Equilibria in Social Belief Removal

Richard Booth
Mahasarakham University
Faculty of Informatics
Mahasarakham 44150, Thailand
richard.b@msu.ac.th

Thomas Meyer
Meraka Institute
CSIR, Pretoria
South Africa
tommie.meyer@meraka.org.za

Abstract

In studies of multi-agent interaction, especially in game theory, the notion of *equilibrium* often plays a prominent role. A typical scenario for the *belief merging* problem is one in which several agents pool their beliefs together to form a consistent "group" picture of the world. The aim of this paper is to define and study new notions of equilibria in belief merging. To do so, we assume the agents arrive at consistency via the use of a *social belief removal* function, in which each agent, using his own *individual* removal function, removes some belief from his stock of beliefs. We examine several notions of equilibria in this setting, assuming a general framework for individual belief removal due to Booth et al. We look at their inter-relations as well as prove their existence or otherwise. We also show how our equilibria can be seen as a generalisation of the idea of taking maximal consistent subsets of agents.

Introduction

The problem of multi-agent belief merging has received a lot of attention in KR in recent years (Konieczny and Gregoire 2006; Konieczny and Pino Pérez 2002; Booth 2006). The problem occurs when several agents each have their own beliefs, and want to combine or pool their beliefs into a consistent "group" picture of the world. A problem arises when two or more agents have conflicting beliefs. Then such conflicts need to be smoothed out. In studies of multi-agent interaction the notion of *equilibrium* often plays a prominent role (most famously in (Nash 1950)). It would therefore seem natural to investigate such notions in belief merging. The purpose of this paper is to define and study some possible notions of equilibria in a belief merging setting.

To enable a clear formulation of such notions, we will employ the approach to merging advocated in (Booth 2006) and inspired by the contraction+expansion approach to belief revision (Gärdenfors 1988; Levi 1991), in which the merging operation is explicitly broken down into two sub-operations. In the first stage, the agents each modify their own beliefs in such a way as to make them jointly consistent. This is called *social contraction* in (Booth 2006). In the second, trivial, stage, the beliefs thus obtained are conjoined. In this approach, the crucial question becomes "how do the agents modify their beliefs in the first stage?" In this paper we assume agents do so by *removing* some sentence from their stock of beliefs. More precisely we associate to each agent i its very own *individual* removal function $\divideontimes_i$ which computes the result of removing any given sentence. A *social belief removal* function is then a function which, given a profile of individual removal functions as input, returns a (consistent) profile consisting of the results of each agent's removal. The central question studied in this paper is "*when can the outcome of a social removal function be said to be in equilibrium?*".

How can we express the idea of equilibrium in social removal? As our starting point we would like to propose the following general principle for multi-agent interaction:

Principle of Equilibrium
Each agent simultaneously makes the appropriate *response to what all the other agents do.*

It remains to formalise what "appropriate" means. In the theory of *strategic games* (see, e.g., (Osborne and Rubinstein 1994) as well as the section "Entrenchment equilibria" of the present paper) agents are assumed to have their own preferences over the set of all outcomes. Then a *Nash equilibrium* (Nash 1950) is a profile consisting of each agent's selected action, in which no agent can achieve a more preferred outcome by changing his action, given the actions of the other agents are held fixed. Hence in this setting "appropriate" may be equated with "best" in a precise sense. We will see that the framework of social belief removal offers up new and interesting ways of formalising what "appropriate" might mean.

Of course the explicit introduction of individuals' removal functions raises the question of what kind of belief removal function we should assume is being used. Do agents use AGM contraction (Alchourrón, Gärdenfors, and Makinson 1985), or severe withdrawal (Rott and Pagnucco 1999), or perhaps a belief liberation function (Booth et al. 2005)? Luckily there exists a general family, called *basic removal* (Booth et al. 2004) which contains *all* these families and more besides. Thus we find it convenient to use this family as a basis.

The plan of the paper is as follows. In the next section we set up the framework of social removal functions. Then we

Copyright © 2008, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

focus on the agents' individual removal functions, reviewing some results about basic removal functions and giving some concrete examples of such functions. Next, we introduce our first equilibrium notion, that of a *removal equilibrium*, and examine its compatibility with some plausible minimal change properties, before proving the existence of such equilibria for arbitrary basic removal profiles. We also briefly look at the notion of *perfect removal equilibria*. After this we move on to *entrenchment equilibria*, which can be thought of as Nash equilibria of the strategic game where agent preferences over outcomes are derived from their entrenchment orderings, and examine their relationship with removal equilibria. We also suggest a possible refinement of this idea, the *strong entrenchment equilibrium*. Next we show how our equilibria can be thought of as generalising the idea of taking maximal consistent subsets of agent, before looking at some related work by Meyer, Zhang et al. on logical models of negotiation. We finish with a concluding section.

Preliminaries: We work in a finitely-generated propositional language L. Classical logical consequence and logical equivalence are denoted by $\vdash$ and $\equiv$ respectively. W denotes the set of possible worlds/interpretations for L. Given $\theta \in L$, we denote the set of worlds in which θ is true by $[\theta]$. The set of non-tautologous sentences in L is denoted by L_*. We will usually talk of belief *sets*, but assume a belief set is always represented by a single sentence standing for its set of logical consequences. We assume a set of *agents* $\mathbb{A} = \{1, \ldots, n\}$. A *belief profile* is any n-tuple of belief sets. Given two belief profiles we shall write $(\phi_i)_{i\in\mathbb{A}} \equiv (\phi'_i)_{i\in\mathbb{A}}$ iff $\phi_i \equiv \phi'_i$ for all i, and write $(\phi_i)_{i\in\mathbb{A}} \equiv_\wedge (\phi'_i)_{i\in\mathbb{A}}$ iff $\bigwedge_{i\in\mathbb{A}} \phi_i \equiv \bigwedge_{i\in\mathbb{A}} \phi'_i$. Clearly we have $\equiv \subseteq \equiv_\wedge$ for belief profiles. We say the belief profile is consistent iff the conjunction of its elements is consistent.

Social belief removal

As we said above, we assume each agent $i \in \mathbb{A}$ comes equipped with its own *removal function* $\divideontimes_i$, which tells it how to remove any given sentence from its belief set. In this paper we view $\divideontimes_i$ as a *unary function* on the set L_* of non-tautologous sentences, i.e., agents are never required to remove $\top$. The result of removing $\lambda \in L_*$ from i's belief set is denoted by $\divideontimes_i(\lambda)$. We assume i's *initial* belief set can always be recaptured from $\divideontimes_i$ alone by just removing the contradiction, i.e., i's initial belief set is $\divideontimes_i(\bot)$. We call any n-tuple $(\divideontimes_i)_{i\in\mathbb{A}}$ of removal functions a *removal profile*.

Definition 1 *A* social removal function $\mathbf{F}$ *(relative to $\mathbb{A}$) is any function which takes as input any removal profile $(\divideontimes_i)_{i\in\mathbb{A}}$ and outputs a consistent belief profile $\mathbf{F}((\divideontimes_i)_{i\in\mathbb{A}}) = (\phi_i)_{i\in\mathbb{A}}$ such that, for each $i \in \mathbb{A}$, there exists $\lambda_i \in L_*$ such that $\phi_i \equiv \divideontimes_i(\lambda_i)$.*

Each social removal function yields a merging operator for removal profiles – we just take the conjunction $\bigwedge_{i\in\mathbb{A}} \phi_i$ of the agents' new belief profile. However in this paper our main interest will be in the profile itself.

The above definition differs from Booth's social contraction in two main ways. First, here we *explicitly* associate from the outset an individual removal function to each i, whereas this was only implicit in (Booth 2006). More importantly, unlike in social contraction, we will allow agents to use removal functions which don't necessarily satisfy the Inclusion property, i.e., removing a sentence *may* lead to new beliefs entering i's belief set. As is argued in (Booth et al. 2005), this situation can arise quite naturally. This motivates the use of the term social *removal* rather than social *contraction*.

What properties might we expect from a social removal function $\mathbf{F}$? Throughout the paper we will mention various postulates for $\mathbf{F}$, but to begin with the following two properties have – on the face of it – a strong appeal from a "minimal change" viewpoint:

(FVac) If $(\divideontimes_i(\bot))_{i\in\mathbb{A}}$ is consistent then $\mathbf{F}((\divideontimes_i)_{i\in\mathbb{A}}) \equiv (\divideontimes_i(\bot))_{i\in\mathbb{A}}$

(FVac$_\wedge$) If $(\divideontimes_i(\bot))_{i\in\mathbb{A}}$ is consistent then $\mathbf{F}((\divideontimes_i)_{i\in\mathbb{A}}) \equiv_\wedge (\divideontimes_i(\bot))_{i\in\mathbb{A}}$

Both these rules deal with the case the initial belief sets of the agents are already jointly consistent. **(FVac)** says that in this case the agents' beliefs should remain unchanged. Although intuitively appealing, we will later have grounds for believing this rule is a touch too strong (specifically in contexts where the agents' individual removal functions might not adhere to the Vacuity rule – see next section). Rule **(FVac$_\wedge$)** is weaker. It requires only that the result should be conjunction-equivalent to the profile of the agents' initial belief sets.

Basic and hyperregular removal

What properties should be assumed of the individual removal functions $\divideontimes_i$? We will assume agents always use *basic* removal.

Definition 2 *A function $\divideontimes : L_* \to L$ is a* basic removal function *iff it satisfies the following rules (Booth et al. 2004):*

($\divideontimes$1) $\divideontimes(\lambda) \nvdash \lambda$

($\divideontimes$2) *If* $\lambda_1 \equiv \lambda_2$ *then* $\divideontimes(\lambda_1) \equiv \divideontimes(\lambda_2)$

($\divideontimes$3) *If* $\divideontimes(\chi \wedge \lambda) \vdash \chi$ *then* $\divideontimes(\chi \wedge \lambda \wedge \psi) \vdash \chi$

($\divideontimes$4) *If* $\divideontimes(\chi \wedge \lambda) \vdash \chi$ *then* $\divideontimes(\chi \wedge \lambda) \vdash \divideontimes(\lambda)$

($\divideontimes$5) $\divideontimes(\chi \wedge \lambda) \vdash \divideontimes(\chi) \vee \divideontimes(\lambda)$

($\divideontimes$6) *If* $\divideontimes(\chi \wedge \lambda) \nvdash \lambda$ *then* $\divideontimes(\lambda) \vdash \divideontimes(\chi \wedge \lambda)$

All these rules are familiar from the literature on belief removal. Rule **($\divideontimes$1)** is the Success postulate which says the sentence to be removed is no longer implied by the new belief set, while **($\divideontimes$2)** is a syntax-irrelevance property. Rule **($\divideontimes$3)** is sometimes known as Conjunctive Trisection (Hansson 1993a; Rott 1992). It says if χ is believed after removing the conjunction $\chi \wedge \lambda$, then it should also be believed when removing the longer conjunction $\chi \wedge \lambda \wedge \psi$. Rule **($\divideontimes$4)** is closely-related to the rule Cautious Monotony from the area of non-monotonic reasoning (Kraus, Lehmann, and Magidor 1991), while **($\divideontimes$5)** and **($\divideontimes$6)** are the two AGM supplementary postulates for contraction (Alchourrón, Gärdenfors, and Makinson 1985).

Note the non-appearance in this list of the AGM contraction postulates Vacuity ($\divideontimes(\bot) \nvdash \lambda$ implies $\divideontimes(\lambda) \equiv \divideontimes(\bot)$),

Inclusion ($\divideontimes(\bot) \vdash \divideontimes(\lambda)$) and Recovery ($\divideontimes(\lambda) \wedge \lambda \vdash \divideontimes(\bot)$), none of which are valid in general for basic removal. Inclusion has been questioned as a general requirement for removal in (Booth et al. 2005), while Recovery has long been noted as controversial (see, e.g., (Hansson 1991)). Vacuity is a little harder to argue against. It says if the sentence to be removed is not in the intial belief set, then the belief set should remain unchanged. Nevertheless we feel there *are* plausible removal scenarios in which it may fail, one of which will be described in our examples of basic removals below when we introduce the subclass of *prioritised* removal functions. For basic removals Inclusion actually implies Vacuity (Booth et al. 2004).

Note: The postulates are the same ones as in (Booth et al. 2004), but their appearance is changed to take into account the fact we take $\divideontimes$ to be a unary operator which returns a sentence (rather than a logically-closed set of sentences). We also leave out one rule from the list in (Booth et al. 2004), which in our reformulation corresponds to "$\divideontimes(\bot) \wedge \neg\lambda \vdash \divideontimes(\lambda)$". This rule turns out to be redundant, being derivable mainly from ($\divideontimes$**3)**.

As well as the above postulates, Booth et al. (2004) also gave a semantic account of basic removal. A *context* is any pair $\mathcal{C} = (\leq, \preceq)$ of binary relations over W such that *(i)* $\leq$ is a total preorder, i.e., transitive and connected, and *(ii)* $\preceq$ is a reflexive sub-relation of $\leq$. From any such $\mathcal{C}$ we may define a removal operator $\divideontimes_{\mathcal{C}}$ by setting

$$[\divideontimes_{\mathcal{C}}(\lambda)] = \{w \in W \mid w \preceq w' \text{ for some } w' \in \min_{\leq}([\neg\lambda])\}.$$

That is, the set of worlds following removal of λ is determined by first locating the $\leq$-minimal worlds in $[\neg\lambda]$, and then taking along with these all worlds which are less than them according to $\preceq$. We call $\divideontimes_{\mathcal{C}}$ the removal function *generated by* $\mathcal{C}$. Booth et al. (2004) showed $\divideontimes_{\mathcal{C}}$ is a basic removal function and that in fact *every* basic removal function is generated from a unique context. For another, closely-related, family of belief removal functions see (Cantwell 2003).

Hyperregular removal

In this paper, another property which we will find useful, especially for technical reasons, is *Hyperregularity* (Hansson 1993b):

$$\text{If } \divideontimes(\lambda \wedge \chi) \nvdash \lambda \text{ then } \divideontimes(\lambda \wedge \chi) \equiv \divideontimes(\lambda).$$

This rule says if the removal of $\lambda \wedge \chi$ excludes λ then removing $\lambda \wedge \chi$ is the same as removing just λ. This property is very strong. Not only does it imply Vacuity, but in the presence of ($\divideontimes$**1)** and ($\divideontimes$**2)** it implies ($\divideontimes$**3)**-($\divideontimes$**6)**. It is probably *too* strong to be required in general. Indeed given ($\divideontimes$**1)** and ($\divideontimes$**2)** it can be shown to imply the following "Decomposition" property of removal, which has been noted as overly strong in (Gärdenfors 1988, p66):

$$\divideontimes(\lambda \wedge \chi) \equiv \divideontimes(\lambda) \text{ or } \divideontimes(\lambda \wedge \chi) \equiv \divideontimes(\chi).$$

Despite this it is nevertheless still satisifed by several interesting sub-classes of basic removal (see the examples below), and when proving results we will sometimes find it a useful stepping-stone towards the more general basic removal. In terms of contexts, it corresponds to requiring the following condition on $(\leq, \preceq)$, for all $w_1, w_2, w_3 \in W$:

($\mathcal{C}$-hyp) If $w_1 \preceq w_2$ and $w_2 \sim w_3$ then $w_1 \preceq w_3$

(where $\sim$ is the symmetric closure of $\leq$). In other words, whether or not $w_1 \preceq w_2$ depends only on the $\leq$-rank of w_2.

Definition 3 *A* hyperregular removal function *is any basic removal function satisfying Hyperregularity.*

In (Booth et al. 2004) it was shown that hyperregular removal functions correspond precisely to the class of *linear liberation* operators from (Booth et al. 2005).

Some examples of basic removal functions

We now give three concrete families of operators, all of which come under the umbrella of basic removal. These families will be useful when we come to describing examples of equilibria.

(i). **Prioritised removal** Let $\langle \Sigma, \sqsubseteq \rangle$ be any finite set of *consistent* sentences Σ, totally preordered by a relation $\sqsubseteq$ over Σ. Intuitively the different sentences in Σ correspond to different possible *extensions*, prioritised by $\sqsubseteq$ (and with sentences lower down in the ordering given higher priority). Given such a set, for any $\lambda \in L_*$ let $\Sigma(\lambda) = \{\gamma \in \Sigma \mid \gamma \nvdash \lambda\}$. Then we define $\divideontimes_{\langle \Sigma, \sqsubseteq \rangle}$ from $\langle \Sigma, \sqsubseteq \rangle$ by setting:

$$\divideontimes_{\langle \Sigma, \sqsubseteq \rangle}(\lambda) = \begin{cases} \bigvee \min_{\sqsubseteq} \Sigma(\lambda) & \text{if } \bigvee \Sigma \nvdash \lambda \\ \top & \text{otherwise.} \end{cases}$$

In other words, after removing λ, the new belief set is just the disjunction of all the $\sqsubseteq$-minimal elements in Σ which do not entail λ. In case there is no sentence in Σ which fails to imply λ, then the result is just $\top$. We will call any removal function definable in this way a *prioritised removal* function. A similar family of removal has also been studied in (Bochman 2001).

One can easily check that $\divideontimes_{\langle \Sigma, \sqsubseteq \rangle}$ satisfies ($\divideontimes$**1)**-($\divideontimes$**6)** and so forms a basic removal function. Note however that $\divideontimes_{\langle \Sigma, \sqsubseteq \rangle}$ will fail to satisfy Vacuity (hence also Hyperregularity) in general. For example suppose $\Sigma = \{p, \neg p\}$ but $\sqsubseteq$ is the "flat" ordering on Σ which ranks both sentences equally. This would correspond to a situation in which an agent has equally good reasons to believe p and $\neg p$. The belief set corresponding to this is then $\divideontimes_{\langle \Sigma, \sqsubseteq \rangle}(\bot) = p \vee \neg p$, i.e., since the agent cannot choose between p and $\neg p$, he commits to neither. But $\divideontimes_{\langle \Sigma, \sqsubseteq \rangle}(p) = \neg p$. That is, the direction to remove p tips the balance in favour of $\neg p$, and the agent thus comes to believe $\neg p$, even though p was not in the initial belief set. We take this plausible removal scenario as indication that the Vacuity rule may be too strong in general.

(ii). **Severe withdrawal** (Rott and Pagnucco 1999). A *severe withdrawal* function may be represented by a *logical chain* $\rho = \beta_1 \vdash \beta_2 \vdash \cdots \vdash \beta_m$, with $\divideontimes_\rho(\lambda) = \beta_i$, where i is minimal such that $\beta_i \nvdash \lambda$ (equals $\top$ if no such i exists). Severe withdrawal functions always satisfy Inclusion and Hyperregularity. It is easy to see they form a special case of prioritised removal. Severe withdrawal functions also have

a simple representation in terms of their generating contexts $(\leq, \preceq)$. They are just those basic removals for which $\leq = \preceq$.

(iii). σ**-liberation** (Booth et al. 2005). *σ-liberation* functions again use a sequence of sentences $\sigma = (\alpha_1, \ldots, \alpha_s)$. Given such σ and $\lambda \in L_*$, define a sequence of sentences $f_i(\sigma, \lambda)$ inductively on i by setting $f_0(\sigma, \lambda) = \top$, and then for $i > 0$,

$$f_i(\sigma, \lambda) = \begin{cases} f_{i-1}(\sigma, \lambda) \wedge \alpha_i & \text{if } f_{i-1}(\sigma, \lambda) \wedge \alpha_i \not\vdash \lambda \\ f_{i-1}(\sigma, \lambda) & \text{otherwise.} \end{cases}$$

In other words, $f_s(\sigma, \alpha)$ is obtained by starting with $\top$, and then working through σ from left to right, adding each sentence provided doing so does not lead to the inference of λ. (In (Booth et al. 2005) the direction was right-to-left, but this difference is inessential.) Then $\divideontimes_\sigma(\lambda) = f_s(\sigma, \lambda)$. (This is very closely-related to the "linear base-revision" of (Nebel 1994).) σ-liberation functions do not satisfy Inclusion in general, but they do satisfy Hyperregularity (and hence also Vacuity). In terms of their generating contexts, σ-liberation functions correspond to those contexts $(\leq, \preceq)$ which satisfy the Hyperregularity condition **($\mathcal{C}$-hyp)** and for which $\preceq$ is transitive.

The three families described above are inter-related as follows:

severe withdrawal $\subset$ σ-liberation $\subset$ prioritised removal.

The inclusions are strict. In addition to these three, Booth et al. (2004) showed basic removal includes many other well-known families of removal functions, including systematic withdrawal (Meyer et al. 2002), AGM contraction and even AGM *revision*.[1]

In the rest of the paper we shall assume the domain of a social removal function is the set of all n-tuples of basic removal functions.

Removal equilibria

When is the outcome $(\phi_i)_{i \in \mathbb{A}}$ of an operation of social removal an *equilibrium* point? Our first idea is the following.

Definition 4 *$(\phi_i)_{i \in \mathbb{A}}$ is a* removal equilibrium *for $(\divideontimes_i)_{i \in \mathbb{A}}$ iff it is consistent and, for each $i \in \mathbb{A}$, $\phi_i \equiv \divideontimes_i(\neg \bigwedge_{j \neq i} \phi_j)$.*

This definition is a direct formulation of the idea that each agent removes precisely the "right" sentence to be consistent with every other agent. As such this seems like a good candidate for a first formalisation of the word "appropriate" in our Principle of Equilibrium from the introduction.

Example 1 *Assume $\mathbb{A} = \{1, 2\}$ and suppose both agents use severe withdrawal to remove beliefs. Let $\divideontimes_1$ and $\divideontimes_2$ be specified by the logical chains $(p \wedge q) \vdash q$ and $(\neg p \wedge \neg q) \vdash (\neg p \vee \neg q)$ resp. Then there are three possible removal equilibria for the profile $(\divideontimes_1, \divideontimes_2)$: (1) $(p \wedge q, \top)$, corresponding to a case where 1 removes nothing and 2 removes everything, (2) $(\top, \neg p \wedge \neg q)$, corresponding to the opposite case, and (3) $(q, \neg p \vee \neg q)$, corresponding to the case where both agents give up* something, *but not everything.*

[1]The fact that basic removal also covers AGM revision is what motivated our choice of the contraction-revision "hybrid" symbol $\divideontimes$ to denote removal functions.

We might be interested in requiring the following property for social removal functions:

(FREq) $\mathbf{F}((\divideontimes_i)_{i \in \mathbb{A}})$ is a removal equilibrium for $(\divideontimes_i)_{i \in \mathbb{A}}$.

Is **(FREq)** even consistent? In other words, do removal equilibria always exist for *any* profile of basic removal functions? We shall shortly answer this question in the affirmative. But before that we examine such equilibria in the special case when $(\divideontimes_i(\bot))_{i \in \mathbb{A}}$ is consistent, and examine the compatibility of **(FREq)** with **(FVac)** and **(FVac$_\wedge$)**. First, the following example shows **(FREq)** is *not* compatible with **(FVac)**.

Example 2 *Again suppose $\mathbb{A} = \{1, 2\}$. Suppose agent 1 uses the prioritised removal function $\divideontimes_{\langle \Sigma, \sqsubseteq \rangle}$ where $\Sigma = \{p, \neg p\}$ and $\sqsubseteq$ is the flat priority ordering, and suppose agent 2 uses the severe withdrawal function specified by the single element logical chain (p). We have $\divideontimes_1(\bot) \equiv \top$ and $\divideontimes_2(\bot) = p$. Then $\divideontimes_1(\bot) \wedge \divideontimes_2(\bot)$ is equivalent to p and so is clearly consistent, but $(\divideontimes_1(\bot), \divideontimes_2(\bot))$ is not a removal equilibrium. This is because, while we do have $\divideontimes_2(\neg\top) \equiv p$, we have $\divideontimes_1(\neg p) \equiv p \not\equiv \top$.*

Thus for general basic removal profiles, we cannot require **both (FREq)** and **(FVac)**. At first glance it might be thought **(FVac)** is unquestionable, and so it is **(FREq)** which must be given up. However we believe that as soon as one takes the step – as we do – to relax Vacuity for *individual* removal $\divideontimes$, then **(FVac)** itself becomes less "untouchable". Thus we believe this incompatibility with **(FVac)** should not by itself be taken as reason to reject **(FREq)**. Furthermore the next result (which is actually a consequence of Prop. 12 below) shows **(FREq)** *is* compatible with **(FVac$_\wedge$)**.

Proposition 1 *If $(\divideontimes_i(\bot))_{i \in \mathbb{A}}$ is consistent then there exists a removal equilibrium $(\phi_i)_{i \in \mathbb{A}}$ for $(\divideontimes_i)_{i \in \mathbb{A}}$ such that $(\phi_i)_{i \in \mathbb{A}} \equiv_\wedge (\divideontimes_i(\bot))_{i \in \mathbb{A}}$.*

In Example 2 we do indeed have a removal equilibrium which is conjunction-equivalent to $(\divideontimes_1(\bot), \divideontimes_2(\bot))$, namely (p, p).

Note that in Example 2, agent 1 uses a removal function which does not satisfy Vacuity. The next result says that if we *do* insist on Vacuity for individual removal functions, then we do achieve compatibility with **(FVac)**.

Proposition 2 *Suppose each $\divideontimes_i$ satisfies Vacuity, and suppose $(\divideontimes_i(\bot))_{i \in \mathbb{A}}$ is consistent. Then $(\divideontimes_i(\bot))_{i \in \mathbb{A}}$ is a removal equilibrium for $((\divideontimes_i)_{i \in \mathbb{A}})$.*

However, even if the $\divideontimes_i$ satisfy Vacuity, this might not be the *only* removal equilibrium. In other words even in this restricted domain case, **(FREq)** is not enough by itself to imply **(FVac)** or even **(FVac$_\wedge$)**.

Example 3 *Let $\divideontimes$ be the σ-liberation function determined by the sequence $(p, \neg p)$. Then the belief set associated to $\divideontimes$ is $\divideontimes(\bot) = p$. Now suppose we have n agents, all using this* same *removal function $\divideontimes$. Then for the resulting removal profile there are two removal equilibria. As well as the expected $(p)_{i \in \mathbb{A}}$ we also get $(\neg p)_{i \in \mathbb{A}}$!*

It might seem bizarre that $(\neg p)_{i\in\mathbb{A}}$ should be recognised as an equilibrium in this example. Why should the agents all jump across to $\neg p$ when they can just as well stay with the comfort of p? In fact the situation is analogous to that with Nash equilibrium itself. We shall expand on this point later after we introduce the notion of entrenchment equilibria.

By restricting the domain of $\mathbf{F}$ further, we *do* force a unique removal equilibrium in the case when the initial belief sets are jointly consistent.

Proposition 3 *Suppose each $*_i$ satisfies Inclusion (and hence also Vacuity). Then if $(*_i(\bot))_{i\in\mathbb{A}}$ is consistent then it is the* only *removal equilibrium for $(*_i)_{i\in\mathbb{A}}$.*

Existence of removal equilibria

In this section we prove that removal equilibria are guaranteed to exist when the agents use basic removal functions to remove beliefs. First we concentrate on the case when all agents use hyperregular removal, providing two concrete social removal operators which satisfy **(FREq)**. We will build on this case to prove existence in the general basic removal case.

The hyperregular case: First method

Our first social removal function $\mathbf{F}_1$ requires the upfront specification of a linear order on $\mathbb{A}$. Without loss we take this order here to be just the numerical one on $\mathbb{A} = \{1, 2, \ldots, n\}$. Given a removal profile $(*_i)_{i\in\mathbb{A}}$, we define $\mathbf{F}_1((*_i)_{i\in\mathbb{A}}) = (\phi_i)_{i\in\mathbb{A}}$ inductively by setting

$$\phi_i = *_i(\neg \bigwedge_{j<i} \phi_j).$$

In other words, ϕ_1 is just taken to be agent 1's initial belief set $*_1(\bot)$, and then each agent takes his turn to remove the negation of the conjunction of the belief sets of all those agents whose turn has already passed. By an easy induction on i, and using the fact each $*_i$ satisfies **(*1)**, we know $\neg \bigwedge_{j<i} \phi_j \in L_*$ and so $*_i(\neg \bigwedge_{j<i} \phi_j)$ is well-defined. In particular we know from **(*1)** that $\phi_n = *_n(\neg \bigwedge_{j<n} \phi_j) \nvdash \neg \bigwedge_{j<n} \phi_j$ and so $(\phi_i)_{i\in\mathbb{A}}$ is consistent.

Proposition 4 *If all the $*_i$ satisfy Hyperregularity then $\mathbf{F}_1$ returns a removal equilibrium for $(*_i)_{i\in\mathbb{A}}$.*

Proof. From the above remarks we know $(\phi_i)_{i\in\mathbb{A}}$ is consistent. It remains to show, for all i, $\phi_i \equiv *_i(\neg \bigwedge_{j\neq i} \phi_j)$. We know $\phi_i = *_i(\neg \bigwedge_{j<i} \phi_j)$. Since $\neg \bigwedge_{j<i} \phi_j \vdash \neg \bigwedge_{j\neq i} \phi_j$ this is equivalent to

$$\phi_i \equiv *_i((\neg \bigwedge_{j<i} \phi_j) \wedge (\neg \bigwedge_{j\neq i} \phi_j)).$$

Since $(\phi_i)_{i\in\mathbb{A}}$ is consistent we have $\phi_i \nvdash \neg \bigwedge_{j\neq i} \phi_j$ and so we may apply Hyperregularity to deduce the required conclusion. □

$\mathbf{F}_1$ might not return a removal equilibrium for general basic removal profiles. This can be seen on Example 2, where running the above procedure returns the non-equilibrium $(\top, p)$.

What other properties does $\mathbf{F}_1$ satisfy? Well to begin, it can be shown to satisfy **(FVac)** (in the hyperregular case). Also, let's say two removal functions $*$ and $*'$ are *revision-equivalent* iff $*(\lambda) \wedge \neg\lambda \equiv *'(\lambda) \wedge \neg\lambda$ for all $\lambda \in L_*$. (i.e., the revision functions defined from them via the Levi Identity (Levi 1991) are the same). Then we have:

Proposition 5 *$\mathbf{F}_1$ satisfies the following rule for social removal functions:*

(FRev$_\wedge$) *If $*_i$ and $*'_i$ are revision-equivalent for each $i \in \mathbb{A}$ then $\mathbf{F}((*_i)_{i\in\mathbb{A}}) \equiv_\wedge \mathbf{F}((*'_i)_{i\in\mathbb{A}})$.*

In fact $\mathbf{F}_1$ satisfies this property even in the general basic removal case. Letting $\mathbf{F}_1((*_i)_{i\in\mathbb{A}}) = (\phi_i)_{i\in\mathbb{A}}$ and $\mathbf{F}_1((*'_i)_{i\in\mathbb{A}}) = (\phi'_i)_{i\in\mathbb{A}}$, the proof proceeds by induction on i that $\bigwedge_{j\leq i} \phi_j \equiv \bigwedge_{j\leq i} \phi'_j$. This result implies that if we are only interested in the result of *merging*, we could just focus on revision functions only.

One questionable property of $\mathbf{F}_1$ is that we *always* get $\phi_1 = *_1(\bot)$ for any input removal profile. Thus agent 1 never leaves his initial belief set. He assumes a dictator-like role. Our second construction aims at rectifying this.

The hyperregular case: Second method

Our second construction is just like the first, except now, at the start of the process, agent 1 removes some fixed, possibly consistent, sentence χ (chosen independently of the given removal profile) rather than remove $\bot$ as before. Formally, the function $\mathbf{F}_2$ makes use of an auxilliary function $\mathbf{s}$ which takes as arguments a removal profile $(*_i)_{i\in\mathbb{A}}$ together with a sentence $\chi \in L_*$, and outputs a belief profile $(\eta_i)_{i\in\mathbb{A}}$. The η_i are defined inductively by setting $\eta_1 = *_1(\chi)$, and then for $i > 1$,

$$\eta_i = *_i(\neg \bigwedge_{j<i} \eta_j).$$

Note that if $\chi \equiv \bot$ then this is just $\mathbf{F}_1((*_i)_{i\in\mathbb{A}})$. Is this a removal equilibrium? In fact the result of this operation will be a removal equilibrium for agents $2, \ldots, n$, but not necessarily for agent 1.

Proposition 6 *Assume all $*_i$ satisfy Hyperregularity and let $\mathbf{s}(\chi \mid (*_i)_{i\in\mathbb{A}}) = (\eta_i)_{i\in\mathbb{A}}$. Then for each $i > 1$, $\eta_i \equiv *_i(\neg \bigwedge_{j\neq i} \eta_j)$, but in general $\eta_1 \not\equiv *_1(\neg \bigwedge_{j>1} \eta_j)$.*

In case $\eta_1 \not\equiv *_1(\neg \bigwedge_{j>1} \eta_j)$ we just try again with $\mathbf{s}(\chi \wedge \neg \bigwedge_{j>1} \eta_j \mid (*_i)_{i\in\mathbb{A}})$. Precisely, $\mathbf{F}_2$ is defined via the following iterative procedure:

1. Calculate $\mathbf{s}(\chi \mid (*_i)_{i\in\mathbb{A}}) = (\eta_i)_{i\in\mathbb{A}}$.
2. If $\eta_1 \equiv *_1(\neg \bigwedge_{j>1} \eta_j)$ then STOP and output $\mathbf{F}_2((*_i)_{i\in\mathbb{A}}) = (\eta_i)_{i\in\mathbb{A}}$. Otherwise set $\chi := \chi \wedge \neg \bigwedge_{j>1} \eta_j$ and go to step 1.

In case the termination condition in step 2 is not met, it can be shown $\chi \not\equiv \chi \wedge \neg \bigwedge_{j>1} \eta_j$, so we generate a stricly stronger sentence to input back into $\mathbf{s}(\cdot \mid (*_i)_{i\in\mathbb{A}})$ in step 1. Hence the process continues at most until we input $\bot$. But in this case $\mathbf{s}(\bot \mid (*_i)_{i\in\mathbb{A}}) = \mathbf{F}_1((*_i)_{i\in\mathbb{A}})$ as we have seen. Hence:

Proposition 7 *If all the $\circledast_i$ satisfy Hyperregularity then* $\mathbf{F}_2$ *satisfies (***FREq***).*

For example, if we run this method on Example 3, taking $\chi = p$, we obtain the 2^{nd} equilibrium $\mathbf{F}_2((\circledast)_{i\in\mathbb{A}}) = (\neg p)_{i\in\mathbb{A}}$. Hence we see $\mathbf{F}_2$ does not validate **(FVac$_\wedge$)**. It also does not satisfy **(FRev$_\wedge$)**, since it can be shown the σ-liberation function from Example 3 is revision-equivalent to the severe withdrawal function $\circledast_\rho$ determined by the 1-element chain $\rho = (p)$. But if we again take $\chi = p$ then $\mathbf{F}_2((\circledast_\rho)_{i\in\mathbb{A}}) = (p)_{i\in\mathbb{A}}$.

Note although agent 1 no longer has dictator-like powers in $\mathbf{F}_2$, agent j still *dominates* all agents k for which $2 \leq j < k$, in the sense that if $\mathbf{F}_2((\circledast_i)_{i\mathbb{A}}) = (\phi_i)_{i\in\mathbb{A}}$, we *always* end up with $\phi_j = \circledast_j(\neg\bigwedge_{s<j}\phi_s)$. This means j *never* takes into account the beliefs of $k > j$ when calculating his new beliefs.

A natural question to ask is: is *every* removal equilibrium for $(\circledast_i)_{i\in\mathbb{A}}$ obtainable by the above iterative method for appropriate choices of ordering of agents and starting points χ? The next example shows the answer is generally no.

Example 4 *Suppose three agents, all using severe withdrawal functions specified respectively by the following logical chains:*

$$\begin{array}{ll}\circledast_1: & (p \leftrightarrow \neg q) \vdash (p \vee q),\\ \circledast_2: & \neg q \vdash (p \vee \neg q),\\ \circledast_3: & \neg p \vdash (\neg p \vee q).\end{array}$$

Then the reader may check

$$(\phi_1, \phi_2, \phi_3) = (p \vee q, p \vee \neg q, \neg p \vee q)$$

is a removal equilibrium (giving a merging result of $\phi_1 \wedge \phi_2 \wedge \phi_3 \equiv p \wedge q$). However, note this equilibrium has the special property that for each i, there is no proper subset $X \subset \{j \in \mathbb{A} \mid j \neq i\}$ *such that $\phi_i \equiv \circledast_i(\neg\bigwedge_{j\in X}\phi_j)$. Hence this point cannot be reached using* $\mathbf{F}_2$*, since as we just remarked, there we* always *end up with $\phi_2 \equiv \circledast_2(\neg\phi_1)$.*

In the above example it could be said that at the point $(p \vee q, p \vee \neg q, \neg p \vee q)$ the three agents are all in a state of *perfect tension* with regard to one another. Each agent contributes equally to the equilibrium. We make the following definition:

Definition 5 *Let $(\phi_i)_{i\in\mathbb{A}}$ be a removal equilibrium for $(\circledast_i)_{i\in\mathbb{A}}$. Then it is a* perfect removal equilibrium *iff for each i, there is no* proper subset $X \subset \{j \in \mathbb{A} \mid j \neq i\}$ *such that $\phi_i \equiv \circledast_i(\neg\bigwedge_{j\in X}\phi_j)$.*

The next question is: do perfect removal equilibria always exist for any given removal profile? The answer is no, because according to the definition we may *not* have $\phi_i \equiv \circledast_i(\neg\bigwedge_{j\in\emptyset}\phi_j)$, i.e., we may not have $\phi_i \equiv \circledast_i(\bot)$. However, we may conceive of examples in which, for *every* removal equilibrium there exists at least one agent i for which $\phi_i \equiv \circledast_i(\bot)$. Indeed this will typically happen in the case of *drastic* removal profiles, see the section "Equilibria as maxiconsistent sets" below.

Existence: The general case

We have established that if all agents use hyperregular removal, then removal equilibria are guaranteed to exist. We now extend this fact to the case of arbitrary basic removal profiles. Given an arbitrary $(\circledast_i)_{i\in\mathbb{A}}$, we first convert each $\circledast_i$ to its *hyperregular version* $\circledast_i^h$, and then show that every removal equilibrium for $(\circledast_i^h)_{i\in\mathbb{A}}$ can be *converted* into an equilibrium for the original profile. To do this we go back to the semantic representation of basic removal functions which was mentioned after Defn. 2.

Definition 6 *Let $\circledast$ be a basic removal function and $(\leq, \preceq)$ its generating context. Then the* hyperregular version *of $\circledast$ is the the removal operator $\circledast^h$ generated by the context $(\leq, \preceq^h)$, where $\preceq^h$ is defined by:*

$$w_1 \preceq^h w_2 \textit{ iff } w_1 \preceq w_3 \textit{ for some } w_3 \textit{ s.t. } w_3 \sim w_2.$$

(where $\sim$ is the symmetric closure of $\leq$).

The following are the relevant properties of $\circledast^h$:

Proposition 8
(i). *$\circledast^h$ satisfies Hyperregularity.*
(ii). *For all $\lambda \in L_*$, $\circledast(\lambda) \vdash \circledast^h(\lambda)$.*
(iii). *$\circledast$ and $\circledast^h$ are revision-equivalent.*

Now, suppose we start with arbitrary $(\circledast_i)_{i\in\mathbb{A}}$ and suppose we have found some removal equilibrium $(\phi'_i)_{i\in\mathbb{A}}$ **for the hyperregular versions** $(\circledast_i^h)_{i\in\mathbb{A}}$. Then for each i set

$$\phi_i = \circledast_i(\neg(\bigwedge_{j<i}\phi_j \wedge \bigwedge_{j>i}\phi'_j)).$$

Proposition 9 *$(\phi_i)_{i\in\mathbb{A}}$ is a removal equilibrium for $(\circledast_i)_{i\in\mathbb{A}}$. Furthermore $(\phi_i)_{i\in\mathbb{A}} \equiv_\wedge (\phi'_i)_{i\in\mathbb{A}}$.*

Proof. The proof depends on the following property:

$$\bigwedge_{j\in\mathbb{A}}\phi'_j \vdash \phi_i \vdash \phi'_i \text{ for all } i \in \mathbb{A} \qquad (1)$$

This property is proved by induction on i. For $i = 1$ we have $\phi_1 = \circledast_1(\neg\bigwedge_{j>1}\phi'_j)$ and $\phi'_1 = \circledast_1^h(\neg\bigwedge_{j>1}\phi'_j)$. Hence the first logical implication above reduces to

$$(\bigwedge_{j>1}\phi'_j) \wedge \circledast_1^h(\neg\bigwedge_{j>1}\phi'_j) \vdash \circledast_1(\neg\bigwedge_{j>1}\phi'_j),$$

which holds by Prop. 8*(iii)*, while the second logical implication reduces to

$$\circledast_1(\neg\bigwedge_{j>1}\phi'_j) \vdash \circledast_1^h(\neg\bigwedge_{j>1}\phi'_j),$$

which holds by Prop. 8*(ii)*. This establishes the base case of the induction.

Now let $i > 1$ and assume the property holds for all $j < i$. Note that this implies

$$\bigwedge_{j\in\mathbb{A}}\phi'_j \equiv \bigwedge_{j<i}\phi_j \wedge \bigwedge_{j\geq i}\phi'_j.$$

We have $\phi_i = \divideontimes_i(\neg(\bigwedge_{j<i}\phi_j \wedge \bigwedge_{j>i}\phi'_j))$ and $\phi'_i = \divideontimes^h_i(\neg\bigwedge_{j\neq i}\phi'_j)$. Given all this the first logical implication in (1) above may be rewritten as

$$(\bigwedge_{j<i}\phi_j \wedge \bigwedge_{j>i}\phi'_j) \wedge \divideontimes^h_i(\neg\bigwedge_{j\neq i}\phi'_j) \vdash \divideontimes_i(\neg(\bigwedge_{j<i}\phi_j \wedge \bigwedge_{j>i}\phi'_j)).$$

Now since $\phi_j \vdash \phi'_j$ for all $j < i$ (inductive hypothesis) we know $\neg\bigwedge_{j\neq i}\phi'_j \vdash \neg(\bigwedge_{j<i}\phi_j \wedge \bigwedge_{j>i}\phi'_j)$. Hence, using the following derived property of basic removal functions (which is mainly a consequence of **(⊛3)**),

(⊛A) If $\lambda \vdash \chi$ then $\neg\chi \wedge \divideontimes(\lambda) \vdash \divideontimes(\chi)$,

we see the left-hand side above logically implies $(\bigwedge_{j<i}\phi_j \wedge \bigwedge_{j>i}\phi'_j) \wedge \divideontimes^h_i(\neg(\bigwedge_{j<i}\phi_j \wedge \bigwedge_{j>i}\phi'_j)$. From this we get the right-hand side as a logical conclusion from Prop. 8*(iii)*.

For the second implication in (1) $\phi_i \vdash \phi'_i$ we must show

$$\divideontimes_i(\neg(\bigwedge_{j<i}\phi_j \wedge \bigwedge_{j>i}\phi'_j)) \vdash \divideontimes^h_i(\neg\bigwedge_{j\neq i}\phi'_j).$$

By Prop. 8*(ii)* it suffices to show

$$\divideontimes^h_i(\neg(\bigwedge_{j<i}\phi_j \wedge \bigwedge_{j>i}\phi'_j)) \vdash \divideontimes^h_i(\neg\bigwedge_{j\neq i}\phi'_j).$$

By the indictive hypothesis $\phi_j \vdash \phi'_j$ for all $j < i$ we know $\neg\bigwedge_{j\neq i}\phi'_j \vdash \neg(\bigwedge_{j<i}\phi_j \wedge \bigwedge_{j>i}\phi'_j)$ and so we may use **(⊛6)** to obtain this implication *provided* we can show $\divideontimes^h_i(\neg\bigwedge_{j\neq i}\phi'_j) \not\vdash \neg(\bigwedge_{j<i}\phi_j \wedge \bigwedge_{j>i}\phi'_j)$. Since $\divideontimes^h_i(\neg\bigwedge_{j\neq i}\phi'_j) = \phi'_i$ this just boils down to showing $\bigwedge_{j<i}\phi_j \wedge \bigwedge_{j\geq i}\phi'_j$ is consistent. But as remarked above, the inductive hypothesis implies this is equivalent to $\bigwedge_j \phi'_j$ which is clearly consistent. This completes the inductive step and so our property is proved, namely

$$\bigwedge_{j\in\mathbb{A}}\phi'_j \vdash \phi_i \vdash \phi'_i \text{ for all } i \in \mathbb{A}.$$

Note that this implies, for all i,

$$\bigwedge_{j\in\mathbb{A}}\phi'_j \equiv \bigwedge_{j\leq i}\phi_i \wedge \bigwedge_{j>i}\phi'_j. \quad (2)$$

In particular $\bigwedge_{j\in\mathbb{A}}\phi'_j \equiv \bigwedge_{j\in\mathbb{A}}\phi_j$, which proves the second part of the proposition. Now, we want to show $(\phi_i)_{i\in\mathbb{A}}$ is a removal equilibrium for $(\divideontimes_i)_{i\in\mathbb{A}}$, which means we need to show, for all $i \in \mathbb{A}$, $\phi_i \equiv \divideontimes_i(\neg\bigwedge_{j\neq i}\phi_j)$, i.e.,

$$\divideontimes_i(\neg(\bigwedge_{j<i}\phi_j \wedge \bigwedge_{j>i}\phi'_j)) \equiv \divideontimes_i(\neg\bigwedge_{j\neq i}\phi_j).$$

For simplicity let us write $\sigma = \neg(\bigwedge_{j<i}\phi_j \wedge \bigwedge_{j>i}\phi'_j)$ and $\rho = \neg\bigwedge_{j\neq i}\phi_j$. So we must show $\divideontimes_i(\sigma) \equiv \divideontimes_i(\rho)$. Using the just established fact that $\phi_j \vdash \phi'_j$ for all j we know $\sigma \vdash \rho$ and so $\sigma \equiv \rho \wedge (\sigma \vee \neg\rho)$. Hence $\divideontimes_i(\sigma) \equiv \divideontimes_i(\rho \wedge (\sigma \vee \neg\rho))$. Now using **(⊛4)** and **(⊛6)** we have $\divideontimes_i(\rho\wedge(\sigma\vee\neg\rho)) \equiv \divideontimes_i(\rho)$ *if* $\divideontimes_i(\rho \wedge (\sigma \vee \neg\rho)) \vdash \sigma \vee \neg\rho$. Hence if we can show $\divideontimes_i(\sigma) \vdash \sigma \vee \neg\rho$, equivalently $\divideontimes_i(\sigma) \wedge \neg\sigma \vdash \neg\rho$, then we obtain the desired conclusion. But we have $\divideontimes_i(\sigma) \wedge \neg\sigma \equiv \bigwedge_{j\leq i}\phi_i \wedge \bigwedge_{j>i}\phi'_j$. From property (2) this in turn is equivalent to $\bigwedge_{j\in\mathbb{A}}\phi_j$, and so we obtain $\divideontimes_i(\sigma) \wedge \neg\sigma \vdash \bigwedge_{j\neq i}\phi_j \equiv \neg\rho$ as required. □

The second part of this proposition implies that if we are interested only in the result of *merging*, we might as well just use the Hyperregular versions.

Entrenchment equilibria

In this section we investigate another equilibrium notion for social belief removal, which is more directly comparable to the usual notion of Nash equilibrium in strategic games. To do so we will first show how any removal profile $(\divideontimes_i)_{i\in\mathbb{A}}$ defines a particular strategic game $\mathcal{G}((\divideontimes_i)_{i\in\mathbb{A}})$ and then use the Nash equilibria of this game to define our new notion of equilibrium. We start by recalling the definitions of strategic game and Nash equilibrium. (See, e.g., (Osborne and Rubinstein 1994).)

Definition 7 *A* strategic game (over $\mathbb{A}$) *is a pair* $\langle(A_i)_{i\in\mathbb{A}}, (\succsim_i)_{i\in\mathbb{A}}\rangle$*, where, for each* $i \in \mathbb{A}$*:*

- A_i *is the set of* actions *available to agent* i*,*
- $\succsim_i$ *is a total preorder over* $\times_{i\in\mathbb{A}}A_i$*, i.e., the* preference relation *of agent* i*.*

The set $\times_{i\in\mathbb{A}}A_i$ is the set of *action profiles* for the agents in $\mathbb{A}$, i.e., the set of tuples consisting of a chosen action $a_i \in A_i$ for each agent i. Given two action profiles $(a_i)_{i\in\mathbb{A}}$ and $(b_i)_{i\in\mathbb{A}}$, $(a_i)_{i\in\mathbb{A}} \succsim_j (b_i)_{i\in\mathbb{A}}$ means agent j prefers (the outcome resulting from) the action profile $(b_i)_{i\in\mathbb{A}}$ *at least as much as* $(a_i)_{i\in\mathbb{A}}$.

Definition 8 *A* Nash equilibrium of a strategic game $\langle(A_i)_{i\in\mathbb{A}}, (\succsim_i)_{i\in\mathbb{A}}\rangle$ *is an action profile* $(a^*_i)_{i\in\mathbb{A}}$ *such that, for each* $j \in \mathbb{A}$*, and any* $a_j \in A_j$ *we have* $(a_i)_{i\in\mathbb{A}} \succsim_j (a^*_i)_{i\in\mathbb{A}}$*, where* $a_i = a^*_i$ *for* $i \neq j$*.*

In a Nash equilibrium no single agent can change his action in a way which leads to a more preferred outcome for him, *given* that the other agents' actions remain fixed.

How can we define a strategic game from a removal profile? Well first note in our situation of social belief removal too each agent takes an action – he chooses which sentence to remove. That is, the set of possible actions of agent i may be identified with L_*. What, then, is the preference relation of agent i over the resulting set of action profiles $\times_{j\in\mathbb{A}}L_*$? Clearly each agent prefers any action profile leading to a consistent outcome over one which leads to inconsistency. But what is his preference between different profiles leading to consistent outcomes? One natural idea is that agents prefer to remove *less entrenched* sentences (Gärdenfors 1988). Given agent i is using removal function $\divideontimes_i$, his *entrenchment ordering* (over L_*) $\trianglelefteq^{\mathrm{E}}_i$ is given by

$$\lambda \trianglelefteq^{\mathrm{E}}_i \chi \text{ iff } \divideontimes_i(\lambda \wedge \chi) \not\vdash \lambda.$$

Thus χ is *at least as entrenched as* λ iff the removal of the conjuction causes λ to be excluded. It expresses that agent i finds it *at least as easy to discard* λ as χ.

Proposition 10 *If $\circledast_i$ is a basic removal function, and $\trianglelefteq_i^{\mathrm{E}}$ is defined from $\circledast_i$ as above then $\trianglelefteq_i^{\mathrm{E}}$ forms a standard entrenchment ordering in the sense of (Gärdenfors 1988). In particular $\trianglelefteq_i^{\mathrm{E}}$ is a total preorder over L_*.*

Given this, agent i's preference relation $\precsim_i^{\mathrm{E}}$ over the set $\times_{j\in\mathbb{A}} L_*$ is may be specified completely as follows. Given any two action profiles $(\lambda_j)_{j\in\mathbb{A}}$ and $(\chi_j)_{j\in\mathbb{A}}$, we set $(\lambda_j)_{j\in\mathbb{A}} \precsim_i^{\mathrm{E}} (\chi_j)_{j\in\mathbb{A}}$ iff one of the following two conditions hold:

either *(i).* $(\circledast_j(\lambda_j))_{j\in\mathbb{A}}$ is inconsistent
or *(ii).* $(\circledast_j(\lambda_j))_{j\in\mathbb{A}}$ and $(\circledast_j(\chi_j))_{j\in\mathbb{A}}$ are both consistent and $\chi_j \trianglelefteq_i^{\mathrm{E}} \lambda_j$.

Since $\trianglelefteq_i^{\mathrm{E}}$ is a total preorder over L_*, it is easy to check $\precsim_i^{\mathrm{E}}$ forms a total preorder over the set of all action profiles.

Definition 9 *Given a removal profile $(\circledast_i)_{i\in\mathbb{A}}$, the strategic game $\langle (L_*)_{i\in\mathbb{A}}, (\precsim_i^{\mathrm{E}})_{i\in\mathbb{A}}\rangle$ defined from $(\circledast_i)_{i\in\mathbb{A}}$ as above will be denoted by $\mathcal{G}((\circledast_i)_{i\in\mathbb{A}})$.*

Given all this, we are ready to define our next equilibrium notion.

Definition 10 $(\phi_i)_{i\in\mathbb{A}}$ *is an* entrenchment equilibrium *for $(\circledast_i)_{i\in\mathbb{A}}$ iff it is consistent and $(\phi_i)_{i\in\mathbb{A}} \equiv (\circledast_i(\lambda_i^*))_{i\in\mathbb{A}}$ for some Nash equilibrium $(\lambda_i^*)_{i\in\mathbb{A}}$ of the game $\mathcal{G}((\circledast_i)_{i\in\mathbb{A}})$.*

Put more directly, an entenchment equilibrium is an outcome $(\phi_i)_{i\in\mathbb{A}}$ which is consistent and for which no *single* agent may deviate and remove a less entrenched sentence *without* destroying this consistency.

This brings us to the following social removal property:

(FEEq) $\mathbf{F}((\circledast_i)_{i\in\mathbb{A}})$ is an entrenchment equilibrium for $(\circledast_i)_{i\in\mathbb{A}}$.

What is the relationship between entrenchment equilibria and removal equilibria?

Proposition 11 *Every removal equilibrium for $(\circledast_i)_{i\in\mathbb{A}}$ is an entrenchment equilibrium for $(\circledast_i)_{i\in\mathbb{A}}$. Furthermore if all $\circledast_i$ are hyperregular then every entrenchment equilibrium for $(\circledast_i)_{i\in\mathbb{A}}$ is a removal equilibrium for $(\circledast_i)_{i\in\mathbb{A}}$.*

Proof. Let $(\phi_i)_{i\in\mathbb{A}}$ be a removal equilibrium and let $i \in \mathbb{A}$. Then $\phi_i \equiv \circledast_i(\neg\bigwedge_{j\neq i}\phi_j)$. We will show the profile $(\neg\bigwedge_{j\neq i}\phi_j)_{i\in\mathbb{A}}$ is a Nash equilibrium for $\mathcal{G}((\circledast_i)_{i\in\mathbb{A}})$. Suppose $\chi \lhd_i^{\mathrm{E}} \neg\bigwedge_{j\neq i}\phi_j$, where $\lhd_i^{\mathrm{E}}$ is the strict part of i's entrenchment relation $\trianglelefteq_i^{\mathrm{E}}$. This means

$$\circledast_i(\chi \wedge \neg\bigwedge_{j\neq i}\phi_j) \vdash \neg\bigwedge_{j\neq i}\phi_j. \tag{3}$$

We must show $(\phi_1,\ldots,\circledast_i(\chi),\ldots,\phi_n)$ is inconsistent, i.e., $\circledast_i(\chi)\vdash\neg\bigwedge_{j\neq i}\phi_j$. But since $\chi \trianglelefteq_i^{\mathrm{E}} \neg\bigwedge_{j\neq i}\phi_j$ we know

$$\circledast_i(\chi \wedge \neg\bigwedge_{j\neq i}\phi_j) \nvdash \chi. \tag{4}$$

From this and **(⊛6)** we get $\circledast_i(\chi)\vdash\circledast_i(\chi\wedge\neg\bigwedge_{j\neq i}\phi_j)$ and so from this and (3) we obtain the desired conclusion.

For the second part, let $\circledast_i$ be a hyperregular removal function for each $i\in\mathbb{A}$ and suppose $(\phi_i)_{i\in\mathbb{A}}$ is consistent and logically equivalent to $(\circledast_i(\lambda_i^*))_{i\in\mathbb{A}}$ for some Nash equilibrium $(\lambda_i^*)_{i\in\mathbb{A}}$ for $\mathcal{G}((\circledast_i)_{i\in\mathbb{A}})$. We need to show $\phi_i \equiv \circledast_i(\neg\bigwedge_{j\neq i}\phi_i)$ for all i, i.e., $\circledast_i(\lambda_i^*)\equiv\circledast_i(\neg\bigwedge_{j\neq i}\phi_i)$. First we show λ_i^* and $\neg\bigwedge_{j\neq i}\phi_i$ are equally entrenched according to $\circledast_i$. That $\lambda_i^* \trianglelefteq_i^{\mathrm{E}} \neg\bigwedge_{j\neq i}\phi_i$ holds since $(\phi_1,\ldots,\circledast_i(\neg\bigwedge_{j\neq i}\phi_i),\ldots,\phi_n)$ is consistent (by **(⊛1)**) and so $\neg\bigwedge_{j\neq i}\phi_i \lhd_i^{\mathrm{E}} \lambda_i^*$ would contradict that $(\lambda_i^*)_{i\in\mathbb{A}}$ is a Nash equilibrium for $\mathcal{G}((\circledast_i)_{i\in\mathbb{A}})$. That $\neg\bigwedge_{j\neq i}\phi_i \trianglelefteq_i^{\mathrm{E}} \lambda_i^*$ follows since we have $\circledast_i(\lambda_i^*)\nvdash\neg\bigwedge_{j\neq i}\phi_i$ (since $(\phi_i)_{i\in\mathbb{A}}$ is consistent) and so we may deduce $\neg\bigwedge_{j\neq i}\phi_i \trianglelefteq_i^{\mathrm{E}} \lambda_i^*$ using the following derived rule for basic removals:

(⊛B) If $\circledast(\lambda)\nvdash\chi$ then $\circledast(\lambda\wedge\chi)\nvdash\chi$.

Hence we have shown λ_i^* and $\neg\bigwedge_{j\neq i}\phi_i$ are equally entrenched according to $\circledast_i$, and the result now follows from the fact that for hyperregular removals, removing equally entrenched sentences yields logically equivalent results. □

Thus if all agents use hyperregular removal then the two notions of equilibrium *coincide*. However, in general, not every entrenchment equilibrium is a removal equilibrium, since for example if $(\circledast_i(\bot))_{i\in\mathbb{A}}$ is consistent then it is *always* an entrenchment equilibrium, because $\bot$ is always minimally entrenched for any basic removal function. However we have already seen that it might not be a removal equilibrium. However we can show the following:

Proposition 12 *For every entrenchment equilibrium $(\phi_i')_{i\in\mathbb{A}}$ for $(\circledast_i)_{i\in\mathbb{A}}$ there exists a removal equilibrium $(\phi_i)_{i\in\mathbb{A}}$ for $(\circledast_i)_{i\in\mathbb{A}}$ such that $(\phi_i)_{i\in\mathbb{A}} \equiv_\wedge (\phi_i')_{i\in\mathbb{A}}$.*

Proof. First observe that the definition of entrenchment equilibrium for $(\circledast_i)_{i\in\mathbb{A}}$ depends only on the *entrenchment relations* $\trianglelefteq_i^{\mathrm{E}}$ generated from $\circledast_i$ (because the definition of $\mathcal{G}((\circledast_i)_{i\in\mathbb{A}})$ does). Hence if two removal profiles generate the same tuple of entrenchment relations, they will have the same set of entrenchment equilibria. Now, let $(\phi_i')_{i\in\mathbb{A}}$ be an entrenchment equilibrium for $(\circledast_i)_{i\in\mathbb{A}}$. By Prop. 8*(iii)* we know each $\circledast_i$ is revision-equivalent to its hyperregular version $\circledast_i^h$. Since revision-equivalent removal functions generate the same entrenchment relation this means $(\circledast_i)_{i\in\mathbb{A}}$ and $(\circledast_i^h)_{i\in\mathbb{A}}$ must have the same set of entrenchment equilibria, and so $(\phi_i')_{i\in\mathbb{A}}$ is also an entrenchment equilibrium for $(\circledast_i^h)_{i\in\mathbb{A}}$. We may now apply exactly the same construction as in Prop. 9 to obtain the desired removal equilibrium for $(\circledast_i)_{i\in\mathbb{A}}$. □

Strong entrenchment equilibria

As we saw in Example 3, even in the hyperregular case, $(\circledast_i(\bot))_{i\in\mathbb{A}}$ might not be the *only* entrenchment equilibrium. It might seem irrational for both agents to give up p in this example, when it's possible for both to remove a less entrenched sentence (i.e. $\bot$) while preserving consistency. This kind of counterintuitive result is not restricted to entrenchment equilibria. In fact it is inherent in the concept of Nash equilibrium itself. It has long been recognised that the Nash equilibrium does not rule out sub-optimal solutions

in the case where agents have identical preferences over outcomes. This is illustrated by the following example, taken from (Osborne and Rubinstein 1994, p16).

Example 5 *Suppose two agents* $\{1, 2\}$ *who wish to go to a concert together, but must choose between going to a Mozart (Mo) concert or a Mahler (Ma) concert. Thus the set of actions for both agents is* $A = \{Mo, Ma\}$*. We assume both agents have identical preferences over the four possible action profiles. Firstly, the agents want to reach agreement, so the two profiles in which they choose different actions are the least preferred. Moreover, both agents prefer to see the Mozart concert. Thus the preference relation* $\succsim$ *of both agents is specified completely by*

$$(Mo, Ma) \sim (Ma, Mo) \prec (Ma, Ma) \prec (Mo, Mo).$$

(Just for this example we are using $\sim$ *and* $\prec$ *to denote the symmetric closure and strict part of* $\succsim$ *respectively.) In this game there are two Nash equilibria* (Ma, Ma) *and* (Mo, Mo)*. Even though both agents have a mutual interest in reaching* (Mo, Mo)*, the Nash equilibrium does not rule out the inferior outcome* (Ma, Ma)*.*

This anomaly led several authors to propose refined equilibria concepts for strategic games. One such refinement, the *strong* Nash equilibrium (Aumann 1959), says roughly that no *set* – not just singletons as with Nash – of players can make a joint change in strategy which leads to a more preferred outcome for all players in that set.

Definition 11 *A* strong Nash equilibrium of a strategic game $\langle (A_i)_{i\in\mathbb{A}}, (\succsim_i)_{i\in\mathbb{A}} \rangle$ *is an action profile* $(a_i^*)_{i\in\mathbb{A}}$ *such that, for any* $X \subseteq \mathbb{A}$*, and each tuple* $(a_i)_{i\in X}$*, there exists* $j \in X$ *such that* $(a_i)_{i\in\mathbb{A}} \precsim_j (a_i^*)_{i\in\mathbb{A}}$*, where* $a_i = a_i^*$ *for* $i \notin X$*.*

This leads to the corresponding refinement for entrenchment equilibria.

Definition 12 $(\phi_i)_{i\in\mathbb{A}}$ *is a* strong entrenchment equilibrium *for* $(\circledast_i)_{i\in\mathbb{A}}$ *iff it is consistent and* $(\phi_i)_{i\in\mathbb{A}} \equiv (\circledast_i(\lambda_i^*))_{i\in\mathbb{A}}$ *for some strong Nash equilibrium* $(\lambda_i^*)_{i\in\mathbb{A}}$ *of the game* $\mathcal{G}((\circledast_i)_{i\in\mathbb{A}})$*.*

The following property thus strengthens (**FEEq)**:

(FEEq+) $\mathbf{F}((\circledast_i)_{i\in\mathbb{A}})$ is a strong entrenchment equilibrium for $(\circledast_i)_{i\in\mathbb{A}}$.

In Example 3 the only strong entrenchment equilibrium is $(p)_{i\in\mathbb{A}}$. For hyperregular removal profiles, it can be shown function $\mathbf{F}_1$ defined earlier satisfies **(FEEq+)**, but $\mathbf{F}_2$ does not. Thus strong entrenchment equilibria *always* exist for hyperregular removal profiles. However at the time of writing it is an open problem whether they are guaranteed to exist for general basic removal profiles. It would also be interesting to try and find a necessary and sufficient condition for a removal equilibrium to be a strong entrenchment equilibrium (even in the hyperregular case).

Equilibria as maxiconsistent sets

The simplest kind of removal function is what might be termed *drastic removal*, in which the result of removing λ is $\circledast(\bot)$ if λ is not entailed by the initial belief set, or $\top$ if it *is* entailed. That is, an agent either leaves his belief set unchanged, or throws out *all* beliefs. Drastic removals correspond to the severe withdrawal functions determined by single-element logical chains.

If all agents use drastic removal, then removal/entrenchment equilibria reduce to taking *maximal consistent sets of agents*. $X \subseteq \mathbb{A}$ is maximally consistent iff *(i)* $\bigwedge_{i\in X} \circledast_i(\bot)$ is consistent, and *(ii)* $\bigwedge_{i\in Y} \circledast_i(\bot)$ is inconsistent for all $X \subset Y \subseteq \mathbb{A}$.

Proposition 13 *Suppose all* $\circledast_i$ *are drastic removal functions. Then* $(\phi_i)_{i\in\mathbb{A}}$ *is a removal (or entrenchment) equilibrium for* $(\circledast_i)_{i\in\mathbb{A}}$ *iff* $\{i \mid \phi_i \equiv \circledast_i(\bot)\}$ *is a maximally consistent subset of* $\mathbb{A}$*.*

Thus we see that the main notions of equilibria studied in this paper (removal and entrenchment) can be seen as *generalisations* of the idea of taking maximal consistent sets.

Related work

While this paper is, to our knowledge, the first attempt to define explicit notions of equilibria in a belief merging setting, a proposal that is similar in spirit has been made in the context of *negotiation*. In a series of papers, Zhang et al. (2004) and Meyer et al. (2004b; 2004a) considered the problem of negotiation from a belief change perspective. They consider the case of negotiation involving only two agents, but make it clear that the real interest is in a setting involving a finite number n of agents. The initial demands of agents are represented as (logically closed) belief sets, and are compared to the beliefs of agents. Negotiation is then described as a process in which agents *strike a deal* by modifying their initial demands to obtain new belief sets, say ϕ_1 and ϕ_2. The outcome of the process of negotiation is the conjunction $\phi_1 \wedge \phi_2$ of the modified belief sets. Negotiation in this sense is thus closely related to belief merging and hence, indirectly, to social belief removal.

The basic assumptions in the series of papers differ from those made in this paper. Zhang et al. (2004) define the modified belief sets in terms of belief revision, and in particular, basic AGM revision (i.e. revision operators satisfying the first six AGM revision postulates (Alchourrón, Gärdenfors, and Makinson 1985)). Meyer et al. (2004a) consider modified belief sets in terms of both contraction and revision, but assuming basic AGM revision and contraction. This was also extended to full AGM contraction and revision (Meyer et al. 2004b).

Despite these differences regarding basic assumptions, there are some interesting similarities between their work and the notion of a removal equilibrium. Zhang et al. (2004) characterise the modified belief sets ϕ_1 and ϕ_2 in terms of the following fixed-point definition, using belief revision functions $+_1$ and $+_2$ for agents 1 and 2 respectively:

(FP) $\phi_1 \wedge \phi_2 \equiv +_1(\phi_2) \vee +_2(\phi_1)$

That is, the outcome of a deal $(\phi_1 \wedge \phi_2)$ is equivalent to the disjunction of the result of agent 1 revising with the revised demands of agent 2, and the result of agent 2 revising with the revised demands of agent 1. To compare this with our

results, observe firstly that for the case of two agents, the definition of a removal equilibrium reduces to the following:

Definition 13 $(\phi_i)_{i\in\{1,2\}}$ *is a* removal equilibrium *for* $(\divideontimes_i)_{i\in\{1,2\}}$ *iff it is consistent,* $\phi_1 \equiv \divideontimes_1(\neg\phi_2)$ *and* $\phi_2 \equiv \divideontimes_2(\neg\phi_1)$.

From Defn. 13 it follows immediately that

$$\phi_1 \wedge \phi_2 \equiv \divideontimes_1(\neg\phi_2) \wedge \divideontimes_2(\neg\phi_1)$$

which can almost be seen as the dual of (FP): revision functions are replaced by removal functions, the input to the functions are negated, and we take the conjunction of the result instead of the disjunction.

In fact, there is an even closer link between our work and theirs. For each $i = 1, 2$ let $+_i$ be the revision function defined from $\divideontimes_i$ using the Levi Identity, i.e., $+_i(\phi) = \divideontimes_i(\neg\phi) \wedge \phi$. Now, from Defn. 13 it follows that $\phi_1 \equiv \divideontimes_1(\neg\phi_2)$, and therefore that

$$\phi_1 \wedge \phi_2 \equiv \divideontimes_1(\neg\phi_2) \wedge \phi_2 \equiv +_1(\phi_2).$$

Switching the roles of ϕ_1 and ϕ_2 we also get

$$\phi_1 \wedge \phi_2 \equiv \divideontimes_2(\neg\phi_1) \wedge \phi_1 \equiv +_2(\phi_1).$$

From this it follows that

$$\phi_1 \wedge \phi_2 \equiv +_1(\phi_2) \vee +_2(\phi_1).$$

In other words, assuming the same class of removal functions, and assuming the Levi Identity, the fixed-point construction (FP) actually follows from Defn. 13.

Conclusion

We have defined several notions of equilibrium in the framework of social removal functions, formulated purely in the language of belief removal operators. Assuming all agents use basic removal functions to remove their own beliefs, we proved our equilibria are always guaranteed to exist. We gave several examples to illustrate these notions, and we showed that they generalise in some sense the idea of resolving inconsistency by taking maximal consistent subsets of agents.

For future work, we want to generalise our results to handle social removal under *integrity constraints* (Konieczny and Pino Pérez 2002). An *IC social removal function* is a function taking as arguments a removal profile and a consistent sentence Ψ, which returns a belief profile which is consistent *with* Ψ. The equilibrium notions described in this paper should extend to this setting. For example an IC removal equilibrium could be defined to be any belief profile $(\phi_i)_{i\in\mathbb{A}}$ for which $\phi_i \equiv \divideontimes_i(\neg(\Psi \wedge \bigwedge_{j\neq i} \phi_j))$ for all i.

Social belief removal functions have obvious similarities to *social choice rules* (Arrow, Sen, and Suzumura 2002). A social choice rule takes as input a profile of total preorders over the set of alternatives together with a given subset A of the alternatives, and outputs a subset of A – the *chosen* elements of A for the group. By conjoining the elements of the output of a social belief removal function we obtain an output of the same type as with social choice rules, but the input of a social belief removal function can be viewed as *richer* than that for social choice, since a basic removal function corresponds to a total preorder $\leq$ *plus* a reflexive sub-relation $\preceq$. It would be interesting to explore any (im)possibility theorems for social removal functions.

Acknowledgements

Thanks are due to Alexander Nittka and to three anonymous KR'08 reviewers whose comments greatly helped to improve the paper. This material is based upon work supported by the National Research Foundation under Grant number 65152.

References

Alchourrón, C.; Gärdenfors, P.; and Makinson, D. 1985. On the logic of theory change: Partial meet contraction and revision functions. *Journal of Symbolic Logic* 50(2):510–530.

Arrow, K.; Sen, A.; and Suzumura, K., eds. 2002. *Handbook of Social Choice and Welfare*. Elsevier.

Aumann, R. 1959. Acceptable points in general cooperative n-person games. In *Contributions to the Theory of Games, Vol. IV*, 287–324.

Bochman, A. 2001. *A Logical Theory of Nonmonotonic Inference and Belief Change*. Springer.

Booth, R.; Chopra, S.; Meyer, T.; and Ghose, A. 2004. A unifying semantics for belief change. In *Proceedings of ECAI'04*, 793–797.

Booth, R.; Chopra, S.; Ghose, A.; and Meyer, T. 2005. Belief liberation (and retraction). *Studia Logica* 79(1):47–72.

Booth, R. 2006. Social contraction and belief negotiation. *Information Fusion* 7(1):19–34.

Cantwell, J. 2003. Eligible contraction. *Studia Logica* 73:167–182.

Gärdenfors, P. 1988. *Knowledge in Flux*. MIT Press.

Hansson, S. O. 1991. Belief contraction without recovery. *Studia Logica* 50(2):251–260.

Hansson, S. O. 1993a. Changes on disjunctively closed bases. *Journal of Logic, Language and Information* 2:255–284.

Hansson, S. O. 1993b. Theory contraction and base contraction unified. *Journal of Symbolic Logic* 58:602–625.

Konieczny, S., and Gregoire, E. 2006. Logic-based approaches to information fusion. *Information Fusion* 7(1):4–18.

Konieczny, S., and Pino Pérez, R. 2002. Merging information under constraints: A logical framework. *Journal of Logic and Computation* 12(5):773–808.

Kraus, S.; Lehmann, D.; and Magidor, M. 1991. Nonmonotonic reasoning, preferential models and cumulative logics. *Artificial Intelligence* 44:167–207.

Levi, I. 1991. *The Fixation of Belief and Its Undoing*. Cambridge: Cambridge University Press.

Meyer, T.; Heidema, J.; Labuschagne, W.; and Leenen, L. 2002. Systematic withdrawal. *Journal of Philosophical Logic* 31(5):415–443.

Meyer, T.; Foo, N.; Kwok, R.; and Zhang, D. 2004a. Logical foundations of negotiation: Outcome, concession and adaptation. In *Proceedings of AAAI'04*, 293–298.

Meyer, T.; Foo, N.; Kwok, R.; and Zhang, D. 2004b. Logical foundations of negotiation: strategies and preferences. In *Proceedings of KR'04*, 311–318.

Nash, J. 1950. Equilibrium points in n-person games. *Proceedings of the National Academy of Sciences* 36(1):48–49.

Nebel, B. 1994. Base revision operations and schemes: Semantics, representation and complexity. In *Proceedings of ECAI'94*, 342–345.

Osborne, M., and Rubinstein, A. 1994. *A Course in Game Theory*. MIT Press.

Rott, H., and Pagnucco, M. 1999. Severe withdrawal (and recovery). *Journal of Philosophical Logic* 28:501–547.

Rott, H. 1992. Preferential belief change using generalized epistemic entrenchment. *Journal of Logic, Language and Information* 1:45–78.

Zhang, D.; Foo, N.; Meyer, T.; and Kwok, R. 2004. Negotiation as mutual belief revision. In *Proceedings of AAAI'04*, 317–323.