Opinion Gathering Using a Multi-Agent Systems Approach to Policy Selection

(Extended Abstract)

Katie Atkinson
Department of Computer Science
University of Liverpool
Liverpool, UK
k.m.atkinson@liverpool.ac.uk

Trevor Bench-Capon
Department of Computer Science
University of Liverpool
Liverpool, UK
tbc@csc.liv.ac.uk

Adam Wyner
Department of Computer Science
University of Liverpool
Liverpool, UK
a.z.wyner@Liverpool.ac.uk

ABSTRACT

An important aspect of e-democracy is consultation, in which policy proposals are presented and feedback from citizens is received and assimilated so that these proposals can be refined and made more acceptable to the citizens affected by them. In this paper we present an innovative web-based application that uses recent developments in multi-agent systems, specifically models of computational argument designed to support reasoning about what should be done, to provide intelligent support for opinion gathering. We motivate the work with a discussion of the current difficulties with online opinion gathering systems and the need for a structured model of argumentation about agents and their actions to support opinion gathering. We then present work on an implemented system designed to gather opinions on policy proposals that is underpinned by a theory of argument representation, generation and evaluation based on a model of agents and their actions. We show how the underlying multi-agent system can offer support to improve survey systems, by eliciting a structured critique within a highly usable system.

1. INTRODUCTION

Current web technologies are both fuelling an increase in the desire of members of the public to participate in democratic debate and decision making, and are enabling governments to provide opportunities to do so. Not only do very many people subscribe to blogging and social networking sites, such as Twitter and Facebook, but there has also been a surge in the use of specialist sites that enable people to participate in democratic debate and decision making. For example, in the US, *RegulationRoom* is an academically hosted facility that provides guidelines for effectively commenting on proposed legislation.[1] The United Kingdom's Cabinet Office *Public Reading* website unfolds a proposed bill, allowing online readers to look at and comment on specific portions of draft legislation.[2] Whilst the inclusive nature of these initiatives is a supportive way to bring democracy to the people, many issues arise when one considers how to analyse, evaluate and respond to the volume of data gathered. Here we show how a multi-agent system can support a website gathering opinions about political policy proposals. The agents reason by exchanging arguments for and against proposals for action in relation to a particular policy. As such, the system is underpinned by a formal computational theory of argumentation that takes inspiration from recent work in agent systems on this topic. The model provides support to users in the domain of e-democracy in that policy analysts are facilitated in structuring their justifications of policy proposals that form input to the system. The opinions submitted by end users can then be automatically evaluated by the agents. The feedback given following the evaluation of arguments may lead to refinements of the model and justifications of the policies presented.

The paper is structured as follows. In section 2 we review the state of the art in current online tools for opinion gathering and we draw attention to particular open issues that motivate the approach we have taken in developing our survey tool. In section 3 we describe the main ideas from the literature on computational argumentation and multi-agent systems that we make use of to provide a solid grounding for our tool. In section 4 we introduce the running example that we use to demonstrate how a policy debate can be formally modelled and reasoned about in an agent system. The reasoning involves the automatic generation of arguments by agents, which are then evaluated for acceptability and from the set of acceptable arguments, some are selected for presentation to end users through the survey tool. In section 5 we discuss how our current prototype tool will be taken forward and extended. Finally, we give some concluding remarks in section 6.

2. PROBLEMS WITH ONLINE TOOLS FOR OPINION GATHERING

From a developer's point of view, a key consideration in designing and building online tools for opinion gathering is the trade-off between the amount of structure provided by the tool and its ease of learning and use. Since the target au-

[1] http://regulationroom.org/ Accessed October 7th, 2011.

[2] http://publicreadingstage.cabinetoffice.gov.uk/ Accessed October 7th, 2011.

Appears in: *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems – Innovative Applications Track (AAMAS 2012)*, Conitzer, Winikoff, Padgham, and van der Hoek (eds.), 4-8 June 2012, Valencia, Spain.

dience is the general public, participation must be fostered by making the interactive system as straightforward to use as possible. If, however, the responses are to be meaningfully analysed in terms of their content, then considerable structure needs to be imposed on the data.

Consider some current popular approaches to opinion gathering such as the UK e-petition[3], designed to enable users to create, view and sign petitions online. The motivation is stated on the site as follows: "e-petitions is an easy way for you to influence government policy in the UK". e-Petitions can address anything the government is responsible for and once a petition gets at least 100,000 signatures, it will be eligible for debate in the UK parliament. A similar site has recently been launched by the US government[4] where the policy is to issue an official response once the petition reaches a threshold number of signatures. Whilst these e-petitions are indeed easy to use, easy to respond to and facilitate signature collection (one particular petition in the UK gained over 1.81 million electronic signatures), the *quality* of engagement they offer is questionable. Such e-petitions are simply electronic versions of paper petitions, and they suffer from the same shortcomings as paper versions, the most significant being the conflation of a number of issues into one stock statement. It can only be assumed that by signing, the signatory agrees wholeheartedly with all of the (potentially) multiple points raised in the statement. This makes it easy to over simplify and to blur the issues since it is likely that individuals object for different reasons. Consider, for example, a recent petition proposing a reduction in the UK national speed limit on roads[5]. The petition objects that the reduction will not make a difference to road deaths and that the subsequent cut in carbon emissions will be too insignificant to justify the speed limit reduction. Signing such a petition is thus an 'all-or-nothing' statement with no room to discriminate between (or even acknowledge) the two very different objections raised; in a word, the petitions lack structure. The responses provided by the government are also at a general level and not able to recognise or address particular concerns, and so typically fail to satisfy anyone fully. We need reasons as well as opinions, arguments rather than assertions.

The clear separation of distinct issues is one problem with unstructured systems. A second important difficulty concerns how to assess and evaluate competing opinions. Putting the requirement on the user to provide arguments that are sound and coherent yields no guarantee this will be accomplished. Forming coherent and well expressed arguments is a rare skill, and people, including the highly educated, find it hard even to organise their thoughts into premises and a conclusion that follows validly from these premises. If, additionally, the arguments need to conform to, and be annotated with respect to, a structure requiring some minimum knowledge of argumentation theory (as with some current systems such as [9] and [11]) the difficulties are multiplied. Furthermore, checking the conformity of user input to an appropriate argument structure is itself a difficult task; despite advances in natural language processing, current technologies cannot yet analyse free text to the level required for the tasks under discussion here.

A key issue we have raised here with respect to tools that solicit user input in free text is thus how and where to impose structure to identify the arguments proposed so that the analysis of the opinions can be made meaningful, and even supported through computational analysis. An alternative is to oblige users to conform to a restrictive structure, but this may inhibit their interaction or require them to understand the underlying theory. Users may then make mistakes, and their responses be precise but wrong, which is even worse than being vague. Despite the difficulties, a number of systems have been developed with the intention of providing a better level of support. We briefly discuss some of the currently most popular and the issues they raise.

One category of tool is argument mapping tools. Araucaria [15] is one example which enables users to mark up the premises and conclusions of arguments, and indicate particular argumentation schemes identifying patterns of reasoning. Whilst the markup requires users to think more deeply about the structure of their arguments, there still remains no guarantee that the semantics of the marked up text is coherent and consistent since users simply decide what text to label as premises and conclusions and what the inferences are that follow. Moreover every different user is likely to produce a different, equally valid, markup. Another online argument mapping tool is Debatepedia[6]. It is an online "wiki" containing an ever growing collection arguments and debates within which users can express pros and cons of a range of issues. Although democratic in that users can freely modify others' contributions, the arguments entered are not required to conform to any particular semantics that would support coherence and argument evaluation, and so it is is often difficult to relate the various points made, and to evaluate the status of the debate.

Still more structure is imposed by systems that have been built that use the IBIS (Issue Based Information Systems) model of argument [17]. IBIS enables a particular problem/issue to be decomposed into a number of different positions. Arguments can then be created to attack or defend the positions until the issue is settled (possibly by a vote). A collaborative decision support system that uses this model is HERMES [11] (and its predecessor Zeno [9]). Evaluation showed that although users enjoyed using the system it was not easy to learn and difficulties were experienced understanding the argumentation content of the system, casting doubt on the usefulness of its output.

Given all the issues we have raised, we see a clear need for online opinion gathering tools to be grounded on some solid semantic foundation whilst retaining their usability. To achieve this, we look to multi-agent systems, and in particular how the reasoning of the agents in a system can be support by a computational model of argument. In the next section we pinpoint two key developments from this field that can provide the backbone of support for a tool for online opinion gathering.

3. ARGUMENT IN AGENT SYSTEMS

Since the publication of Dung's seminal paper [7], computational modelling of argument has become increasingly important as a sub-field of AI in general and MAS in particular. Leading AI conferences such as IJCAI, AAAI and AAMAS

[3]http://epetitions.direct.gov.uk/ Accessed October 7th, 2011.

[4]https://wwws.whitehouse.gov/petitions Accessed October 7th, 2011.

[5]See: http://petitions.number10.gov.uk/noNSLreduction/#detail

[6]http://www.debatepedia.org

typically now have *argumentation* as a specific keyword and have workshops (e.g. ArgMas which has been associated with AAMAS since 2004) devoted to the topic. Also there have been special issues of *Artificial Intelligence* and other journals and a successful biennial conference (COMMA) has been run since 2006. Argumentation has been found very useful in MAS both for modelling the dynamics of agent dialogues (e.g. [18]) and for reasoning about action selection (e.g. [14], [4]). Here we exploit two particular advances made in this area. Also relevant to our application from multi agent systems is work on norms for regulating agents and their interaction (e.g. [20], [1], [2]). Here we exploit three particular advances made in this area.

From [7] itself, we take the key notion that evaluating the status of an argument (determining whether it is acceptable or indefensible) takes place in the context of an argumentation framework, containing many arguments, and the status of an argument is *relative to a set of arguments*, which attack and defend it, both directly and indirectly. For Dung, arguments are entirely abstract, having only a defeat relation with each other. Dung's *abstract argument frameworks* [7] are thus a pair $AF = \langle \mathcal{X}, defeat \rangle$, where $\mathcal{X}$ is a set of arguments and *defeat* a binary relation on $\mathcal{X}$. A subset $\mathcal{Y}$ of $\mathcal{X}$ is called *conflict-free* if no argument in $\mathcal{Y}$ defeats an argument in $\mathcal{Y}$, and it is called *admissible* if it is conflict-free and defends itself against any attack, i.e., if argument x_1 is in $\mathcal{Y}$ and argument x_2 defeats x_1, then some argument in $\mathcal{Y}$ defeats x_2. A *preferred extension* is then a maximally (wrt set inclusion) admissible set. Dung defines several other types of extensions but we do not need them here. There is always a preferred extension, although it may be the empty set. There may be several preferred extensions if there are cycles in the defeat relation. If an argument is part of every preferred extension it is *sceptically acceptable*, if it is part of at at least one but not all, it is *credulously acceptable*, and it is *indefensible* otherwise. Important developments of the initial framework have included methods for distinguishing between successful and unsuccessful attacks. The defeat relation is replaced by an attack relation, and then a preference relation on arguments is used to remove unsuccessful attacks leaving only successful attacks (i.e. defeats), and so inducing a standard AF. Several kinds of preference have been suggested: we use an ordering on the social values promoted or demoted by acceptance of an argument [6], which yields Value-based Argumentation Frameworks (VAFs). The preferences have been generalised to the notion of arguments attacking (and possibly defeating) attacks in [12].

A second important development involves Argumentation Schemes, a notion imported from the study of argument in Informal Logic and Critical Thinking, but widely used in MAS (e.g. [16]). An argumentation scheme [19] may be taken as a stereotypical pattern of inference in which a number of premises of particular types give rise to the presumptive truth of some claim of a particular type, subject to a critique provided by a number of critical questions characteristic of the scheme used. The importance from our perspective is that such schemes provide us with guidance on how to construct, and how to attack arguments. Thus an argumentation scheme will tell us what components are needed to compose an argument of this type, and the critical questions will indicate how such an argument may be attacked. The argumentation scheme mainly used in this paper is a scheme for justifying the choice of an action as developed in [5]. This scheme can be stated as

> **PR**: In the current circumstances (R), action ac should be performed, since this will bring about a new set of circumstances (S) in which some goal (g) is realised. Realising g is desirable because it promotes a particular social value (v).

This scheme identifies the various components needed to establish that an action should be performed: an understanding of the current circumstances; the result of the action; a goal, and the value promoted which makes the goal desirable. Distinguishing between the resulting circumstances, S, what is desirable about those circumstances (g), and the social value with respect to which g is desirable (v) is important since all three are needed to justify the action, and all three may be established and attacked in different ways. As well as establishing what components must be assembled to justify an action, the scheme is associated with critical questions (seventeen in [5]), each of which indicates a way in which some other argument could attack the argument. Some of these simply challenge the truth of a premise, or supply a rebuttal the conclusion, but others may establish that the circumstances are such that the scheme cannot be used. With regard to this particular scheme, the existence of better actions is a particular source of challenges: while the argument may provide a justification, and so be a good argument, it can be defeated by a stronger justification. The inclusion of social values in the scheme provides a property of arguments on which their strengths can be compared, using a VAF.

One particular area of study in MAS is the interaction between independent agents and how this interaction can be managed so as ensure that the system as a whole operates in as harmonious and effective manner as possible. The analogy with the laws regulating human societies so as to smooth human-human interactions is obvious and so the idea of regulating agent societies using normative rules has been explored. One example of such work is [20]. In order to provide a semantical basis for modelling agents and their interactions a transition system based on joint actions between agents (Action-Based Alternating Transition Structure (AATS)) is introduced in [20]. These structures are grounded in Alternating-time Temporal Logic (ATL) [3]. AATS is also used in [5] to ground the argumentation scheme PR and its related critical questions by relating instantiations of the scheme and attacks based on the various critical questions to an AATS.

We claim that these theoretical developments taken from agent-based studies of computational argumentation can support our opinion gathering task in the following ways:

- *Modelling the Domain.* The need to underpin the enterprise with an AATS determines the components that we need and structures the task of identifying them.
- *Producing Arguments.* Instantiations of the Argumentation Scheme now give us arguments which can justify various actions in the situation as modelled, and various attacks on these arguments.
- *Selecting an Argument.* The arguments can now be organised into an Argumentation Framework (in particular a VAF). Choosing the best argument from those available requires us to make factual and preference assumptions, which can be modelled by the agents.

- *Receiving Feedback* The chosen argument, and various possible ways of attacking it, can now be offered to the public as a series of simple questions.

- *Evaluating Feedback* Given the precise attacks which various people wish to make, and the relative numbers who wish to make the different attacks, we can record these in the agent system and so reconsider the factual and value assumptions in the light of what is believed by the citizenry.

4. EXAMPLE

We will illustrate each of the aspects identified at the end of the previous section with a running example. We locate our implemented tool at a particular stage of (UK) policy making. Initially a Green Paper is issued which sets out a number of questions and concerns relating to a particular policy topic. Interested groups and individuals respond with their views, and then a concrete policy proposal is argued for in a White Paper. The proposed policy is in turn critiqued by the public and interested parties, whereupon the proposal is modified, abandoned, or implemented through detailed legislation. Our tool is intended to be used in forming the White Paper proposal and in gathering and analysing opinions on this proposal.

Our example concerns UK Road Traffic policy. The number of fatal road accidents is a cause for concern, and in the UK there are speed restrictions on various types of road, in the belief that excessive speed causes accidents. People, however, disobey these limits, and too many speed-related, avoidable, accidents and deaths occur. The policy issue we consider is how to reduce road deaths. The system we describe is to be used after a Green Paper has been issued and responses received. We suppose that among the solutions offered are: the introduction of speed cameras, supported by evidence of success in other countries and by pilot studies; and educational programmes, also supported by small pilot studies, but more expensive than cameras. Motoring lobbies will have denied that accidents can be seen as caused by speeding at all, and civil liberties groups will have denounced the use of cameras as an unacceptable intrusion on privacy. We now construct an AATS to reflect these concerns.

4.1 Modelling the Domain

Understanding the domain is the most effort intensive and intellectually demanding part of the process, whether or not a formal model is constructed. The target AATS, however, does at least provide the process with structure and methodology and provides a consistency check on the evolving understanding. As used in [5][7] an AATS is an $(n + 9)$ tuple $\mathcal{M} = \langle Q, q_0, Ag_1, ..., Ag_n, Ac_i, \rho, \tau, \Phi, \pi, V, \delta\rangle$. We will explain each element in turn, using our example. Remember the modelling is at a very coarse granularity, since we are seeking feedback on very broad brush policy proposals, rather than the fine-grained detail which will be required to implement them in legislation. Thus both propositions and actions may embrace a great deal of fine detail.

- Φ is a finite, non-empty set of *atomic propositions*; we will use three propositions: c, r, s. These are intended to represent *speed cameras are in operation*, *there are excessive road accidents and deaths*, and *speed limits are widely disregarded*, respectively.

- Q is a finite, non-empty set of *states*. These are the eight states corresponding to models of the three propositions.

- $\pi : Q \rightarrow 2^{\Phi}$ 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.

- $q_0 \in Q$ is the *initial state*; in our example, $c = 0, r = 1$ and $s = 1$.

- $Ag = \{1,...,n\}$ is a finite, non-empty set of *agents*. We are interested in three agents, g, (the Government), m (motorists collectively) and n (nature). Nature represents indeterminacy in actions, in particular, whether reducing speed will in practice reduce accidents.

- Ac_i is a finite, non-empty set of *actions*, for each $i \in Ag$ where $Ac_i \cap Ac_j = \emptyset$ for all $i \neq j \in Ag$. For our purposes the Government can introduce speed cameras (g_1), educate the public to make them speed aware and able to recognise the safe speed for given conditions (g_2) and nothing (g_0). Motorists may either cut their speed (m_1) or nothing (m_0). Nature is used since speeding may (n_1) or may not (n_0) reduce accidents.

- $\rho : Ac_{Ag} \rightarrow 2^Q$ 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; Essentially g_1 is only possible when $c = 0$, m_1 and n_1 are only possible when $s = 1$: other actions are always possible.

AATSs are particularly concerned with the joint actions of the set of agents. j_{Ag} is the joint action of the set of k agents Ag, and is a tuple $\langle\alpha_1,...,\alpha_k\rangle$, where for each α_j (where $j \leq k$) there is some $i \in Ag$ such that $\alpha_j \in Ac_i$. Moreover, there are no two different actions α_j and $\alpha_{j'}$ in j_{Ag} that belong to the same Ac_i. The set of all joint actions for the set of agents Ag is denoted by J_{Ag}, so $J_{Ag} = \prod_{i \in Ag} Ac_i$. Given an element j of J_{Ag} and an agent $i \in Ag$, i's action in j is denoted by j_i. The joint actions are shown in Table 1.

- $\tau : Q \times J_{Ag} \rightarrow Q$ is a partial *system transition function*, which defines the state $\tau(q, j)$ that would result by the performance of j from state q – note that, as this function is partial, not all joint actions are possible in all states (cf. the pre-condition function above); See Table 1 for transitions from q_0.

- V is a finite, non-empty set of values. We recognise l, life, b, budget, o obedience to the law, and p privacy.

- $\delta : Q \times Q \times V \rightarrow \{+, -, =\}$ is a *valuation function* which defines the status (promoted (+), demoted (–) or neutral (=)) of a value $v_u \in V$ by the transition between two states. V is set of social values. Moving from a state with $r = 1$ to one with $r = 0$ promotes l; moving to a state where $s = 0$ promotes o, moving from a state with $c = 0$ to one with $c = 1$ demotes p and any transition involving the expensive option g_2 demotes b. Finally if motorists continue to speed when there are

[7] The elements V and δ were added in [5]: the other elements were taken from [20].

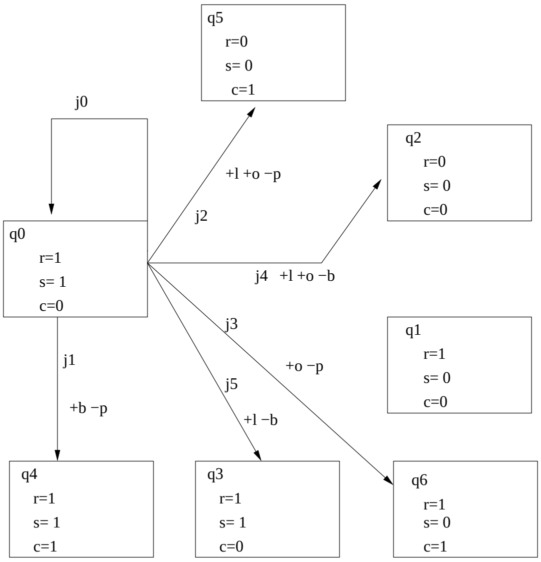

Figure 1: Example AATS $\mathcal{M}$

cameras, the increased detection will promote b. The values promoted and demoted by transitions are shown in Table 1. Figure 1 then shows the initial state, q_0, and the transitions from that state[8].

Table 1: Transition matrix from q_0 for $\mathcal{M}$ showing destination state, values promoted and values demoted.

Joint act.	Gov. action	Motor. action	Nature action	Result state	+	−
j0	g_0	m_0	n_0	$q0$	–	–
j1	g_1	m_0	n_0	$q4$	+b	-p
j2	g_1	m_1	n_1	$q5$	+l+o	-p
j3	g_1	m_1	n_0	$q6$	+o	-p
j4	g_2	m_1	n_1	$q2$	+l+o	-b
j5	g_2	m_0	n_1	$q3$	+l	-b

4.2 Producing Arguments

Given this model, arguments can now be automatically generated. We know that the initial state is q_0, and so every transition marked as promoting a value leading from q_0 will give an instantiation of the practical reasoning scheme:

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 o

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 o

Similarly every transition marked as demoting a value will give an argument against a particular action.

[8]To explain the headings for columns 6 and 7 of the table, these give the values promoted and demoted respectively.

NPR1: We should not perform g_1 to avoid q_4, q_5 and q_6 since this would demote p

NPR2: We should not perform g_2 to avoid q_2 and q_3 since that would demote b

In [5] seventeen ways of attacking an instantiation of PR were given. These attacks were related to three stages of practical reasoning: *problem formulation*, which disputes features of the AATS, an *epistemic* stage which queries either what is currently true, or the assumed action of another agent, and *option selection* which considers alternative proposals to discover which is best for the audience concerned. As $\mathcal{M}$ is given, we move directly to the epistemic stage. Since q_0 is given, the key unknowns are how the motorists will react and whether reduced speed will impact on accidents. That motorists may perform m_0 instead of m_1 represents an attack (Obj1) on PR1, PR2 and PR5. Supposing that accidents do not fall in response to speed (Obj2), attacks PR1, but not PR4, since it is held that the increased skills of motorists will reduce accidents and deaths even if they continue to speed. Finally PR3 assumes that motorists will do m_0 even when cameras are introduced, which can be questioned by Obj3. Thus we have three epistemic 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.

Each of these can be countered by assuming the contrary: Ass1, Ass2 and Ass3, respectively. Only PR4 is immune to epistemic objections, mainly because it succeeds in promoting the desired value whatever motorists and nature do. In order to answer Obj1-3 we need to step outside of our model to find arguments for our assumptions Ass1-3. For example we may quote the experience of other countries, or of pilot studies, to support our view that motorists will (or will not) reduce their speed in the presence of cameras. Accident statistics may be used to establish the proportion of accidents, and of fatal accidents, in which speeding was considered a crucial element. Much of the information we need will be available from the responses to the Green Paper. Considerations such as these - and the argumentation which supports them - takes us outside of deliberation. We are no longer concerned with *choice*, but with scientific questions of fact or causality (physical and psychological). Such questions therefore need to be resolved for the purposes of continuing the deliberation by agreeing to take a view on the facts, or reach consensus using these different argumentation schemes.

We now reach the final stage, when we weigh the merits and demerits of competing options. First we must identify the attack relation between our 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: thus PR1 and PR4 mutually attack, as do PR2 and PR5. Finally we have different actions promoting different values: PR1 and PR5, and PR2 and PR4 mutually attack in this way. Finally we

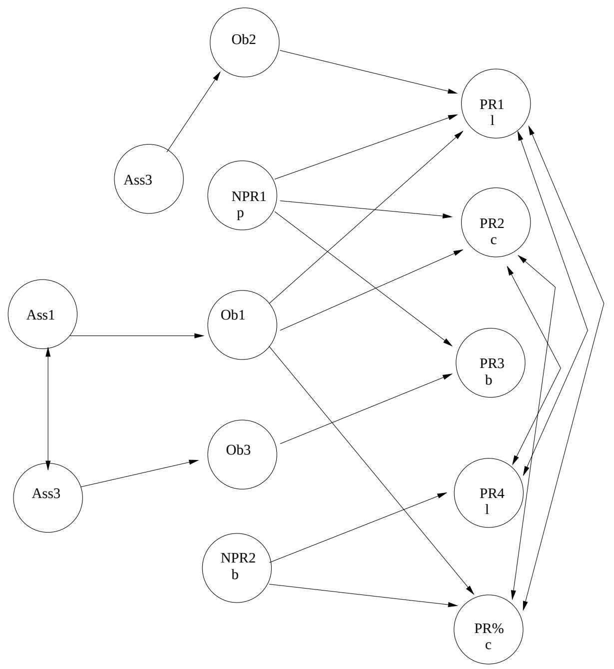

Figure 2: VAF for example

can have attacks which question the motive given: 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 the save money, rather than lives. This, however, does not challenge the action, but the justification, and, since we are interested in the policy itself rather than the particular motivation for it, we will not include these attacks here.

For purposes of clarity, we have specified the above arguments as though we were constructing them by hand. However, there is an implementation that enables an AATS to be constructed and instantiated to model particular scenarios. Given an instantiated AATS, the program can then automatically generate all the arguments and critical questions that can be produced, and subsequently organise them into a VAF and evaluate the arguments to produce the preferred extensions for given audiences. See [13] for the full technical details of this implementation.

Now that we have at hand a set of arguments, these can now be evaluated, as we describe in the next section.

4.3 Selecting an Argument

The arguments generated can now be organised into a Value Based Argumentation framework (VAF), as shown in Figure 2. VAFs are formally defined in [6]. Here we may say that an argument is defeated if it is attacked by an undefeated argument *unless* it relates to a value preferred to that of its attacker. Factual arguments defeat all value based arguments: practical arguments have the value they promote or demote. A set of arguments is mutually acceptable (*admissible*) if none defeats a member of the set, and some member of the set can defeat every argument attacking a member of the set. A maximal admissible set is a *preferred extension*.

To select an argument, assumptions about the response to actions need to be made. We will suppose that motorists do cut their speed if speed cameras appear (there is sufficient evidence from pilot studies), and that reducing speeding will cut accidents and deaths (there is abundant evidence that a large proportion of fatal accidents involve speeding cars). We can now induce a single preferred extension by imposing a value order. Since we cannot exceed our budget, b must be highest, and since the point is to save lives l should be the second highest. Suppose then that our value order is $b \succ l \succ p \succ c$. The preferred extension with this order and the assumptions about behaviour is {Ass1, Ass2, Obj3, NPR1, NPR2, PR1}. Thus we have an argument justifying the introduction of speed cameras. Note that NPR1 attacks, but does not defeat PR1 because of our preference of l to p. Had we preferred p to l, we would have had no justified action, unless we were willing to prefer both p and l to b so that PR4 could resist the attack of NRP2.

So as to provide a specific example to walk through, we have imposed the value order described above. If the value ordering is already settled, this is appropriate. However, since the agents can produce and reason about all arguments automatically, all the possible preferred extensions can be calculated if desired. This enables the full range of alternatives to be considered and the value ordering to be used can then be decided in the light of its consequences.

4.4 Receiving Feedback

The argument that was produced as output of the previous stage can now be presented to the public for feedback using the web-based tool. We solicit feedback on the model, both disagreements and omissions, the assumptions made, and the ordering of values chosen. After an initial statement of PR1 (if this is the selected argument), participants who disagree are led through a series of screens to identify the particular points at which they disagree, or want further justification.

- *Screen 1* asks about the current state. For each proposition, the participant is invited to agree or disagree. If there is disagreement, evidence is presented (e.g accident statistics). If the participant remains unconvinced, the argument supporting the premise can be critiqued. The first screen also asks the participant to list any other relevant facts that need to be considered.

- *Screen 2* asks questions such as *Do you agree that reducing road deaths promotes life*, so that each of the labellings of the relevant transitions in $\mathcal{M}$ can be questioned.

- *Screen 3* relates to the states reached by a transition. Participants are asked if the propositions of the next state (q_5) will result from the action. Disagreement will result in an argument justifying that transition being shown, and either participants will accept this and return, or be led through a critique of this further argument justifying the causal relationship. This screen also offers the opportunity to identify unstated consequences of the action thought relevant and undesirable.

- *Screen 4* offers a range of other actions (such as g_2) which participants may think achieve the aims of the policy. Selecting one of these leads to the reason for rejecting it (NPR2 in the example, together with the strict budget constraint). Any other alternative actions supported by participants may be entered as free text.

- *Screen 5* asks about values: whether participants endorse the values used, or want other values considered, and gives the opportunity to express their ordering of values.

4.5 Evaluating Feedback

When participants have submitted their opinions, we can see whether our proposed policy commands popular support and, if not, exactly why not. Screen 1 should confirm that the number of deaths and accidents are seen as a problem, and solicits factors other than speeding which may be a cause. A substantial write-in for *poor lighting*, coupled with later comments, would indicate that a different approach has popular support. Screen 2 is about the link between goals and values. It may be that people disagree that cameras represent an unacceptable intrusion on privacy, which would be good news for advocates of the policy represented by PR1. Screen 3 allows questioning of the underlying causal model. This is the opportunity to deny that speeding causes accidents, and to offer poor lighting as an alternative cause. Screen 4 gauges support for g_2, and, where people may suggest an alternative, such as improved lighting. The acceptability of the budgetary argument against g_2 is also shown. Finally Screen 5 tests our assessment of value priorities. We ranked life above privacy; this may be endorsed or disputed.

The various answers can now be applied to the VAF: we have numbers to attach to the arguments which will confirm or reject the assumptions made to resolve cycles. We also have responses which confirm or question the value order used. Finally we may be able to remove arguments found unattractive. Other responses will lead us to modify $\mathcal{M}$: modification may be the removal of labels from transitions where disagreement has been expressed, or a more fundamental refinement involving additional propositions and actions, as would be necessary if the lighting argument commanded sufficient support. The new model then generates a new VAF, which can be evaluated by the multi-agent system using the knowledge gained to inform the assumptions and value order.

5. DISCUSSION

We have described the full spectrum of the process that can be followed in order for public policies to be formally represented and reasoned about by agents, taking into account user-supplied opinions on the arguments presented through a survey tool. A version of a survey tool that used the argumentation scheme to organise questions to the public and to assimilate their answers, but which is not underpinned by an AATS or an agent system, is described in [8][9]. This tool was subjected to two separate evaluation studies, one using a group of PhD students and one using a group of academics working on argumentation spread throughout the world. The evaluations focused on issues such as ease of use, ease of expressing disagreement, and ease of expressing alternative opinions. The results here were broadly encouraging with a significant majority finding it easy to use the system and easy to express disagreement, but less successful in allowing alternatives to be expressed. From the analysis side, the responses can readily be assimilated into the structure, but the consequences of the response are less easy to identify. Thus the evaluation confirmed the utility of the argumentation scheme as an organising principle and the use of critical questions characteristic of the scheme to identify what the public should be asked, but some limitations were also identified .

It is to address these limitations that we have integrated the agent system and provided the more formal basis of the AATS. In particular because this allows the generation of all the alternative arguments and objections, based on different assumptions and different value orders, we are now able to suggest alternative arguments supported by the assumptions and value preferences expressed by users for their consideration and endorsement, rather than relying of the users to be able to propose these alternatives themselves, as described for Screen 4 above. Moreover the existence of the underlying model allows for a deeper level of justification. Where users disagree, they can be confronted by a specific counter argument. If this disagreement relates to an element of the model (such as the presence of a transition), this takes the form of the submission or submissions from which that element of the model was derived. Where the disagreement relates to the proposal of an alternative argument, the model allows counter arguments to that alternative to be generated and presented to the user, again as in Screen 4 above. Thus the use of the underlying agent system addresses the limitations of the original tool of [8].

The model also enables a number of software engineering advantages over the original as developed in [8]. We are able to support the model and the resulting argument system using a Protege ontology, so as to make entirely explicit the assumptions on which it is based, and to generate arguments using standard reasoners. Furthermore, although the arguments, being value-based, naturally lend themselves to evaluation through the use of a VAF, other appropriate inference engines could also be used for argument evaluation. For example, we have carried out an exercise to import the data in the semantic model into the Legal Knowledge Interchange Format (LKIF), then represent and reason with it automatically using the Carneades system [10]. This too has given the full range of arguments for and against the policy position. While we have carried out an initial experiment along these lines, the results remain to be finalised, and it is not the main line of this paper.

We believe that the improvements, both from a functional and a software engineering perspective, resulting from the use of the underlying AATS and the supporting agent system it enables, represent a significant advance on what was possible in [8], and represents an innovative and effective use of MAS techniques.

6. CONCLUDING REMARKS

In this paper we have described a web-based application which deploys state of the art of argumentation techniques taken from agent-based research to provide computational support for a particular stage of the policy making process: the production of a White Paper to solicit public feedback on a broadly expressed proposal. Here the effort is shifted from the construction of arguments to justify the proposal, and the understanding of free form responses, to achieving

[9]This tool can be used at SUPPRESSED FOR BLIND REVIEW.

a precise and formal understanding of the problem and its relevant aspects, to provide a model from which arguments can be generated automatically and into which responses can be assimilated. The interactivity offered by the web is exploited by enabling the exact points of objection to be pinpointed so that disagreement becomes specific and hence capable of being addressed specifically: by improved justifications; by modifications to the assumptions and understanding; or even by changes to the policy. The application illustrates how the full potential of the web and agent systems is achieved, not by making available and supporting existing, paper based, procedures, and so perpetuating the flaws in those processes, but rather by rethinking those procedures so that the opportunities offered can be grasped. Our immediate future work concerns further development of our tool to fully integrate the agent system and couple it with the existing web interface, and to make it sufficiently robust to allow it to be fielded for evaluation purposes. Following these extensions, a series of large scale field trials is already planned.

Acknowledgements

This work was partially supported by the FP7-ICT-2009-4 Programme, REMOVED FOR BLIND REVIEW Project, Grant Agreement Number REMOVED FOR BLIND REVIEW The views expressed are those of the authors and are not necessarily representative of the project.

7. REFERENCES

[1] T. Ågotnes, W. van der Hoek, M. Tennenholtz, and M. Wooldridge. Power in normative systems. In C. Sierra, C. Castelfranchi, K. S. Decker, and J. S. Sichman, editors, *AAMAS (1)*, pages 145–152. IFAAMAS, 2009.

[2] T. Ågotnes and M. Wooldridge. Optimal social laws. In W. van der Hoek, G. A. Kaminka, Y. Lespérance, M. Luck, and S. Sen, editors, *AAMAS*, pages 667–674. IFAAMAS, 2010.

[3] R. Alur, T. A. Henzinger, and O. Kupferman. Alternating-time temporal logic. *J. ACM*, 49(5):672–713, 2002.

[4] L. Amgoud, C. Devred, and M.-C. Lagasquie-Schiex. A constrained argumentation system for practical reasoning. In L. Padgham, D. C. Parkes, J. P. Müller, and S. Parsons, editors, *AAMAS (1)*, pages 429–436. IFAAMAS, 2008.

[5] K. Atkinson and T. J. M. Bench-Capon. Practical reasoning as presumptive argumentation using action based alternating transition systems. *Artificial Intelligence*, 171(10-15):855–874, 2007.

[6] T. J. M. Bench-Capon. Persuasion in practical argument using value based argumentation frameworks. *Journal of Logic and Computation*, 13(3):429–48, 2003.

[7] P. M. Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artif. Intell.*, 77(2):321–358, 1995.

[8] A. for Blind Review. *Blank*. PhD thesis, Blank, Blank.

[9] T. F. Gordon and N. Karacapilidis. The Zeno Argumentation Framework. In *Proceedings of the Sixth International Conference on Artificial Intelligence and Law (ICAIL '97)*, pages 10–18. ACM Press, 1997.

[10] T. F. Gordon, H. Prakken, and D. Walton. The carneades model of argument and burden of proof. *Artif. Intell.*, 171(10-15):875–896, 2007.

[11] N. I. Karacapilidis and D. Papadias. Computer supported argumentation and collaborative decision making: the HERMES system. *Information Systems*, 26(4):259–277, 2001.

[12] S. Modgil. Reasoning about preferences in argumentation frameworks. *Artif. Intell.*, 173(9-10):901–934, 2009.

[13] R. L. Parreira Duarte. Modelling arguments in the Dictator Game. Technical Report ULCS-10-006, Department of Computer Science, University of Liverpool, 2009.

[14] I. Rahwan and L. Amgoud. An argumentation based approach for practical reasoning. In H. Nakashima, M. P. Wellman, G. Weiss, and P. Stone, editors, *AAMAS*, pages 347–354. ACM, 2006.

[15] C. Reed and G. Rowe. Araucaria: Software for argument analysis, diagramming and representation. *International Journal on AI Tools (IJAIT)*, 13(4):961–979, 2004.

[16] C. Reed and D. Walton. Towards a formal and implemented model of argumentation schemes in agent communication. In I. Rahwan, P. Moraitis, and C. Reed, editors, *ArgMAS*, volume 3366 of *Lecture Notes in Computer Science*, pages 19–30. Springer, 2004.

[17] H. Rittel and W. Kunz. Issues as elements of information systems. Technical report, University of California, 1970. Working paper #131.

[18] Y. Tang and S. Parsons. Argumentation-based dialogues for deliberation. In F. Dignum, V. Dignum, S. Koenig, S. Kraus, M. P. Singh, and M. Wooldridge, editors, *AAMAS*, pages 552–559. ACM, 2005.

[19] D. Walton. *Argumentation Schemes for Presumptive Reasoning*. Lawrence Erlbaum Associates, NJ, USA, 1996.

[20] 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.