

Belief Revision Games

Nicolas Schwind

Transdisciplinary Research Integration Center
National Institute of Informatics
Tokyo, Japan
email: schwind@nii.ac.jp

Katsumi Inoue

National Institute of Informatics
Tokyo, Japan
email: inoue@nii.ac.jp

Gauvain Bourgne

LIP6
Université Pierre et Marie Curie
Paris, France
email: Gauvain.Bourgne@lip6.fr

Sébastien Konieczny

CRIL - CNRS
Université d'Artois
Lens, France
email: konieczny@cril.fr

Pierre Marquis

CRIL - CNRS
Université d'Artois
Lens, France
email: marquis@cril.fr

Abstract

Belief revision games (BRGs) are concerned with the dynamics of the beliefs of a group of communicating agents. BRGs are “zero-player” games where at each step every agent revises her own beliefs by taking account for the beliefs of her acquaintances. Each agent is associated with a belief state defined on some finite propositional language. We provide a general definition for such games where each agent has her own revision policy, and show that the belief sequences of agents can always be finitely characterized. We then define a set of revision policies based on belief merging operators. We point out a set of appealing properties for BRGs and investigate the extent to which these properties are satisfied by the merging-based policies under consideration.

Introduction

In this paper, we introduce belief revision games (BRGs), that are concerned with the dynamics of the beliefs of a group of communicating agents. At each step of the game each agent has to revise her own beliefs by taking account for the beliefs of her acquaintances. BRGs can be viewed as “zero-player” games: each agent has some beliefs (expressed in some finite propositional language) which depend of her previous beliefs and of the previous beliefs of her acquaintances. The aim is to study the dynamics of the game, i.e., the way the beliefs of a group of agents may evolve depending on how agents are ready to share their beliefs (what we call policies). BRGs could be useful to model the evolution of beliefs in a group of agents in social networks, and to study several interesting notions such as influence, manipulation, gossip, etc. In this first paper we mainly focus on the definition of BRGs, using formal tools coming from belief change theory, and investigate their behavior with respect to a set of expected logical properties. Before providing definitions, let us introduce a motivating example of a BRG.

Example 1 Consider a group of three undergraduate students, Alice, Bob and Charles, following the same CS curriculum. Bob is a friend of both Alice and Charles, but Alice and Charles do not know each other. Alice, Bob and

Charles want to prepare the final exam of the “Basics of programming” course. Each student has some feelings about the topics which will be considered by their teacher for this exam. At start, Alice believes that “Binary search” will not be among the topics of the final exam, unlike “Bubble sort”; Bob believes that “Binary search” will be kept by the teacher, and that if “Bubble sort” is kept then “Quick sort” will be chosen as well by the teacher; finally, Charles just feels that “Binary search” will not be considered by the teacher. Each pair of friends exchange their opinions by sending e-mails in the evening. Each student is ready to make her opinion to evolve by adopting the opinions of her friends when this does not conflict with hers, and by considering as most plausible any state of affairs which is as close as possible to the set of opinions at hand (her own one plus her friends’ ones) in the remaining case. At the end of each day, Alice e-mails to Bob with her feelings, Bob to both Alice and Charles, and Charles to Bob. One is asked now about what can be inferred from this description. Some of the key questions are: (1) How beliefs must be updated? (2) Will agents always agree on some pieces of belief if they agree on it at the beginning of the game? (3) Will they eventually stop changing their beliefs?

In the following, we present a formal setting for BRGs. Our very objective is to provide some answers to the questions above. Thus, we address question (1) by putting forward a set of revision policies which are based on existing belief merging operators from the literature and the induced belief revision operators. We identify a set of valuable properties for BRGs. They include unanimity preservation which models question (2) and convergence which models question (3). For each revision policy under consideration, we determine whether such properties are satisfied or not.

An extended version (including proofs) is available at <http://www.cril.fr/marquis/aaai15.pdf>.

Belief Revision Games

Belief sets are represented using a propositional language $\mathcal{L}_{\mathcal{P}}$ defined from a finite set of propositional variables \mathcal{P} and the usual connectives. \perp (resp. \top) is the Boolean constant always false (resp. true.) An interpretation is a total function

from \mathcal{P} to $\{0, 1\}$. The set of all interpretations is denoted \mathcal{W} . An interpretation ω is a model of a formula $\phi \in \mathcal{L}_{\mathcal{P}}$ if and only if it makes it true in the usual truth functional way. $Mod(\phi)$ denotes the set of models of the formula ϕ , i.e., $Mod(\phi) = \{\omega \in \mathcal{W} \mid \omega \models \phi\}$. \models denotes the logical entailment and \equiv the logical equivalence between formulae, i.e., $\phi \models \psi$ iff $Mod(\phi) \subseteq Mod(\psi)$ and $\phi \equiv \psi$ iff $Mod(\phi) = Mod(\psi)$. Two multisets of formulae $\mathcal{K}_1, \mathcal{K}_2$ are said to be equivalent, denoted $\mathcal{K}_1 \equiv \mathcal{K}_2$ if there is a one-to-one correspondence f from \mathcal{K}_1 to \mathcal{K}_2 such that for every $\phi \in \mathcal{K}_1$, $f(\phi) \equiv \phi$.

Let us now introduce the formal definition of a Belief Revision Game.

Definition 1 (Belief Revision Game) A Belief Revision Game (BRG) is a 5-tuple $G = (V, A, \mathcal{L}_{\mathcal{P}}, B, \mathcal{R})$ where

- $V = \{1, \dots, n\}$ is a finite set;
- $A \subseteq V \times V$ is an irreflexive binary relation on V ;
- $\mathcal{L}_{\mathcal{P}}$ is a finite propositional language;
- B is a mapping from V to $\mathcal{L}_{\mathcal{P}}$;
- $\mathcal{R} = \{R_1, \dots, R_n\}$, where each R_i is a mapping from $\mathcal{L}_{\mathcal{P}} \times \mathcal{L}_{\mathcal{P}}^{in(i)}$ to $\mathcal{L}_{\mathcal{P}}$ with $in(i) = \{(j, i) \mid (j, i) \in A\}$, such that for all vectors of formulae $\mathcal{K}_1, \mathcal{K}_2$, if $\mathcal{K}_1 \equiv \mathcal{K}_2$ then $R_i(\mathcal{K}_1) \equiv R_i(\mathcal{K}_2)$, and such that if $in(i) = \emptyset$, then R_i is the identity function.

Let $G = (V, A, \mathcal{L}_{\mathcal{P}}, B, \mathcal{R})$ be a BRG. The set V represents the set of agents under consideration in G . The set A represents the set of *acquaintances* between the agents. Intuitively, if $(i, j) \in A$ then agent j is “aware” of the beliefs of agent i in the sense that agent i communicates her beliefs to agent j when the game is being played. The set B is used to represent each agent’s beliefs which is expressed by a formula from $\mathcal{L}_{\mathcal{P}}$: for each $i \in V$, the formula $B(i)$ (noted B_i for short) is called a *belief state* and represents the initial beliefs of agent i . Lastly, each element $R_i \in \mathcal{R}$ is called the *revision policy* of agent i . Let us denote \mathcal{C}_i the *context* of i , defined as the sequence $\mathcal{C}_i = B_{i_1}, \dots, B_{i_{in(i)}}$ where $i_1 < \dots < i_{in(i)}$ and $\{i_1, \dots, i_{in(i)}\} = \{i_j \mid (i_j, i) \in A\}$. Then $R_i(B_i, \mathcal{C}_i)$ is the belief state of agent i once revised by taking into account her own current beliefs B_i and her current context. It is assumed by definition that all beliefs are considered up to equivalence (i.e., the syntactical form of the beliefs does not matter) and that an agent’s beliefs do not evolve spontaneously when she has no neighbor.

Playing a BRG consists in determining how the beliefs of each agent evolve each time a revision step is performed. This calls for a notion of “belief sequence”, which makes precise the dynamics of the game:

Definition 2 (Belief Sequence) Given a BRG $G = (V, A, \mathcal{L}_{\mathcal{P}}, B, \mathcal{R})$ and an agent $i \in V$, the belief sequence of i , denoted $(B_i^s)_{s \in \mathbb{N}}$, states how the beliefs of agent i evolve while moves take place. $(B_i^s)_{s \in \mathbb{N}}$ is inductively defined as follows:

- $B_i^0 = B_i$;
- $B_i^{s+1} = R_i(B_i^s, \mathcal{C}_i^s)$ for every $s \in \mathbb{N}$, where \mathcal{C}_i^s is the context of i at step s .

B_i^s denotes the belief state of agent i after s moves. The BRG G at step $s \in \mathbb{N}$ is defined as the BRG $G^s = (V, A, \mathcal{L}_{\mathcal{P}}, B^s, \mathcal{R})$, where $B^s = \{B_i^s \mid i \in V\}$.

Since $\mathcal{L}_{\mathcal{P}}$ is a finite propositional language, there exists only finitely many formulae up to equivalence, hence only finitely many belief states can be reached. In order to make it formal, we need the concept of belief cycle:

Definition 3 (Belief Cycle) A series $(K^s)_{s \in \mathbb{N}}$ of formulae from $\mathcal{L}_{\mathcal{P}}$ is cyclic if there exists a finite subsequence K^b, \dots, K^e such that for every $j > e$, we have $K^j \equiv K^{b + ((j-b) \bmod (e-b+1))}$. In this case, the (characteristic) belief cycle of $(K^s)_{s \in \mathbb{N}}$ is defined by the subsequence K^b, \dots, K^e for which b and e are minimal.

By the above argument, it is easy to prove that:

Proposition 1 For every BRG $G = (V, A, \mathcal{L}_{\mathcal{P}}, B, \mathcal{R})$ and every agent $i \in V$, the belief sequence of i is cyclic.

As a consequence, each agent i is associated with a belief cycle which we simply denote $Cyc(B_i)$: the belief sequence of every agent i (which is an infinite sequence) can always be finitely described, since it is entirely characterized by its initial segment $B_i^0, B_i^1, \dots, B_i^{b-1}$ and its belief cycle $Cyc(B_i) = B_i^b, B_i^{b+1}, \dots, B_i^e$, which will be repeated ad infinitum in the sequence.

While there is no winner in a BRG G , such a game can thus be “stopped” after a finite number of steps $stop(G) = \max_{i \in V} \{e \mid Cyc(B_i) = B_i^b, \dots, B_i^e\}$, since when this step $stop(G)$ is reached the belief cycles of all agents can be determined up to equivalence and the future evolution of the agents’ beliefs can be predicted from the sequences of beliefs reached before $stop(G)$.

In the following, we are interested in determining the pieces of beliefs which result from the interaction of the agents participating in the BRG, while focusing on the agents’ belief cycles. A formula ϕ is considered accepted by an agent when it holds in every state of its belief cycle, which means that from some step s , ϕ will always hold. Then we define the notion of acceptability at the agent level and at the group level:

Definition 4 (Acceptability) Let $G = (V, A, \mathcal{L}_{\mathcal{P}}, B, \mathcal{R})$ be a BRG and $\phi \in \mathcal{L}_{\mathcal{P}}$. ϕ is accepted by $i \in V$ if and only if for every $B_i^s \in Cyc(B_i)$, we have $B_i^s \models \phi$. ϕ is unanimously accepted in G if and only if ϕ is accepted by every agent $i \in V$.

A case of interest is when $|Cyc(B_i)| = 1$, i.e., the belief cycle of agent i has length 1. In such a case, the beliefs of agent i “stabilize” once the belief cycle is reached. A specific case is achieved by *stable* BRGs:

Definition 5 (Stability) Let $G = (V, A, \mathcal{L}_{\mathcal{P}}, B, \mathcal{R})$ be a BRG. A belief state $B_i \in B$ is said to be stable in G if $|Cyc(B_i)| = 1$. The BRG G is said to be stable iff each $B_i \in B$ is stable in G .

Stability of a game is an interesting property, since it says in a sense that we reach some equilibrium point, where no agent further changes her belief. These two concepts will take part of some further properties on BRGs which we will introduce and investigate in the following.

Merging-Based Revision Policies

While the general definition of BRG allows all kinds of possible revision policies, we now focus on revision policies R that are rationalized by theoretical tools from Belief Change Theory (see e.g. (Alchourrón, Gärdenfors, and Makinson 1985)), in particular belief merging and belief revision operators. Before introducing specific classes of revision policies of interest, let us introduce some necessary background on belief merging and belief revision.

Formally, given a propositional language $\mathcal{L}_{\mathcal{P}}$ a merging operator Δ is a mapping from $\mathcal{L}_{\mathcal{P}} \times \mathcal{L}_{\mathcal{P}}^n$ to $\mathcal{L}_{\mathcal{P}}$. It associates any formula μ (the *integrity constraints*) and any multiset $\mathcal{K} = \langle K_1, \dots, K_n \rangle$ of belief states (the *profile*) with a new formula $\Delta_{\mu}(\mathcal{K})$ (the *merged state*). A merging operator Δ aims at defining the merged state as the beliefs of a group of agents represented by the profile, under some integrity constraints.

A set of nine standard properties denoted **(IC0)–(IC8)** are expected for merging operators (Konieczny and Pino Pérez 2002). Such operators are called *IC merging operators*. For space reasons, we just recall those used in the rest of the paper:

- (IC0)** $\Delta_{\mu}(\mathcal{K}) \models \mu$;
- (IC1)** If $\mu \not\models \perp$, then $\Delta_{\mu}(\mathcal{K}) \not\models \perp$;
- (IC2)** If $\bigwedge_{K \in \mathcal{K}} K \wedge \mu \not\models \perp$, then $\Delta_{\mu}(\mathcal{K}) \equiv \bigwedge_{K \in \mathcal{K}} K \wedge \mu$;
- (IC3)** If $\mathcal{K}_1 \equiv \mathcal{K}_2$ and $\mu_1 \equiv \mu_2$, then $\Delta_{\mu_1}(\mathcal{K}_1) \equiv \Delta_{\mu_2}(\mathcal{K}_2)$;
- (IC4)** If $K_1 \models \mu$, $K_2 \models \mu$ and $\Delta_{\mu}(\langle K_1, K_2 \rangle) \wedge K_1 \not\models \perp$, then $\Delta_{\mu}(\langle K_1, K_2 \rangle) \wedge K_2 \not\models \perp$.

A couple of additional postulates have been investigated in the literature, which are appropriate for some merging scenarios. We recall below one of them, namely **(Disj)** (Everaere, Konieczny, and Marquis 2010):

- (Disj)** If $\bigvee \mathcal{K} \wedge \mu$ is consistent, then $\Delta_{\mu}(\mathcal{K}) \models \bigvee \mathcal{K}$.

(Disj) is not satisfied by all IC merging operators but is expected in the case when it is assumed that (at least) one of the agent is right (her beliefs hold in the actual world), but we do not know which one.

Distance-based merging operators $\Delta^{d,f}$ are characterized by a pseudo-distance d (i.e., triangular inequality is not mandatory) between interpretations and an (aggregation) function f from $\mathbb{R}^+ \times \dots \times \mathbb{R}^+$ to \mathbb{R}^+ (some basic conditions are required on f , including symmetry and non-decreasingness conditions, see (Konieczny, Lang, and Marquis 2004) for more details). They associate with every formula μ and every profile \mathcal{K} a belief state $\Delta_{\mu}^{d,f}(\mathcal{K})$ which satisfies $Mod(\Delta_{\mu}^{d,f}(\mathcal{K})) = \min(Mod(\mu), \leq_{\mathcal{K}}^{d,f})$, where $\leq_{\mathcal{K}}^{d,f}$ is the total preorder over interpretations induced by \mathcal{K} defined by $\omega \leq_{\mathcal{K}}^{d,f} \omega'$ if and only if $d^f(\omega, \mathcal{K}) \leq d^f(\omega', \mathcal{K})$, where $d^f(\omega, \mathcal{K}) = f_{K \in \mathcal{K}}\{d(\omega, K)\}$ and $d(\omega, K) = \min_{\omega' \models K} d(\omega, \omega')$. Usual distances are d_D , the drastic distance ($d_D(\omega, \omega') = 0$ if $\omega = \omega'$ and 1 otherwise), and d_H the Hamming distance ($d_H(\omega, \omega') = n$ if ω and ω' differ on n variables).

ω	K_1	K_2	K_3	$d_H^{\Sigma}(\omega, \mathcal{K})$	$d_H^{GMin}(\omega, \mathcal{K})$
11	0	2	2	4	(0, 2, 2)
10	1	1	1	3	(1, 1, 1)
01	1	1	1	3	(1, 1, 1)

Table 1: The merging operators $\Delta^{d_H, \Sigma}$ and $\Delta^{d_H, GMin}$.

IC merging operators include some distance-based ones. We mention here two subclasses of them: the summation operators $\Delta^{d, \Sigma}$ (i.e., the aggregation function is the sum Σ) and the GMin operators $\Delta^{d, GMin}$. GMin operators¹ associate with every formula μ and every profile \mathcal{K} a belief state $\Delta_{\mu}^{d, GMin}(\mathcal{K})$ which satisfies $Mod(\Delta_{\mu}^{d, GMin}(\mathcal{K})) = \min(Mod(\mu), \leq_{\mathcal{K}}^{d, GMin})$, where $\leq_{\mathcal{K}}^{d, GMin}$ is the total preorder over interpretations induced by \mathcal{K} defined by $\omega \leq_{\mathcal{K}}^{d, GMin} \omega'$ if and only if $d^{GMin}(\omega, \mathcal{K}) \leq^{lex} d^{GMin}(\omega', \mathcal{K})$ (where \leq^{lex} is the lexicographic ordering induced by the natural order) and $d^{GMin}(\omega, \mathcal{K})$ is the vector of numbers d_1, \dots, d_n obtained by sorting in a non-decreasing order the multiset $\langle d(\omega, K_i) \mid K_i \in \mathcal{K} \rangle$.

Example 2 Let $\mathcal{P} = \{a, b\}$, $\mathcal{K} = \langle K_1, K_2, K_3 \rangle$ where $K_1 = a \wedge b$, $K_2 = K_3 = \neg a \wedge \neg b$, and $\mu = a \vee b$. We consider both summation and GMin operators based on the Hamming distance. Table 1 shows for each interpretation $\omega \in Mod(\mu)$ the distances $d_H(\omega, K_i)$ for $i \in \{1, 2, 3\}$, and the distances $d_H^{\Sigma}(\omega, \mathcal{K})$ and $d_H^{GMin}(\omega, \mathcal{K})$ (interpretations ω are denoted as binary sequences following the ordering $a < b$). We get that $\Delta^{d_H, \Sigma} \equiv (a \wedge \neg b) \vee (\neg a \wedge b)$ and $\Delta^{d_H, GMin} \equiv a \wedge b$.

Noteworthy, summation operators and GMin operators satisfy all **(IC0)–(IC8)** postulates (whatever the pseudo-distance under consideration), and additionally, GMin operators satisfy **(Disj)**, as well as the operator $\Delta^{d_D, \Sigma} = \Delta^{d_D, GMin}$ (**(Disj)** is not satisfied by $\Delta^{d_H, \Sigma}$).

Belief revision operators can be viewed as belief merging operators restricted to singleton profiles: the revision $K_1 \circ K_2$ of a belief state K_1 by another belief state K_2 consists in “merging” the singleton profile $\langle K_1 \rangle$ under the integrity constraints K_2 . Accordingly, the revision operator \circ_{Δ} induced by Δ defined for all states K_1, K_2 as $K_1 \circ_{\Delta} K_2 = \Delta_{K_2}(\langle K_1 \rangle)$ satisfies the standard AGM revision postulates (Alchourrón, Gärdenfors, and Makinson 1985; Katsuno and Mendelzon 1992) if Δ is an IC merging operator.

We are now ready to introduce several classes of revision policies R_i which are parameterized by an IC merging operator Δ and for some of them, by the corresponding revision operator \circ_{Δ} .² Let $G = (V, A, \mathcal{L}_{\mathcal{P}}, B, \mathcal{R})$ be a BRG. In the following, we assume for the sake of simplicity that all

¹Here we give an alternative definition of $\Delta^{d, GMin}$ by means of lists of numbers. However using Ordered Weighted Averages, one could fit the definition of a distance-based operator (Konieczny, Lang, and Marquis 2004).

²When using a merging operator without integrity constraints we just note $\Delta(\mathcal{K})$ instead of $\Delta_{\top}(\mathcal{K})$ for shorter notations.

agents $i \in V$ take advantage of the same revision policy, i.e., given an IC merging operator Δ , for all $R_i \in \mathcal{R}$, $R_i = R_\Delta$. Then let us consider the following revision policies, for any BRG G^s and all agents i who have a non-empty context \mathcal{C}_i :

Definition 6 (Merging-Based Revision Policies)

- $R_\Delta^1(B_i^s, \mathcal{C}_i^s) = \Delta(\langle \mathcal{C}_i^s \rangle)$;
- $R_\Delta^2(B_i^s, \mathcal{C}_i^s) = \Delta_{\Delta(\langle \mathcal{C}_i^s \rangle)}(\langle B_i^s \rangle) \quad [= B_i^s \circ_\Delta \Delta(\langle \mathcal{C}_i^s \rangle)]$;
- $R_\Delta^3(B_i^s, \mathcal{C}_i^s) = \Delta(\langle B_i^s, \mathcal{C}_i^s \rangle)$;
- $R_\Delta^4(B_i^s, \mathcal{C}_i^s) = \Delta(\langle B_i^s, \Delta(\langle \mathcal{C}_i^s \rangle) \rangle)$;
- $R_\Delta^5(B_i^s, \mathcal{C}_i^s) = \Delta_{B_i^s}(\Delta(\langle \mathcal{C}_i^s \rangle)) \quad [= \Delta(\langle \mathcal{C}_i^s \rangle) \circ_\Delta B_i^s]$;
- $R_\Delta^6(B_i^s, \mathcal{C}_i^s) = \Delta_{B_i^s}(\langle \mathcal{C}_i^s \rangle)$.

First of all, please note that since **(IC3)** requires Δ to be syntax-independent (i.e., profiles and integrity constraints are considered up to equivalence), these revision policies are all consistent with the conditions given in Definition 1. As a consequence, each agent is associated with a belief cycle as expected.

Intuitively, these strategies are ranked according to the relative importance given to each agent’s beliefs compared to her neighbors’ opinion. For R_Δ^1 , only the aggregated opinion of the neighbors is relevant. For R_Δ^2 , the current opinion of the agent is revised by the aggregated opinion of the neighbors; doing so, an agent is ready to adopt the part of the merged beliefs of her neighbors which are as close as possible as her own current beliefs. For R_Δ^3 the agent just considers that her opinion is as important as each one of her neighbors. For R_Δ^4 the agent considers that her opinion is as important as the aggregated opinion of her neighbors. R_Δ^5 and R_Δ^6 are similar in the sense that the agent refuses to reject her current beliefs and just accepts additional information which are compatible with them. Noteworthy, R_Δ^5 and R_Δ^6 are not equivalent: for R_Δ^5 the agent first aggregates her neighbors’ opinion, and then revise the merged result by her own opinion; whereas for R_Δ^6 the agent proceeds with her neighbors’ opinion and her own one in a single step.³

Example 1 (continued) We formalize the example presented in the introduction as the BRG $G = (V, A, \mathcal{L}_P, B, \mathcal{R})$ defined as follows. Let $V = \{1, 2, 3\}$ where 1 corresponds to Alice, 2 to Bob, and 3 to Charles. $A = \{(1, 2), (2, 1), (2, 3), (3, 2)\}$ expresses that Alice and Bob are connected, and that Bob and Charles are connected. \mathcal{L}_P is built up from the set of propositional variables $\mathcal{P} = \{s, b, q\}$, where s stands for “Binary Search”, b for “Bubble Sort” and q for “Quick Sort”. The initial beliefs of agents are expressed as $B_1 = \neg s \wedge b$, $B_2 = s \wedge (b \Rightarrow q)$ and $B_3 = \neg s$. Since in the case of conflicting beliefs, each agent considers to merge her friends’ opinions and her own one together, revision policies R_Δ^3 are appropriate candidates for each agent. Let us consider the summation operator based on the Hamming distance, then we have $R_1 = R_2 = R_3 = R_{\Delta^{d_H, \Sigma}}^3$. The belief sequences associated with the three agents are given in Table 2: the belief cycle of

step i	B_1^i	B_2^i	B_3^i
0	$\neg s \wedge b$	$s \wedge (b \Rightarrow q)$	$\neg s$
1	$b \wedge q$	$\neg s \wedge b \wedge q$	$b \Rightarrow q$
≥ 2	$\neg s \wedge b \wedge q$	$\neg s \wedge b \wedge q$	$\neg s \wedge b \wedge q$

Table 2: The belief sequences of Alice, Bob and Charles.

agent 1 (resp. 2, 3) is given by (B_1^2) (resp. (B_2^1) , (B_3^2)). G is a stable game. Note that $\neg s \wedge b \wedge q$ is unanimously accepted in G (as well as all formulae entailed by it).

Logical Properties for Belief Revision Games

In this section, we introduce some expected logical properties on BRGs, and investigate which BRGs satisfy them depending on the chosen revision policy. While the properties hereafter are relevant to all BRGs, we investigate the behavior of some specific classes of BRGs with respect to each introduced property: we focus on BRGs which are instantiated with revision policies from the six classes defined in the previous section, and assume that the same revision policy is applied for each agent. Given a revision policy R_Δ^k , $\mathcal{G}(R_\Delta^k)$ is the set of all BRGs $(V, C, \mathcal{L}_P, C, \mathcal{R})$ where for each $R_i \in \mathcal{R}$, $R_i = R_\Delta^k$. Additionally, for any property P on BRGs, the revision policy R_Δ^k is said to satisfy P if all BRGs from $\mathcal{G}(R_\Delta^k)$ satisfy P .

We start with a set of “preservation” properties which are counterparts of some postulates on belief merging operators (cf. previous section). These properties express the idea that the interaction between agents should not lead them to “degrade” their belief states.

Definition 7 (Consistency Preservation (CP)) A BRG $G = (V, A, \mathcal{L}_P, B, \mathcal{R})$ satisfies **(CP)** if for each $B_i \in B$, if B_i is consistent then all beliefs from $(B_i^s)_{s \in \mathbb{N}}$ are consistent.

This property requires that each agent who has consistent initial beliefs never becomes self-conflicting in her belief sequence. This property is the direct counterpart of **(IC1)** for merging operators. Indeed:

Proposition 2 For every $k \in \{1, \dots, 6\}$, R_Δ^k satisfies **(CP)** if Δ satisfies **(IC1)**.

Definition 8 (Agreement Preservation (AP)) A BRG $G = (V, A, \mathcal{L}_P, B, \mathcal{R})$ satisfies **(AP)** if given any consistent formula $\varphi \in \mathcal{L}_P$, if for each $B_i \in B$, $\varphi \models B_i$ then for each $B_i \in B$ and at every step $s \geq 0$, $\varphi \models B_i^s$.

(AP) requires that if all agents initially agree on some alternatives, then they will not change their mind about them. It corresponds to **(IC2)** for merging operators:

Proposition 3 For every $k \in \{1, \dots, 6\}$, R_Δ^k satisfies **(AP)** if Δ satisfies **(IC2)**.

Definition 9 (Unanimity Preservation (UP)) A BRG $G = (V, A, \mathcal{L}_P, B, \mathcal{R})$ satisfies **(UP)** if given any formula $\varphi \in \mathcal{L}_P$, if for each $B_i \in B$, $B_i \models \varphi$ then for each $B_i \in B$ and at every step $s \geq 0$, $B_i^s \models \varphi$.

³ Consider for instance $\mathcal{C}_i = p \wedge q, \neg p, \neg p \wedge \neg q$ and $B_i = p$. Then $R_{\Delta^{d_D, \Sigma}}^5(B_i, \mathcal{C}_i) \equiv p \wedge \neg q$ whereas $R_{\Delta^{d_D, \Sigma}}^6(B_i, \mathcal{C}_i) \equiv p \wedge q$.

(UP) states that every formula which is a logical consequence of the initial beliefs of each agent should remain so in the belief sequence of each agent; note that in such a case, the formula is unanimously accepted in the BRG under consideration (cf. Definition 4). It is interesting to note that the statements of (AP) and (UP) have quite a similar structure. However, (AP) expresses a unanimity on models whereas (UP) is concerned with unanimity on formulae. The counterpart of these properties for merging operators has been presented in (Everaere, Konieczny, and Marquis 2010). Moreover, it has been shown in (Everaere, Konieczny, and Marquis 2010) that the corresponding postulate of unanimity on formulae for merging operators is equivalent to (Disj) (cf. previous section).

Proposition 4 For every $k \in \{1, \dots, 6\}$, R_{Δ}^k satisfies (UP) if Δ satisfies:

- (IC0) when $k \in \{5, 6\}$;
- (Disj) when $k \in \{1, 3, 4\}$;
- (IC0) and (Disj) when $k = 2$.

In the general case, revision policies R_{Δ}^k with $k \in \{1, 2, 3, 4\}$ do not satisfy (UP) for merging operators Δ which do not satisfy (Disj). This comes from the fact that such merging operators allow for generating new beliefs from the ones forming the profile under consideration: some interpretations that do not satisfy any of the input belief states can be models of the merged state. However, for R_{Δ}^5 and R_{Δ}^6 , Δ is not required to satisfy (Disj). Indeed, in the presence of (IC0) alone these policies are the most change-reluctant ones: each agent who accepts φ at some step will keep accepting φ at the next step since she will only refine her own beliefs. We address more precisely the behavior of all merging-based revision policies in terms of agents' responsiveness with respect to her neighbors:

Definition 10 (Responsiveness (Resp)) A BRG $G = (V, A, \mathcal{L}_{\mathcal{P}}, B, \mathcal{R})$ satisfies (Resp) if for each $B_i \in B$ such that \mathcal{C}_i is not empty, for every step $s \geq 0$, if (i) for every $B_{i_j}^s \in \mathcal{C}_i^s$, $B_{i_j}^s \wedge B_i^s \models \perp$, and (ii) $\bigwedge_{B_{i_j}^s \in \mathcal{C}_i^s} B_{i_j}^s \not\models \perp$, then $B_i^{s+1} \not\models B_i^s$.

Informally, (Resp) demands that an agent should take into consideration the beliefs of her neighbors whenever (i) her beliefs are inconsistent with the beliefs of each one of her neighbors, and (ii) her neighbors agree on some alternatives. Accordingly, (Resp) is not satisfied by R_{Δ}^5 and R_{Δ}^6 :

Proposition 5 If Δ satisfies (IC0), then R_{Δ}^5 and R_{Δ}^6 do not satisfy (Resp).

But (Resp) is satisfied by most of the remaining revision policies R_{Δ}^k under some basic conditions on Δ :

Proposition 6 For every $k \in \{1, 2, 4\}$, R_{Δ}^k satisfies (Resp) if Δ satisfies:

- (IC2) when $k = 1$;
- (IC0) and (IC2) when $k = 2$;
- (IC2) and (IC4) when $k = 4$.

Intuitively, R_{Δ}^3 seems to be less change-reluctant than R_{Δ}^4 , since for R_{Δ}^3 the agent considers her beliefs as being as important as each one of her neighbors whereas for R_{Δ}^4 , she considers her beliefs as being as important as the aggregated beliefs of her neighbors. However, surprisingly R_{Δ}^3 does not satisfy (Resp) even when some “fully rational” IC merging operators Δ are used:

Proposition 7 $R_{\Delta^{d_H, \Sigma}}^3$ does not satisfy (Resp).

Recall that $\Delta^{d_H, \Sigma}$ is the summation merging operator based on the Hamming distance and that it satisfies all the standard IC postulates (IC0)–(IC8). Thus, the fact that Δ satisfies those postulates is not enough for R_{Δ}^3 to satisfy (Resp). However, we show below that these postulates are consistent with (Resp), in the sense that there exists a merging operator Δ satisfying (IC0)–(IC8) (and (Disj)) which makes R_{Δ}^3 a responsive policy:

Proposition 8 For any aggregation function f , $R_{\Delta^{d_D, f}}^3$ satisfies (Resp).

In particular, the revision policy $R_{\Delta^{d_D, \Sigma}}^3 = R_{\Delta^{d_D, GMin}}^3$ satisfies (Resp).

Given a BRG $G = (V, A, \mathcal{L}_{\mathcal{P}}, B, \mathcal{R})$, a formula φ and an agent $i \in V$, let us denote $G_{i \rightarrow \varphi}$ the BRG $(V, A, \mathcal{L}_{\mathcal{P}}, B', \mathcal{R})$ defined as $B'_i = B_i \wedge \varphi$ and for every $j \in V$, $j \neq i$, $B'_j = B_j$.

Definition 11 (Monotonicity (Mon)) A BRG $G = (V, A, \mathcal{L}_{\mathcal{P}}, B, \mathcal{R})$ satisfies (Mon) if whenever φ is unanimously accepted in G , φ is also unanimously accepted in $G_{i \rightarrow \varphi}$ for every $i \in V$.

(Mon) is similar to the *monotonicity criterion* in Social Choice Theory. It is expressed in (Woodall 1997) as the condition where a candidate should not be harmed if she is raised on some ballots without changing the orders of the other candidates. In the BRG context, a formula φ which is unanimously accepted should still be unanimously accepted if some agent's initial beliefs were “strengthened” by φ .

For each revision policy R_{Δ}^k , $k \in \{1, \dots, 6\}$, (Mon) is not guaranteed even when the merging operator under consideration satisfies the postulates (IC0)–(IC8):

Proposition 9 For every $k \in \{1, \dots, 6\}$, $R_{\Delta^{d_H, \Sigma}}^k$ does not satisfy (Mon).

The existence of revision policies R_{Δ}^k which satisfy (Mon) remains an open issue. However, one conjectures that for every $k \in \{1, \dots, 6\}$, $R_{\Delta^{d_D, \Sigma}}^k$ satisfies (Mon). This claim is supported by some empirical evidence. We have conducted non-exhaustive tests when four propositional symbols were considered in the language $\mathcal{L}_{\mathcal{P}}$, for various graph topologies up to 10 agents and for $k \in \{1, \dots, 6\}$. All the tested instances supported the claim.

The last property we introduce concerns the stability issue:

Definition 12 (Convergence) A BRG satisfies (Conv) if it is stable.

Proposition 10 The revision policies R_{Δ}^5 and R_{Δ}^6 satisfy (Conv) if Δ satisfies (IC0).

	(CP)	(AP)	(UP)	(Resp)	(Mon)	(Conv)
R_{Δ}^1	\checkmark (IC1)	\checkmark (IC2)	\checkmark (Disj)	\checkmark (IC2)	\times ($\Delta^{d_H, \Sigma}$)	\times (IC2)
R_{Δ}^2	\checkmark (IC1)	\checkmark (IC2)	\checkmark (IC0) & (Disj)	\checkmark (IC0) & (IC2)	\times ($\Delta^{d_H, \Sigma}$)	\times (IC0) & (IC2)
R_{Δ}^3	\checkmark (IC1)	\checkmark (IC2)	\checkmark (Disj)	\checkmark ($\Delta^{d_D, f}$) / \times ($\Delta^{d_H, \Sigma}$)	\times ($\Delta^{d_H, \Sigma}$)	\times (IC1) & (IC2) & (IC4)
R_{Δ}^4	\checkmark (IC1)	\checkmark (IC2)	\checkmark (Disj)	\checkmark (IC2) & (IC4)	\times ($\Delta^{d_H, \Sigma}$)	\times (IC1) & (IC2) & (IC4)
R_{Δ}^5	\checkmark (IC1)	\checkmark (IC2)	\checkmark (IC0)	\times (IC0)	\times ($\Delta^{d_H, \Sigma}$)	\checkmark (IC0)
R_{Δ}^6	\checkmark (IC1)	\checkmark (IC2)	\checkmark (IC0)	\times (IC0)	\times ($\Delta^{d_H, \Sigma}$)	\checkmark (IC0)

Table 3: Properties satisfied by the revision policies R_{Δ}^k for $k \in \{1, \dots, 6\}$.

None of the remaining revision policies R_{Δ}^k , $k \in \{1, 2, 3, 4\}$ satisfy **(Conv)** in the general case. In fact, for these policies the stability of BRGs cannot be guaranteed as soon as the merging operator under consideration satisfies some basic IC postulates.

Proposition 11 *For every $k \in \{1, 2, 3, 4\}$, R_{Δ}^k does not satisfy **(Conv)** if Δ satisfies:*

- **(IC2)** when $k = 1$;
- **(IC0)** and **(IC2)** when $k = 2$;
- **(IC1)**, **(IC2)** and **(IC4)** when $k \in \{3, 4\}$.

The results presented so far are summarized in Table 3. For each class R_{Δ}^k of revision policies and each property on revision policies, for some (set of) postulate(s) **(P)** on merging operators or directly for some merging operators, \checkmark (**P**) (resp. \times (**P**)) means that R_{Δ}^k satisfies (resp. does not satisfy) the corresponding property when Δ satisfies **(P)** or is one of the merging operators which are specified. One can observe that under some basic conditions on Δ , for $k \in \{1, 2, 4\}$ the revision policies R_{Δ}^k are well-behaved in terms of responsiveness but do not guarantee the stability of all BRGs, while the converse applies for the revision policies R_{Δ}^5 and R_{Δ}^6 .

Before closing the section, we go further in the investigation of the convergence property by considering a subclass of so-called *directed acyclic* BRGs $(V, A, \mathcal{L}_P, B, \mathcal{R})$ which require the underlying graph (V, A) not to contain any cycle:

Proposition 12 *For $k \in \{1, 2, 3, 4\}$, all directed acyclic BRGs from $\mathcal{G}(R_{\Delta}^k)$ satisfy **(Conv)** when $k = 1$ or if:*

- when $k = 2$, Δ satisfies **(IC0)** and **(IC2)**;
- when $k = 3$, Δ is a distance-based merging operator;
- when $k = 4$, Δ satisfies **(IC2)**, **(IC4)** and **(Disj)**, or Δ is a distance-based merging operator.

Related Work

Our belief revision games are somehow related to many settings where some interacting "agents" are considered, including cellular automata (Wolfram 1983), Boolean networks (Kauffman 1969; Kauffman 1993; Aldana 2003), opinion dynamics (Tsang and Larson 2014), and many complex systems (Latane and Nowak 1997; Kacpersky and Holyst 2000; Olshevsky and Tsitsiklis 2009; Bloembergen et al. 2014; Ranjbar-Sahraei et al. 2014).

In (Gauwin, Konieczny, and Marquis 2007), the authors introduce and study families of so-called iterated merging conciliation operators. Such operators are considered to rule the dynamics of the profile \mathcal{K} of belief states associated with a group of agents. At each step the state B_i of agent i is modified, by revising the merged state $\Delta(\mathcal{K})$ by B_i (skeptical approach), or by revising B_i by the merged state $\Delta(\mathcal{K})$ (credulous approach). Clearly enough, such merge-then-revise change functions are closely related to our merging-based revision policies R^2 (for the credulous one) and R^5 (for the skeptical one). They do not coincide with them nevertheless since in our approach B_i does not belong to its context \mathcal{C}_i ; clearly enough, this amounts to giving more importance to B_i when majoritarian merging operators are considered, and as a consequence the states obtained after the "revision" of B_i may differ. Notwithstanding the merging-based revision policies used, such conciliation processes correspond to specific belief revision games, where the topology is the clique one. One of the main issues considered in (Gauwin, Konieczny, and Marquis 2007) is the stationarity of the process (i.e., the convergence of the policies), which is proved in the skeptical approach. However, preservation issues, as well as responsiveness and monotonicity are not studied in (Gauwin, Konieczny, and Marquis 2007), which remains also focused on the clique topology.

Our work also is relevant to the opinion dynamics problem, which raises an abundant literature in philosophy for the last two decades. One of the most influential model to opinion dynamics is Hegselmann-Krause's one (see e.g. (Hegselmann and Krause 2005)). In the original Hegselmann-Krause's model, a set of agents aims at determining the value of a given parameter $p \in (0, 1]$. Each agent i has some belief p_i , her estimate of the right value of p . Each agent updates her belief p_i by replacing it by the average of p_i with the beliefs of its "neighbors", i.e., the set of all values p_j which are sufficiently close to p_i , i.e., $|p_i - p_j| \leq \epsilon$ where ϵ is a preset constant. Available results take the form of analytical results or of empirical results achieved using computer simulations and show the existence of diverging converging groups in the basic model. Many extensions of it have been pointed out so far, a closest one to our work being Riegler-Douven's one (Riegler and Douven 2009). Indeed, in Riegler-Douven's model, the belief states take the form of propositional theories. Proximity between belief states is evaluated as the minimal Hamming distance between their propositional models. The objective of the agents is to track the truth, which is rendered possible by incorporating at each

update step some piece of evidence e_i supposed to be true in the actual state of affairs. Update proceeds by a specific way of averaging over the "neighbors" beliefs together with the evidence. Thus, this work departs from our own one in many dimensions; mainly the way beliefs are revised, the handling of pieces of evidence, the concept of neighborhood which depends on the proximity of the belief states and the nature of the results (which mainly amounts here to determining using computer simulations for which values of the parameters used in the model the beliefs are converging to the truth).

Conclusion

In this paper, we formalized the concept of belief revision game (BRG) for modeling the dynamics of the beliefs of a group of agents. We pointed out a set of appealing properties for BRGs which address several preservation issues, as well as responsiveness, monotonicity and convergence. As a first attempt to investigate the behavior of BRGs with respect these properties, we introduced several classes of revision policies which are based on belief merging operators. We considered the case where all agents use the same revision policy and investigated the extent to which the BRGs concerned with these policies satisfy the properties. Additionally, we developed a software available online at <http://www.cril.fr/marquis/BRG.jar>. It consists of a user-friendly, graphical interface which allows one to play BRGs considering any of the 18 revision policies from $\{R_{\Delta}^k \mid k \in \{1, \dots, 6\}, \Delta \in \{\Delta^{d_D, \Sigma}, \Delta^{d_H, \Sigma}, \Delta^{d_H, Gmin}\}\}$. Some instances of BRGs are provided together with the software, including the BRG from our motivating example (Example 1) and the counter-examples used in the proofs of some propositions.

Practical applications of the BRG model are numerous. For instance, in brand crisis management, negative content regarding a brand could disseminate rapidly over social media and generate negative perceptions (Dawar and Pillutla 2000). In such a case, identifying how information is propagated within a social network and which are the influential agents (the opinion leaders) is a hot research topic. As a consequence, our general framework leaves the way open to many extensions and additional theoretical studies. Perspectives include a further investigation of the robustness of BRGs in terms of belief manipulation. For example, one could investigate how "controllable" a BRG is with respect to some piece of belief. An intuitive notion of controllability with respect to some given piece of belief would consider the minimal number of controlled agents the role of which is to make this piece of belief unanimously accepted in the BRG.

References

- [Alchourrón, Gärdenfors, and Makinson 1985] Alchourrón, C. E.; 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.
- [Aldana 2003] Aldana, M. 2003. Boolean dynamics of networks with scale-free topology. *Physica D: Nonlinear Phenomena* 185(1):45–66.
- [Bloembergen et al. 2014] Bloembergen, D.; Sahraei, B. R.; Bou-Ammar, H.; Tuyls, K.; and Weiss, G. 2014. Influencing social networks: An optimal control study. In *In Proceedings of the 21st European Conference on Artificial Intelligence*, 105–110.
- [Dawar and Pillutla 2000] Dawar, N., and Pillutla, M. 2000. Impact of product-harm crises on brand equity: the moderating role of consumer expectations. *Journal of Marketing Research* 37.
- [Everaere, Konieczny, and Marquis 2010] Everaere, P.; Konieczny, S.; and Marquis, P. 2010. Disjunctive merging: Quota and gmin merging operators. *Artificial Intelligence* 174(12-13):824–849.
- [Gauwin, Konieczny, and Marquis 2007] Gauwin, O.; Konieczny, S.; and Marquis, P. 2007. Conciliation through iterated belief merging. *Journal of Logic and Computation* 17(5):909–937.
- [Hegselmann and Krause 2005] Hegselmann, R., and Krause, U. 2005. Opinion dynamics driven by various ways of averaging. *Computational Economics* 25:381–405.
- [Kacpersky and Holyst 2000] Kacpersky, K., and Holyst, J. 2000. Phase transition as a persistent feature of groups with leaders in models of opinion formation. *Physica A* 287:631–643.
- [Katsuno and Mendelzon 1992] Katsuno, H., and Mendelzon, A. O. 1992. Propositional knowledge base revision and minimal change. *Artificial Intelligence* 52(3):263–294.
- [Kauffman 1969] Kauffman, S. A. 1969. Metabolic stability and epigenesis in randomly constructed genetic nets. *Journal of Theoretical Biology* 22(3):437–467.
- [Kauffman 1993] Kauffman, S. A. 1993. *The Origins of Order: Self-Organization and Selection in Evolution*. Oxford University Press, USA.
- [Konieczny and Pino Pérez 2002] Konieczny, S., and Pino Pérez, R. 2002. Merging information under constraints: a logical framework. *Journal of Logic and Computation* 12(5):773–808.
- [Konieczny, Lang, and Marquis 2004] Konieczny, S.; Lang, J.; and Marquis, P. 2004. DA^2 merging operators. *Artificial Intelligence* 157:49–79.
- [Latane and Nowak 1997] Latane, B., and Nowak, A. 1997. *Progress in Communication Sciences*. Ablex Publishing Corporation. chapter "Self-organizing social systems: necessary and sufficient conditions for the emergence of clustering, consolidation, and continuing diversity".
- [Olshevsky and Tsitsiklis 2009] Olshevsky, A., and Tsitsiklis, J. 2009. Convergence speed in distributed consensus

and averaging. *SIAM Journal on Control and Optimization* 48(1):33–55.

[Ranjbar-Sahraei et al. 2014] Ranjbar-Sahraei, B.; Ammar, H. B.; Bloembergen, D.; Tuyls, K.; and Weiss, G. 2014. Evolution of cooperation in arbitrary complex network. In *In Proceedings of the 13th International Conference on Autonomous Agents and Multi-Agent Systems*, 677–684.

[Riegler and Douven 2009] Riegler, A., and Douven, I. 2009. Extending the Hegselmann-Krause model III: from single beliefs to complex belief states. *Episteme* 6:145–163.

[Tsang and Larson 2014] Tsang, A., and Larson, K. 2014. Opinion dynamics of skeptical agents. In *In Proceedings of the 13th International Conference on Autonomous Agents and Multi-Agent Systems*, 277–284.

[Wolfram 1983] Wolfram, S. 1983. Statistical mechanics of cellular automata. *Reviews of Modern Physics* 55(3):601–644.

[Woodall 1997] Woodall, D. R. 1997. Monotonicity of single-seat preferential election rules. *Discrete Applied Mathematics* 77(1):81–98.