it was found that using value profiles representing these three life styles predicted offers and rejections that are very close to the empirical results of [Henrich *et al.*, 2001].

4.5 Other applications

As well as these examples, value-based reasoning has been demonstrated using examples in medicine [Atkinson *et al.*, 2006c], health advice [van der Weide *et al.*, 2009] and [Di Tullio and Grasso, 2011], ontology alignment [Laera *et al.*, 2007] and [Trojahn *et al.*, 2007], an account of the emergence of norms [Bench-Capon and Modgil, 2017] and discussions of ethics [Atkinson and Bench-Capon, 2008]. Most recently in [Bench-Capon, 2020] value-based reasoning has been used as the basis of a novel computational account of virtue ethics in agent systems.

In general, value-based reasoning can be used to model argumentation and reasoning in any domain where the direction of fit is from an agent's desires or needs to the world; any situation in which reasoning about action is required. Such reasoning is pervasive, covering many of the most important aspects of life: from everyday choices such as where to eat or how to travel, to law and politics, and fundamental questions of how we should live.

5 Value-based reasoning at the meta-level

Modgil [2009] introduced an elegant and general way of handling preferences: instead of assigning different strengths to arguments, he permitted attacks to themselves be attacked. Such frameworks he termed *Extended Argumentation Frameworks* (EAF). This meant that an attack was unsuccessful not according to whether it was attacking a stronger argument, but according to whether it was itself defeated by some other argument.

The relation between VAFs and EAFs was explored in [Modgil and Bench-Capon, 2008]. A conflict between two arguments is shown as an EAF in Figure 10. There the value preferences are represented as arguments, attacking attacks which require the other preference to succeed. These value preference arguments will, of course, mutually attack. The desired audience represented as an ordering on the values will attack one of these attacks, resolving the framework.

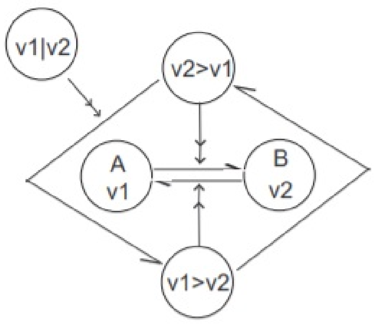

Figure 10. Value-based conflict in extended argumentation framework, as given in [Modgil and Bench-Capon, 2008]

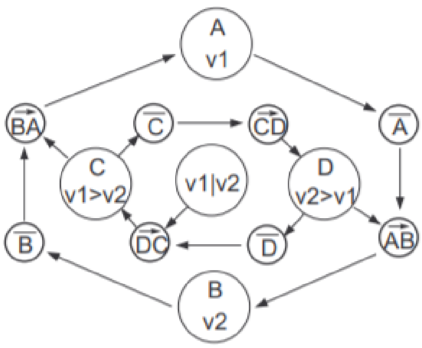

Figure 11. Meta level rewriting of Figure 10 to get a standard AF, as given in [Modgil and Bench-Capon, 2008]

Frameworks of the sort shown in Figure 10 can now be rewritten as standard Dung-style argumentation frameworks using meta level arguments. If we replace arguments by the fact that they are acceptable, e.g. A by *A holds*, and introduce arguments that arguments do not hold ($\overline{A}$) and that one argument defeats another ($\overrightarrow{AB}$), we can rewrite Figure 10 as Figure 11.

Now an attack $\overrightarrow{AB}$ may fail in two ways: either A may be defeated so that $\overline{A}$ defeats it, or there may be a preference argument that defeats it. There are clear simplifications in this rewriting in that standard AFs can be used instead of the more complicated VAFs and EAFs. The use of EAFs in value-based reasoning was discussed in [Bench-Capon and Atkinson, 2009], and its application to the representation of norms in [Bench-Capon and Modgil, 2019]. A full discussion of meta level argumentation can be found in [Modgil and Bench-Capon, 2011].

6 Concluding remarks

In this chapter we have discussed value-based reasoning. Philosophically it models reasoning where the direction of fit is from an agent's desires to the world: that is where an agent is choosing how to act in order to promote its values, and this covers all domains involving an element of practical reasoning, reasoning about what should be done.

Value-based reasoning was originally presented as a form of abstract argumentation extending Dung's original framework by giving arguments the additional property of *promoting a value*, and evaluating the arguments according to an ordering on those values.

Although there are some theoretical results, the main motivation for value-based reasoning was always applications, especially law where [Berman and Hafner, 1993] had drawn attention to the role of values in legal decisions, and [Bench-Capon and Sartor, 2003] had incorporated values into theories of case law for particular areas of law. This emphasis on applications was facilitated by the development of a means of doing structured value-based argumentation, based on an argumentation scheme and critical questions semantically underpinned by a form of state transition diagram, AATS+V.

Because of the importance of applications, we have devoted much of this chapter to a detailed discussion of four application domains: general problem solving, law, e-participation and behavioural economics.

Extended argumentation frameworks [Modgil, 2009] offer a means of generalising argumentation involving preferences. Value-based argumentation frameworks fit very well with this framework, since they can be systematically rewritten as standard AFs using meta level arguments describing the status of arguments in the VAF, the value preferences, and the audience concerned. Moving to meta level argumentation, however, does not remove the need for structured value-based argumentation, which is still needed to generate the arguments and attacks. This combination is used in [Bench-Capon and Atkinson, 2009].

The theory of value-based argumentation is fairly well understood, but its potential for modelling applications continues. As a means of representing problems in areas where values are crucial, such as ethics, law and politics, value-based reasoning offers a tried and tested solution.

Acknowledgments

In the many years that we have been working on computational models of value-based argument, we have collaborated with a wide range of people spanning many of the topics covered in this chapter. We are grateful to all colleagues with whom we have worked to progress research in this area: Latifa Al-Abdulkarim, Floris Bex, Elizabeth Black, Dan Cartwright, Alison Chorley, Sylvie Doutre, Paul Dunne, Floriana Grasso, Yanko Kirchev, Loredana Laera, Peter McBurney, Rolando Medellin-Gasque, Sanjay Modgil, Fahd Saud Nawwab, Samer Nofal, Henry Prakken, Giovanni Sartor, Valentina Tamma, Doug Walton, Maya Wardeh, Adam Wyner.

BIBLIOGRAPHY

[Aleven and Ashley, 1995] Vincent Aleven and Kevin D Ashley. Doing things with factors. In *Proceedings of the 5th International Conference on Artificial Intelligence and Law*, pages 31–41, 1995.

[Alur *et al.*, 2002] Rajeev Alur, Thomas Henzinger, and Orna Kupferman. Alternating-time temporal logic. *Journal of the ACM (JACM)*, 49(5):672–713, 2002.

[Amgoud and Cayrol, 1998] Leila Amgoud and Claudette Cayrol. On the acceptability of arguments in preference-based argumentation. In *Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence*, pages 1–7. Morgan Kaufmann Publishers Inc., 1998.

[Amgoud and Cayrol, 2002] Leila Amgoud and Claudette Cayrol. A reasoning model based on the production of acceptable arguments. *Annals of Mathematics and Artificial Intelligence*, 34(1-3):197–215, 2002.

[Amgoud and Vesic, 2011] Leila Amgoud and Srdjan Vesic. A new approach for preference-based argumentation frameworks. *Annals of Mathematics and Artificial Intelligence*, 63(2):149–183, 2011.

[Anscombe, 1978] G.E.M. Anscombe. On practical reasoning. In Joseph Raz, editor, *Practical Reasoning*, pages 33–45. Oxford University Press, 1978.

[Atkinson and Bench-Capon, 2005] Katie Atkinson and Trevor Bench-Capon. Legal case-based reasoning as practical reasoning. *Artificial Intelligence and Law*, 13(1):93–131, 2005.

[Atkinson and Bench-Capon, 2007a] Katie Atkinson and Trevor Bench-Capon. Action-based alternating transition systems for arguments about action. In *AAAI*, volume 7, pages 24–29, 2007.

[Atkinson and Bench-Capon, 2007b] Katie Atkinson and Trevor Bench-Capon. Argumentation and standards of proof. In *Proceedings of the 11th international conference on Artificial intelligence and law*, pages 107–116, 2007.

[Atkinson and Bench-Capon, 2007c] Katie Atkinson and Trevor Bench-Capon. Practical reasoning as presumptive argumentation using Action based Alternating Transition Systems. *Artificial Intelligence*, 171(10-15):855–874, 2007.

[Atkinson and Bench-Capon, 2008] Katie Atkinson and Trevor Bench-Capon. Addressing moral problems through practical reasoning. *Journal of Applied Logic*, 6(2):135–151, 2008.

[Atkinson and Bench-Capon, 2014] Katie Atkinson and Trevor JM Bench-Capon. Taking the long view: Looking ahead in practical reasoning. In *COMMA*, pages 109–120, 2014.

[Atkinson and Bench-Capon, 2018] Katie Atkinson and Trevor Bench-Capon. Taking account of the actions of others in value-based reasoning. *Artificial Intelligence*, 254:1–20, 2018.

[Atkinson *et al.*, 2004] Katie Atkinson, Trevor Bench-Capon, and Peter McBurney. Justifying practical reasoning. In *Proceedings of the Fourth International Workshop on Computational Models of Natural Argument (CMNA 2004)*, pages 87–90, 2004.

[Atkinson *et al.*, 2006a] Katie Atkinson, Trevor Bench-Capon, and Peter McBurney. Computational representation of practical argument. *Synthese*, 152(2):157–206, 2006.

[Atkinson *et al.*, 2006b] Katie Atkinson, Trevor Bench-Capon, and Peter McBurney. Parmenides: facilitating deliberation in democracies. *Artificial Intelligence and Law*, 14(4):261–275, 2006.

[Atkinson *et al.*, 2006c] Katie Atkinson, Trevor Bench-Capon, and Sanjay Modgil. Argumentation for decision support. In *International Conference on Database and Expert Systems Applications*, pages 822–831. Springer, 2006.

[Atkinson *et al.*, 2011] Katie M Atkinson, Trevor JM Bench-Capon, Dan Cartwright, and Adam Z Wyner. Semantic models for policy deliberation. In *Proceedings of the 13th International Conference on Artificial Intelligence and Law*, pages 81–90, 2011.

[Atkinson *et al.*, 2012] Katie Atkinson, Priscilla Bench-Capon, and Trevor Bench-Capon. A strategy for efficient persuasion dialogues. In *International Conference on Agents and Artificial Intelligence*, pages 332–347. Springer, 2012.

[Atkinson *et al.*, 2013] Katie Atkinson, Trevor Bench-Capon, and Douglas Walton. Distinctive features of persuasion and deliberation dialogues. *Argument & Computation*, 4(2):105–127, 2013.

[Atkinson, 2005] Katie Atkinson. *What Should We Do?* PhD thesis, University of Liverpool, 2005.

[Atkinson, 2012] Katie Atkinson. Introduction to special issue on modelling Popov v. Hayashi. *Artificial Intelligence and Law*, 20(1):1–14, 2012.

[Bench-Capon and Atkinson, 2009] Trevor Bench-Capon and Katie Atkinson. Abstract argumentation and values. In *Argumentation in artificial intelligence*, pages 45–64. Springer, 2009.

[Bench-Capon and Modgil, 2009] Trevor Bench-Capon and Sanjay Modgil. Case law in Extended Argumentation Frameworks. In *Proceedings of the 12th International Conference on Artificial Intelligence and Law*, pages 118–127. ACM, 2009.

[Bench-Capon and Modgil, 2017] Trevor Bench-Capon and Sanjay Modgil. Norms and value based reasoning: justifying compliance and violation. *Artificial Intelligence and Law*, 25(1):29–64, 2017.

[Bench-Capon and Modgil, 2019] Trevor Bench-Capon and Sanjay Modgil. Norms and extended argumentation frameworks. In *Proceedings of the Seventeenth International Conference on Artificial Intelligence and Law*, pages 174–178, 2019.

[Bench-Capon and Rissland, 2001] Trevor Bench-Capon and Edwina Rissland. Back to the future: Dimensions revisited. In *Legal Knowledge and Information Systems - JURIX 2001*, pages 41–52, 2001.

[Bench-Capon and Sartor, 2001] Trevor Bench-Capon and Giovanni Sartor. Theory based explanation of case law domains. In *Proceedings of the 8th international conference on Artificial intelligence and law*, pages 12–21, 2001.

[Bench-Capon and Sartor, 2003] Trevor Bench-Capon and Giovanni Sartor. A model of legal reasoning with cases incorporating theories and values. *Artificial Intelligence*, 150(1-2):97–143, 2003.

[Bench-Capon *et al.*, 2005] Trevor Bench-Capon, Katie Atkinson, and Alison Chorley. Persuasion and value in legal argument. *Journal of Logic and Computation*, 15(6):1075–1097, 2005.

[Bench-Capon *et al.*, 2007] Trevor Bench-Capon, Sylvie Doutre, and Paul Dunne. Audiences in argumentation frameworks. *Artificial Intelligence*, 171(1):42–71, 2007.

[Bench-Capon *et al.*, 2012] Trevor Bench-Capon, Katie Atkinson, and Peter McBurney. Using argumentation to model agent decision making in economic experiments. *Autonomous Agents and Multi-Agent Systems*, 25(1):183–208, 2012.

[Bench-Capon *et al.*, 2015] Trevor Bench-Capon, Katie Atkinson, and Adam Wyner. Using argumentation to structure e-participation in policy making. In *Transactions on Large-Scale Data-and Knowledge-Centered Systems XVIII*, pages 1–29. Springer, 2015.

[Bench-Capon, 2002a] Trevor Bench-Capon. Representation of case law as an argumentation framework. *Proceedings of JURIX 2002*, pages 103–112, 2002.

[Bench-Capon, 2002b] Trevor Bench-Capon. Value based argumentation frameworks. *arXiv preprint cs/0207059. Originally presented at NMR 2002*, 2002.

[Bench-Capon, 2002c] Trevot Bench-Capon. Agreeing to differ: modelling persuasive dialogue between parties with different values. *Informal Logic*, 22(3), 2002.

[Bench-Capon, 2003a] Trevor Bench-Capon. Persuasion in practical argument using value-based argumentation frameworks. *Journal of Logic and Computation*, 13(3):429–448, 2003.

[Bench-Capon, 2003b] Trevor Bench-Capon. Try to see it my way: Modelling persuasion in legal discourse. *Artificial Intelligence and Law*, 11(4):271–287, 2003.

[Bench-Capon, 2012] Trevor Bench-Capon. Representing Popov v Hayashi with dimensions and factors. *Artificial Intelligence and Law*, 20(1):15–35, 2012.

[Bench-Capon, 2014] Trevor Bench-Capon. Dilemmas and paradoxes: cycles in argumentation frameworks. *Journal of Logic and Computation*, 26(4):1055–1064, 2014.

[Bench-Capon, 2017] Trevor Bench-Capon. HYPO's legacy: Introduction to the virtual special issue. *Artificial Intelligence and Law*, 25(2):1–46, 2017.

[Bench-Capon, 2020] Trevor Bench-Capon. Ethical approaches and autonomous systems. *Artificial Intelligence*, page 103239, 2020.

[Berman and Hafner, 1993] Donald H Berman and Carole D Hafner. Representing teleological structure in case-based legal reasoning: the missing link. In *Proceedings of the 4th International Conference on Artificial Intelligence and Law*, pages 50–59. ACM, 1993.

[Cartwright and Atkinson, 2009] Dan Cartwright and Katie Atkinson. Using computational argumentation to support e-participation. *IEEE Intelligent Systems*, 24(5):42–52, 2009.

[Chorley and Bench-Capon, 2005a] Alison Chorley and Trevor Bench-Capon. Agatha: Using heuristic search to automate the construction of case law theories. *Artificial Intelligence and Law*, 13(1):9–51, 2005.

[Chorley and Bench-Capon, 2005b] Alison Chorley and Trevor Bench-Capon. An empirical investigation of reasoning with legal cases through theory construction and application. *Artificial Intelligence and Law*, 13(3-4):323–371, 2005.

[Di Tullio and Grasso, 2011] Eugenio Di Tullio and Floriana Grasso. A model for a motivational system grounded on value based abstract argumentation frameworks. In *International Conference on Electronic Healthcare*, pages 43–50. Springer, 2011.

[Dunne and Bench-Capon, 2002] Paul E. Dunne and Trevor J. M. Bench-Capon. Coherence in finite argument systems. *Artif. Intell.*, 141(1/2):187–203, 2002.

[Dunne and Bench-Capon, 2004] Paul E. Dunne and Trevor J. M. Bench-Capon. Complexity in value-based argument systems. In José Júlio Alferes and João Alexandre Leite, editors, *Logics in Artificial Intelligence, 9th European Conference, JELIA 2004, Lisbon, Portugal, September 27-30, 2004, Proceedings*, volume 3229 of *Lecture Notes in Computer Science*, pages 360–371. Springer, 2004.

[Dunne, 2010] Paul E. Dunne. Tractability in value-based argumentation. In Pietro Baroni, Federico Cerutti, Massimiliano Giacomin, and Guillermo Ricardo Simari, editors, *Computational Models of Argument: Proceedings of COMMA 2010, Desenzano del Garda, Italy, September 8-10, 2010*, volume 216 of *Frontiers in Artificial Intelligence and Applications*, pages 195–206. IOS Press, 2010.

[Engel, 2011] Christoph Engel. Dictator games: a meta study. *Experimental Economics*, 14(4):583–610, 2011.

[Finkelman, 2001] Paul Finkelman. Fugitive baseballs and abandoned property: Who owns the home run ball. *Cardozo L. Rev.*, 23:1609, 2001.

[Forsythe *et al.*, 1994] Robert Forsythe, Joel L Horowitz, Nathan E Savin, and Martin Sefton. Fairness in simple bargaining experiments. *Games and Economic behavior*, 6(3):347–369, 1994.

[Greenwood *et al.*, 2003] Katie Greenwood, Trevor Bench-Capon, and Peter McBurney. Towards a computational account of persuasion in law. In *Proceedings of the 9th International Conference on Artificial Intelligence and Law*, pages 22–31. ACM, 2003.

[Henrich *et al.*, 2001] Joseph Henrich, Robert Boyd, Samuel Bowles, Colin Camerer, Ernst Fehr, Herbert Gintis, and Richard McElreath. In search of homo economicus: behavioral experiments in 15 small-scale societies. *The American Economic Review*, 91(2):73–78, 2001.

[Horty and Bench-Capon, 2012] John Horty and Trevor Bench-Capon. A factor-based definition of precedential constraint. *Artificial Intelligence and Law*, 20(2):181–214, 2012.

[Kenny, 1978] A.J.P. Kenny. Practical reasoning and rational appetite. In Joseph Raz, editor, *Practical Reasoning*, pages 63–80. Oxford University Press, 1978.

[Kirchev *et al.*, 2019] Yanko Kirchev, Katie Atkinson, and Trevor J. M. Bench-Capon. Demonstrating the distinctions between persuasion and deliberation dialogues. In Max Bramer and Miltos Petridis, editors, *Artificial Intelligence XXXVI - 39th SGAI International Conference on Artificial Intelligence, AI 2019, Cambridge, UK, December 17-19, 2019, Proceedings*, volume 11927 of *Lecture Notes in Computer Science*, pages 93–106. Springer, 2019.

[Laera *et al.*, 2007] Loredana Laera, Ian Blacoe, Valentina Tamma, Terry Payne, Jerôme Euzenat, and Trevor Bench-Capon. Argumentation over ontology correspondences in mas. In *Proceedings of the 6th international joint conference on Autonomous agents and multiagent systems*, pages 1–8, 2007.

[Lindahl and Odelstad, 2004] Lars Lindahl and Jan Odelstad. Normative positions within an algebraic approach to normative systems. *Journal of Applied Logic*, 2(1):63–91, 2004.

[Maslow, 1943] Abraham Harold Maslow. A theory of human motivation. *Psychological review*, 50(4):370, 1943.

[Mill, 1844] John Stuart Mill. On the definition of political economy, and on the method of investigation proper to it. *Essays on some unsettled questions of Political Economy*, page 326, 1844.

[Modgil and Bench-Capon, 2008] Sanjay Modgil and Trevor Bench-Capon. Integrating object and meta-level value based argumentation. *Computational Models of Argument: Proceedings of COMMA 2008*, 172:240, 2008.

[Modgil and Bench-Capon, 2011] Sanjay Modgil and Trevor Bench-Capon. Metalevel argumentation. *Journal of Logic and Computation*, 21(6):959–1003, 2011.

[Modgil, 2009] Sanjay Modgil. Reasoning about preferences in argumentation frameworks. *Artificial Intelligence*, 173(9-10):901–934, 2009.

[Nofal *et al.*, 2014] Samer Nofal, Katie Atkinson, and Paul E. Dunne. Algorithms for decision problems in argument systems under preferred semantics. *Artificial Intelligence*, 207:23–51, 2014.

[Nowak *et al.*, 2000] Martin A Nowak, Karen M Page, and Karl Sigmund. Fairness versus reason in the ultimatum game. *Science*, 289(5485):1773–1775, 2000.

[Oosterbeek *et al.*, 2004] Hessel Oosterbeek, Randolph Sloof, and Gijs Van De Kuilen. Cultural differences in ultimatum game experiments: Evidence from a meta-analysis. *Experimental Economics*, 7(2):171–188, 2004.

[Perelman and Olbrechts-Tyteca, 1969] Chaim Perelman and Lucie Olbrechts-Tyteca. The new rhetoric: a treatise on argumentation, trans. *John Wilkinson and Purcell Weaver (Notre Dame, IN: University of Notre Dame Press, 1969)*, 190:411–12, 1969.

[Perelman, 1980] Chaim Perelman. *Justice, law, and argument: Essays on moral and legal reasoning*. Reidal, Dordrecht, 1980.

[Rissland and Ashley, 1987] Edwina L Rissland and Kevin D Ashley. A case-based system for Trade Secrets law. In *Proceedings of the 1st International Conference on Artificial Intelligence and Law*, pages 60–66. ACM, 1987.

[Schwartz, 1992] Shalom H Schwartz. Universals in the content and structure of values: Theoretical advances and empirical tests in 20 countries. *Advances in experimental social psychology*, 25(1):1–65, 1992.

[Searle, 2003] John R Searle. *Rationality in action*. MIT press, 2003.

[Trojahn *et al.*, 2007] Cássia Trojahn, Paulo Quaresma, and Renata Vieira. An extended value-based argumentation framework for ontology mapping with confidence degrees. In *International Workshop on Argumentation in Multi-Agent Systems*, pages 132–144. Springer, 2007.

[van der Weide *et al.*, 2009] Thomas L van der Weide, Frank Dignum, J-J Ch Meyer, Henry Prakken, and GAW Vreeswijk. Practical reasoning using values. In *International Workshop on Argumentation in Multi-Agent Systems*, pages 79–93. Springer, 2009.

[Walton and Krabbe, 1995] Douglas Walton and Erik Krabbe. *Commitment in dialogue: Basic concepts of interpersonal reasoning*. SUNY press, 1995.

[Walton, 1996] Douglas Walton. *Argumentation schemes for presumptive reasoning*. Lawrence Erlbaum Associates, 1996.

[Wardeh *et al.*, 2013] Maya Wardeh, Adam Wyner, Katie Atkinson, and Trevor Bench-Capon. Argumentation based tools for policy-making. In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Law*, pages 249–250, 2013.

[Wooldridge and van der Hoek, 2005] M. Wooldridge and W. van der Hoek. On obligations and normative ability: Towards a logical analysis of the social contract. *Journal of Applied Logic*, 3:396–420, 2005.

[Wyner *et al.*, 2007] Adam Wyner, Trevor Bench-Capon, and Katie Atkinson. Arguments, values and baseballs: Representation of Popov v. Hayashi. In *Proceedings of JURIX 2007*, pages 151–160, 2007.

[Wyner *et al.*, 2012] Adam Wyner, Maya Wardeh, Trevor Bench-Capon, and Katie Atkinson. A model-based critique tool for policy deliberation. In *JURIX*, pages 167–176, 2012.

Katie Atkinson
Department of Computer Science
University of Liverpool, UK
Email: katie@liverpool.ac.uk

Trevor Bench-Capon
Department of Computer Science

University of Liverpool, UK
Email: tbc@liverpool.ac.uk