Measuring minimal change in argument premise revision

Mark Snaith and Chris Reed

Argumentation Research Group, School of Computing, University of Dundee, Dundee, UK, DD1 4HN
marksnaith@computing.dundee.ac.uk

Abstract. The field of belief revision studies how information can be given up in the face of new, conflicting information, while argumentation provides methods through which conflict can be modelled and the resultant acceptability of arguments evaluated. Prominent theories of belief revision depend on the notion of minimal change, measured in terms of epistemic entrenchment, to determine what beliefs to give up. In this paper, we take an initial look at the effects of removing an argument from a system of structured argumentation, in terms of both argument construction and acceptability, and how these can be used in the determination of minimal change.

1 Introduction

If a software agent is forced to accept information that conflicts with information that it currently possesses, it may be forced into giving up the original information. Conflict is a key area in argumentation, with the highly influential work of Dung [3] abstracting the nature of arguments and attacks between them. Dung's theory has been built on and expanded since the seminal paper; one recent development has been to instantiate the abstract approach by providing the arguments with structure, through the application of strict and defeasible inference rules to a knowledge base [7].

The process of removing an argument in a Dung-style framework is relatively straightforward, due to arguments being represented as single, abstract entities with no consideration for structure. However, when the arguments are given structure, a greater degree of flexibility is provided, in that giving up an entire argument can be done by, for instance, giving up a single premise. But with this flexibility comes a problem — complex arguments will contain multiple premises: exactly what premise(s) should be given up in order to remove the argument?

The field of belief revision aims to answer a more general version of this question in terms of belief sets — when an agent is required to give up a belief, and faces a choice as to exactly which belief, how does it make the choice? One of the most influential theories in belief revision is the AGM theory, which provides a set of postulates that describe valid *revisions*, *contractions* and *expansions* of belief sets [1]. These three processes are additionally guided by the concept of

minimal change, with "minimal" being measured in terms of epistemic entrenchment — those beliefs with the lowest degree of entrenchment are more willingly given up [4, 5].

Connections between argumentation and belief revision have recently found new momentum. The work of [8, 9, 6] on Argument Theory Change sees belief revision techniques employed to revise an argumentation system when a new argument is added, such that the argument becomes warranted. We wish to take a different approach to connecting argumentation and belief revision, by considering the application of belief revision techniques to the removal of arguments from a system of structured argumentation.

In this paper, we take an initial look at effects of removing an argument from an ASPIC$^+$ argumentation system, in terms of both argument construction and acceptability, and how these can be used the determination of minimal change.

The paper proceeds as follows: in section 2 we provide a brief introduction to the system of [7]; in section 3 we identify the effects of a change to argument premises and show how these can be used to realise an entrenchment ordering; in section 4 we provide an example to demonstrate the concepts that we presented and in section 5 we outline our conclusions and areas for possible future work.

2 Preliminaries

The ASPIC$^+$ framework [7] further developed the work of [2] and instantiates the abstract approach to argumentation in [3]. The basic notion of the framework is an argumentation system, $AS = \langle \mathcal{L}, ^-, \mathcal{R}, \leq \rangle$ where $\mathcal{L}$ is a logical language, $^-$ is a contrariness function from $\mathcal{L}$ to $2^{\mathcal{L}}$, $\mathcal{R} = \mathcal{R}_s \cup \mathcal{R}_d$ is a set of strict ($\mathcal{R}_s$) and defeasible ($\mathcal{R}_d$) inference rules such that $\mathcal{R}_s \cap \mathcal{R}_d = \emptyset$ and $\leq$ is a partial preorder on $\mathcal{R}_d$.

An argumentation system contains a knowledge base, $\langle \mathcal{K}, \leq' \rangle$ where $K \subseteq \mathcal{L}$ and $\leq'$ is a partial preorder on $\mathcal{K}/\mathcal{K}_n$. $\mathcal{K} = \mathcal{K}_n \cup \mathcal{K}_p \cup \mathcal{K}_a \cup \mathcal{K}_i$ where $\mathcal{K}_n$ is a set of (necessary) axioms, $\mathcal{K}_p$ is a set of ordinary premises, $\mathcal{K}_a$ is a set of assumptions and $\mathcal{K}_i$ is a set of issues.

From the knowledge base ($\mathcal{K}$) and rules ($\mathcal{R}$) arguments are constructed. For an argument A, $Prem(A)$ is a function that returns all premises in A; $Conc(A)$ is a function that returns the conclusion of A; $Sub(A)$ is a function that returns all sub-arguments of A; $DefRules(A)$ is a function that returns all defeasible rules in A; and $TopRule(A)$ is a function that returns the last inference rule used in A.

On the basis of these functions, A is:

1. p if $p \in \mathcal{K}$ with: $Prem(A) = \{p\}$, $Conc(A) = p$, $Sub(A) = p$, $DefRules(A) = \emptyset$,$TopRule(A) =$ undefined
2. $A_1, \cdots, A_n \rightarrow \psi$ if $A_1, \cdots, A_n$ are arguments such that there exists a strict rule $Conc(A_1), \cdots, Conc(A_n) \rightarrow \psi$ in R_s; $Prem(A) = Prem(A_1) \cup \cdots \cup Prem(A_n)$, $Conc(A) = \psi$, $Sub(A) = Sub(A_1) \cup \cdots \cup Sub(A_n) \cup \{A\}$, $DefRules(A) = DefRules(A_1) \cup \cdots \cup DefRules(A_n)$,$TopRule(A) = Conc(A_1), \cdots, Conc(A_n) \rightarrow \psi$

3. $A_1, \cdots, A_n \rightarrow \psi$ if $A_1, \cdots, A_n$ are arguments such that there exists a defeasible rule $Conc(A_1), \cdots, Conc(A_n) \Rightarrow \psi$ in R_s; $Prem(A) = Prem(A_1) \cup \cdots \cup Prem(A_n)$, $Conc(A) = \psi$, $Sub(A) = Sub(A_1) \cup \cdots \cup Sub(A_n) \cup \{A\}$, $DefRules(A) = DefRules(A_1) \cup \cdots \cup DefRules(A_n) \cup \{Conc(A_1), \cdots, Conc(A_n) \Rightarrow \psi\}$, $TopRule(A) = Conc(A_1), \cdots, Conc(A_n) \Rightarrow \psi$

An argument can be attacked in three ways: on a (non-axiom) premise (undermine), on a defeasible inference rule (undercut) or on a conclusion (rebuttal).

Given an argumentation system $\mathcal{AS}$ and a knowledge base $KB = \langle \mathcal{K}, \leq' \rangle$, an argumentation theory is $AT = \langle \mathcal{AS}, KB \leq \rangle$, where $\leq$ is an argument ordering on the set of all arguments that can be constructed from KB in $\mathcal{AS}$.

In this paper, we will use the following notations: $Args(\mathcal{AS})$ is the set of all arguments in $\mathcal{AS}$; when considering the acceptability of arguments in an argumentation theory, AT, on the basis of $\mathcal{AS}$ ($AT_{\mathcal{AS}}$), we will leave the semantics unspecified and instead refer to a (possibly empty or unit) set of S-extensions $E(AT_{\mathcal{AS}})$; $\mathcal{K}(\mathcal{AS})$ is the knowledge base in an argumentation system $\mathcal{AS}$ and $\mathcal{AS} \backslash D$ is an argumentation system such that $\mathcal{K}(\mathcal{AS} \backslash D) = \mathcal{K}(\mathcal{AS}) \backslash D$.

3 Measuring minimal change

In classic theories of belief revision, the process is guided by *minimal change*, which is measured not just in terms of the logical consequences of removing a belief, but by an *entrenchment ordering* placed on beliefs — those beliefs with a lower degree of entrenchment will be more willingly given up [4, 5]. Nevertheless, logical consequences play an important part in arriving at this ordering — intuitively, an agent will be less likely to give up a belief that is fundamental to a significant number of its other beliefs.

The process of removing an argument from a system of structured argumentation involves making some modification to the system such that the argument can no longer be constructed. One of these modifications is to remove elements from the knowledge base, such that at least one of the premises required to construct the argument are no longer present. In the same way that removing beliefs from a belief set can have an impact on other beliefs, removing elements from the knowledge base of an argumentation system can have an impact on other arguments, aside from the one that is actively being removed.

This impact, however, is not solely structural when using the ASPIC$^+$ framework. Being built on Dung's abstract theory, the framework evaluates the acceptability of arguments using various sceptical and credulous semantics. In broad terms, an argument is acceptable if it is not defeated by other arguments and an argument is not acceptable if it is. Arguments can defend other arguments by defeating defeaters (for instance, an argument A defends an argument C if A defeats B, which in turn defeats C).

Thus, we must consider at least three effects when removing premises from a knowledge base in order to remove an argument from an argumentation system — **Structural**: those previously acceptable arguments that can no longer be

constructed in the system; **acceptability loss**: those arguments that remain in the system, but have become unacceptable; and **acceptability gain**: those arguments that remain in the system and have gained acceptability. It is possible to formally define these effects, and we do so now in the form of three functions.

Our first function, the *argument drop function*, considers the structural effects on an argumentation system of removing a set of propositions, $D \subseteq \mathcal{K}$:

Definition 1. *The* ***argument drop function*** Δ_A *of* $D \subseteq \mathcal{K}$*:*

$\Delta_A \colon 2^{\mathcal{K}} \to 2^{Args}$,
$\Delta_A(D) = \{A \mid A \in \bigcup E(AT_{\mathcal{AS}}), A \notin \mathcal{AS}\backslash D\}$

Our next two possible changes relate to argument acceptability, with currently acceptable arguments losing their acceptability (but remaining in $\mathcal{AS}$ as defeated arguments) and currently unacceptable arguments gaining acceptability.

We define two functions to capture these changes; first, the *acceptability drop function*, which identifies all acceptable arguments in $\mathcal{AS}$ that, while still capable of being constructed in $\mathcal{AS}\backslash D$, are no longer acceptable:

Definition 2. *The* ***acceptability drop function****,* Δ_S *of* $D \subseteq \mathcal{K}$*:*

$\Delta_S \colon 2^{\mathcal{K}} \to 2^{Args}$,
$\Delta_S(D) = \{A \mid A \in \bigcup E(AT_{\mathcal{AS}}), A \notin \bigcup E(\mathcal{AS}\backslash D), A \in \mathcal{AS}\backslash D\}$

Secondly, the *acceptability gain function*, which identifies those arguments that are not acceptable in $\mathcal{AS}$, but are acceptable in $\mathcal{AS}\backslash D$:

Definition 3. *The* ***acceptability gain function*** Λ_S *of* $D \subseteq \mathcal{K}$*:*

$\Lambda_S \colon 2^{\mathcal{K}} \to 2^{Args}$,
$\Lambda_S(D) = \{A \mid A \notin \bigcup E(AT_{\mathcal{AS}}), A \in \bigcup E(\mathcal{AS}\backslash D)\}$

There is no "argument gain" function, because we assume an open world, and thus do not consider it possible for an argumentation system to gain arguments when removing an argument. We are already considering all arguments (acceptable or otherwise) and thus the removal of an argument cannot cause new arguments to be constructed (but can influence acceptability, as captured by the acceptability drop and gain functions).

The outputs of these three functions can now be used in realising an entrenchment ordering over $2^{\mathcal{K}}$. The different functions are measuring different effects of a change and to simply combine them would be to remove this context. We therefore keep the components separate by representing them as a vector, Υ, with Υ' being a numeric vector, with the sizes of the functions as its components:

$$\Upsilon(D) = \begin{pmatrix} \Delta_A(D) \\ \Delta_S(D) \\ \Lambda_S(D) \end{pmatrix} \qquad \Upsilon'(D) = \begin{pmatrix} \mid \Delta_A(D) \mid \\ \mid \Delta_S(D) \mid \\ \mid \Lambda_S(D) \mid \end{pmatrix}$$

We arrive at an entrenchment ordering over $2^{\mathcal{K}}$ by considering the size of Υ', computed using the standard formula for the length of a vector (the square root of the sum of the squares of the components). If for some $D_1 \subseteq \mathcal{K}$ and $D_2 \subseteq \mathcal{K}$, $\mid \Upsilon'(D_1) \mid < \mid \Upsilon'(D_2) \mid$, then we have an entrenchment ordering, $<_e$ where $D_1 <_e D_2$ (that is, the set D_2 is *more* entrenched than the set D_1).

4 Example

Consider an argumentation system $\mathcal{AS}$ with knowledge base $\mathcal{K} = \{p, q, t, v, x\}$ such that $t < q$ and $v < s$; defeasible rule set $\mathcal{R}_d = \{p, q \Rightarrow r; p \Rightarrow s; t \Rightarrow u; v \Rightarrow w\}$; and contrariness relations $q \in \overline{t}$, $s \in \overline{v}$ and $w \in \overline{x}$.

In addition to atomic arguments on the basis of $\mathcal{K}$, the following arguments can be constructed in $\mathcal{AS}$: $\langle \{p, q\}; p, q \Rightarrow r; r \rangle$, $\langle \{p\}; p \Rightarrow s; s \rangle$, $\langle \{t\}; t \Rightarrow u; u \rangle$, $\langle \{v\}; v \Rightarrow w; w \rangle$, and there exists only one complete extension in $AT_{\mathcal{AS}}$: $\{p, q, r, s, x\}$.

Assume that we must remove the argument for r. This can be done by removing one of two premises: p or q. Consider the outputs of the functions for each premise:

	Δ_A	Δ_S	Λ_S
p	$\{r, s\}$	$\{x\}$	$\{v, w\}$
q	$\{r\}$	$\{\}$	$\{t, u\}$

These yield the following vectors for p and q:

$$\Upsilon(\{p\}) = \begin{pmatrix} \{r, s\} \\ \{x\} \\ \{v, w\} \end{pmatrix} \qquad \Upsilon'(\{p\}) = \begin{pmatrix} 2 \\ 1 \\ 2 \end{pmatrix}$$

$$\Upsilon(\{q\}) = \begin{pmatrix} \{r\} \\ \{\} \\ \{t, u\} \end{pmatrix} \qquad \Upsilon'(\{q\}) = \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix}$$

By using the sizes of $\Upsilon'(\{p\})$ and $\Upsilon'(\{q\})$, we can determine the entrenchment ordering:

$$\mid \Upsilon'(\{p\}) \mid = \sqrt{2^2 + 1^2 + 2^2} = \sqrt{9}$$

$$\mid \Upsilon'(\{q\}) \mid = \sqrt{1^2 + 0^2 + 2^2} = \sqrt{3}$$

Since $\mid \Upsilon'(\{q\}) \mid < \mid \Upsilon'(\{p\}) \mid$, our entrenchment ordering is $\{q\} <_e \{p\}$; that is, the agent, when using structural and semantic considerations, would choose to remove q instead of p in order to remove the argument for r from $\mathcal{AS}$.

5 Conclusions & future work

We have in this paper explored the concept of minimal change when removing an argument from a system of structured argumentation. We identified that an argument can be removed by removing one or more of its premises, which in turn will have an effect on other arguments.

Other arguments can be affected in one of three ways: through their removal from the system (thanks to sharing premises with the originally removed argument); through losing acceptability (but remaining constructable in the system); or gaining acceptability (thanks to a defeater either being removed, or losing acceptability).

The work presented here is an initial step towards appreciating the effects of removing an argument from an argumentation system, and forms only a small part of a larger study into the connection between belief revision and argumentation. In future work, we aim to further refine our notion of "minimal change" by incorporating preferences between arguments, and exploring the role of acceptability semantics. In terms of preferences, we currently consider all arguments identified by the drop and gain functions to be of equal weight. However, ASPIC$^+$ incorporates a preference ordering over arguments, which intuitively should influence an agent's choice when deciding what argument to sacrifice in a revision process. Acceptability semantics are divided into two broad groups: sceptical and credulous. An argument that is sceptically accepted has gone through a more rigorous process in order to determine its acceptability, and thus could be considered more important to an agent than argument that is only credulously accepted.

Beyond measures of minimal change, it is also our intention to develop a set of postulates that describe valid expansions, contractions and revisions of an argumentation system, similar in principle to the AGM postulates of [1]. We envisage these postulates to capture not only the concepts described by the AGM postulates, but also features that are unique to systems of structured argumentation.

Acknowledgements

This work was supported by the Engineering and Physical Sciences Research Council (EPSRC) of the UK government under grant number EP/G060347/1.

References

[1] C.E. Alchourrón, P. Gärdenfors, and D. Makinson. On the logic of theory change: Partial meet contraction and revision functions. *The Journal of Symbolic Logic*, 50:2:510–530, 1985.

[2] L. Amgoud, L. Bodenstaff, M. Caminada, P. McBurney, S. Parsons, H. Prakken, J. van Veenen, and G.A.W. Vreeswijk. Final review and report on formal argumentation system. Deliverable D2.6, ASPIC IST-FP6-002307, 2006.

[3] P. M. Dung. On the acceptability of arguments and its fundemental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial Intelligence*, 77:321–357, 1995.
[4] P. Gärdenfors. *Knowledge in Flux*. MIT Press, 1988.
[5] P. Gärdenfors. Belief revision: An introduction. In Peter Gärdenfors, editor, *Belief Revision*. Cambridge University Press, 1992.
[6] M.O. Moguillansky, N.D. Rotstein, M.A. Falappa, A.J. Garcia, and G.R. Simari. Argument theory change through defeater activation. In *Proceedings of the 3rd International Conference on Computational Models of Argument (COMMA 2010)*, 2010.
[7] H. Prakken. An abstract framework for argumentation with structured arguments. *Argument and Computation*, 1:2:93–124, 2010.
[8] N.D. Rotstein, M.O. Moguillansky, M.A. Falappa, A.J. Garcia, and G.R. Simari. Argument theory change: Revision upon warrant. In *Proceedings of the 2nd International Conference on Computational Models of Argument (COMMA 2008)*, 2008.
[9] N.D. Rotstein, M.O. Moguillansky, A.J. Garcia, and G.R. Simari. A dynamic argumentation framework. In *Proceedings of the third international conference on Computational Models of Argument (COMMA 2010)*, 2010.