Jia Chen
Tianqun Pan

Logic for Describing Strong Belief-Disagreement Between Agents

Abstract. The result of an interaction is influenced by its epistemic state, and several epistemic notions are related to multiagent situations. Strong belief-disagreement on a certain proposition between agents means that one agent believes the proposition and the other believes its negation. This paper presents a logical system describing strong belief-disagreement between agents and demonstrates its soundness and completeness. The notion of belief-disagreement as well as belief-agreement can facilitate gaining a clearer understanding of the acts of trade and speech.

Keywords: Belief-disagreement, Doxastic possible worlds, Epistemic state, Interaction.

1. Introduction

The result of an interaction is influenced by its epistemic state. Some epistemic notions related to multiagent situations(e.g., shared knowledge and shared belief, implicit or distributed knowledge, and common knowledge and common belief) have been formally studied [1,2,6], (Fagin et al. 2003). In this paper, our interest is centered on a "negative" epistemic state among several agents, or strictly speaking, between agents: belief-disagreement.

In general, epistemic disagreement means that people have different epistemic attitudes toward a certain proposition. For example, if one person believes a proposition p and another person doubts p, then they disagree on p, regardless of whether they know each other's attitude on this proposition. Aumann [1] demonstrated an agreement theorem that indicates that if two people have the same priors and their posteriors for an event A are common knowledge, then these posteriors are equal, namely, they cannot disagree with each other. For Aumann, the posteriors as well as the priors are knowledge and quantitatively expressed in probability. Since knowledge must be true, such a disagreement that one person knows p and another person knows its negation cannot be permitted although such a disagreement is permitted

Presented by **Daniele Mundici**; *Received* January 21, 2017

Studia Logica (2018) 106: 35–47
https://doi.org/10.1007/s11225-017-9724-1

when one person knows a proposition and another person does not. However, such a disagreement is permitted that one person believes p and another believes its negation. In this paper, we concern belief-disagreement between agents.

Strong belief-disagreement among agents as well as moderate belief-disagreement among agents and weak belief-disagreement among agents was defined in Pan [10]. We, in this paper, consider two agents only and focus on strong belief-disagreement between agents. Strong belief-disagreement between agents, or strong belief-disagreement for short, means that one agent believes one proposition and the other believes its negation. Strong belief-disagreement between agents is different from moderate belief-disagreement, which means that one agent believes one proposition and the other fails to believe it, and different from weak belief-disagreement, which means that one agent fails to believe one proposition and the other fails to believe its negation. Provided that agent's beliefs are consistent, a strong belief-disagreement implies a moderate belief-disagreement, and a moderate belief-disagreement implies a weak belief-disagreement.[1] What we concern here is only the strong belief-disagreement between agents. In the following section, we present a logical system describing the strong belief-disagreement between agents and demonstrate its soundness and completeness by using possible world semantics.

2. A Logical System for Disagreement Between Agents and its Soundness and Completeness

We consider two agents involved in an interaction. A strong belief-disagreement on p between them means that one agent believes p and the other believes not-p. Let $\circledcirc$ denote a strong belief-disagreement operator, and let $\circledcirc p$ denote the existence of a strong belief-disagreement on p between the agents.

Let P be a countable set of propositional variables. The formal language L in Backus–Naur Form is defined as follows:

$$\varphi ::= p | \neg\varphi | (\varphi \wedge \varphi) | \circledcirc \varphi$$

[1] Since "An agent does not believe p" is often understood as "An agent believes not-p", we use "An agent fails to believe p" to express "p is not a belief of an agent". We can see that the notions of strong belief-disagreement and moderate disagreement are meaningful and deserve exploring. And the notion of weak belief-disagreement is too weak and trivial because a weak belief-disagreement on p could not exclude the possibility that both agents fail to believe p (or not-p) though the weak belief-disagreement works formally.

where $p \in P$. The other connectives $\vee$, $\rightarrow$ and $\leftrightarrow$ are defined in the standard way in terms of $\neg$ and $\wedge$. We employ $\top$ to stand for constant truth, and the constant falsity $\bot$ is defined as $\neg\top$.

The language in L is interpreted using the standard possible world semantics.

Definition 2.1. (*Frames, model and satisfaction*) A frame is a triple $F = <W, R_1, R_2>$ where W is a non-empty set, whose elements we will call possible worlds, and R_1 and R_2 are binary relations over W. A model is a pair $M = <F, V>$, where F is a frame, and V is a valuation function assigning a set of worlds $V(p) \subseteq W$ to each $p \in P$. A formula φ is true in M at w (written as $M, w \models \varphi$), and it is inductively defined as follows:

$M, w \models p$	iff	$w \in V(p)$;
$M, w \models \neg\varphi$	iff	$M, w \not\models \varphi$;
$M, w \models \varphi \wedge \psi$	iff	$M, w \models \varphi$ and $M, w \models \psi$;
$M, w \models \circledcirc\varphi$	iff	either, for all w' such that R_1ww', $M, w' \models \varphi$ and for all w'' such that R_2ww'', $M, w'' \models \neg\varphi$; or, for all w' such that R_1ww', $M, w' \models \neg\varphi$ and for all w'' such that R_2ww'', $M, w'' \models \varphi$.

The other connectives are evaluated as expected.

Two notes regarding Definition 2.1 are as follows:

(1) Two accessibility relations R_1, R_2, which respectively represent doxastic accessibility relations of agents 1 and 2, are involved in Definition 2.1. Because we do not want to make any special assumptions about how either of our agents' beliefs are related to the other's, we do not impose any conditions on how the accessibility relations R_1, R_2 in these frames are related to each other.

(2) W in a frame $F = <W, R_1, R_2>$ is the set of doxastically possible worlds, and V in $M = <F, V>$ assigns a set of worlds to each propositional variables, with $V(p)$ to be thought as the set of worlds at which the model stipulates that p is true. If an agent believes p at a world w, then p is true at all worlds doxastically accessible (for that agent: i.e., related by R_1 or R_2, as appropriate) to w. If two agents have a common doxastically accessible world, they have no disagreement on any proposition, indicating that $\circledcirc p$ must be false, while if they have no such common accessible world, this leaves open the question of whether they are in disagreement on any proposition.

Definition 2.2. A system SD is axiomatized as follows:

Ax.0 All substitution instances of tautologies
Ax.1 $\circledcirc\varphi \rightarrow \circledcirc\neg\varphi$
Ax.2 $\circledcirc\varphi \wedge \circledcirc\psi \rightarrow \circledcirc(\varphi \wedge \psi) \vee \circledcirc(\varphi \wedge \neg\psi)$
Ax.3 $\circledcirc\varphi \wedge \circledcirc(\varphi \wedge \psi) \wedge \circledcirc(\varphi \wedge \chi) \rightarrow \circledcirc(\varphi \wedge \psi \wedge \chi)$
Ax.4 $\neg\circledcirc\top$
MP $\vdash_{SD} \varphi, \vdash_{SD} \varphi \rightarrow \psi \Rightarrow \vdash_{SD} \psi$
RS $\vdash_{SD} \varphi \rightarrow \psi, \vdash_{SD} \psi \rightarrow \chi \Rightarrow \vdash_{SD} \circledcirc\varphi \wedge \circledcirc\chi \rightarrow \circledcirc\psi$

In what follows, $\vdash_{SD} \varphi$ is simply expressed as $\vdash \varphi$.

Definition 2.3. (*Serial frame*) A frame $F = <W, R_1, R_2>$ is serial iff for any $w \in W$, there exist w' and w'' such that wR_1w' and wR_2w''.

In Definition 2.3, w' and w'' may be identical. Validity in serial frames is denoted as $\models_s \varphi$.

Because consistency is a basic requirement for an agent's beliefs, no agent can believe constant falsity. Therefore, the logic system SD should be established on the basis of arbitrary serial frames. Specifically, Ax.4, $\neg\circledcirc\top$, is valid on arbitrary serial frames, which is shown in Theorem 2.1. Moreover, if the frame is not serial, $\neg\circledcirc\top$ is false at a dead end.

Notably, because two accessibility relations are involved in $F =< W, R_1, R_2 >$, the seriality defined in Definition 2.3 is "strong serial": For any $w \in W$, there exist w' and w'' as wR_1w' and wR_2w''. This seriality is different from a "weak seriality": For any $w \in W$, there exists w' either wR_1w' or wR_2w'.

Theorem 2.1. (Soundness) *The system SD is sound with respect to the class of serial frames.*

Proof. We need to show that in the class of serial frames, all axioms are valid and all the transformation rules preserve validity. Ax.0 and MP are obviously valid. Here, we only prove that Ax.1–4 are valid, and RS preserves validity in the class of serial frames.

Let M be an arbitrary model based on a serial frame and w be a world of M.

For Ax.1, suppose that $M, w \models \circledcirc\varphi$. According to Definition 2.1, we have that either $\forall_{w',w'' \in W}((R_1ww' \Rightarrow M, w' \models \varphi)$ and $(R_2ww'' \Rightarrow M, w'' \models \neg\varphi))$, or $\forall_{w',w'' \in W}((R_1ww' \Rightarrow M, w' \models \neg\varphi)$ and $(R_2ww'' \Rightarrow M, w'' \models \varphi))$. Because $\models \varphi \leftrightarrow \neg\neg\varphi$, we have either $\forall_{w',w'' \in W}((R_1ww' \Rightarrow M, w' \models \neg\neg\varphi)$ and $(R_2ww'' \Rightarrow M, w'' \models \neg\varphi))$, or $\forall_{w',w'' \in W}((R_1ww' \Rightarrow M, w' \models \neg\varphi)$ and

$(R_2ww'' \Rightarrow M, w'' \models \neg\neg\varphi))$. Hence, according to Definition 2.1, we have $M, w \models \circledcirc\neg\varphi$.

For Ax.2, suppose that $M, w \models \circledcirc\varphi \wedge \circledcirc\psi$. Without loss of generality, we assume that $\forall_{w',w'' \in W}$ $((R_1ww' \Rightarrow M, w' \models \varphi)$ and $(R_2ww'' \Rightarrow M, w'' \models \neg\varphi))$. Therefore, either $\forall_{w',w'' \in W}((R_1ww' \Rightarrow M, w' \models \varphi \wedge \psi)$ and $(R_2ww'' \Rightarrow M, w'' \models \neg\varphi \wedge \neg\psi))$, or $\forall_{w',w'' \in W}((R_1ww' \Rightarrow M, w' \models \varphi \wedge \neg\psi)$ and $(R_2ww'' \Rightarrow M, w'' \models \neg\varphi \wedge \psi))$. Hence, in the former case, $M, w \models \circledcirc(\varphi \wedge \psi)$, and in the latter case, $M, w \models \circledcirc(\varphi \wedge \neg\psi)$. We thus derive $M, w \models \circledcirc(\varphi \wedge \psi) \vee \circledcirc(\varphi \wedge \neg\psi)$.

For Ax.3, suppose that $M, w \models \circledcirc\varphi \wedge \circledcirc(\varphi \wedge \psi) \wedge \circledcirc(\varphi \wedge \chi)$. Hence, $M, w \models \circledcirc\varphi$, $M, w \models \circledcirc(\varphi \wedge \psi)$ and $M, w \models \circledcirc(\varphi \wedge \chi)$. Without loss of generality, we assume that $\forall_{w',w'' \in W}((R_1ww' \Rightarrow M, w' \models \varphi)$ and $(R_2ww'' \Rightarrow M, w'' \models \neg\varphi))$. Because $M, w \models \circledcirc(\varphi \wedge \psi)$, either $\forall_{w' \in W}(R_1ww' \Rightarrow M, w' \models \varphi \wedge \psi)$ or $\forall_{w' \in W}(R_2ww' \Rightarrow M, w' \models \varphi \wedge \psi)$. Nevertheless, because $\forall_{w' \in W}(R_2ww' \Rightarrow M, w' \models \neg\varphi)$ and M is a serial frame, $\forall_{w' \in W}(R_2ww' \Rightarrow M, w' \models \varphi \wedge \psi)$ does not hold. Therefore, $\forall_{w' \in W}(R_1ww' \Rightarrow M, w' \models \varphi \wedge \psi)$. Similarly, we derive that $\forall_{w' \in W}(R_1ww' \Rightarrow M, w' \models \varphi \wedge \chi)$. Hence, we have $\forall_{w' \in W}(R_1ww' \Rightarrow M, w' \models \varphi \wedge \psi \wedge \chi)$. However, since $\forall_{w' \in W}(R_2ww' \Rightarrow M, w' \models \neg\varphi)$, $\forall_{w' \in W}(R_2ww' \Rightarrow M, w' \models \neg(\varphi \wedge \psi \wedge \chi))$. Therefore, as desired, $M, w \models \circledcirc(\varphi \wedge \psi \wedge \chi)$.

For Ax.4, suppose that $M, w \models \circledcirc\top$. Hence, either $\forall_{w' \in W}(R_1ww' \Rightarrow M, w' \models \neg\top)$ or $\forall_{w' \in W}(R_2ww' \Rightarrow M, w \models \neg\top)$. This means that there would be no w' with R_1ww' or no w' with R_2ww', which contradicts the fact that M is serial.

For RS, suppose $\models_s \varphi \to \psi$, $\models_s \psi \to \chi$ and $M, w \models \circledcirc\varphi \wedge \circledcirc\chi$. From $M, w \models \circledcirc\varphi \wedge \circledcirc\chi$, we have $M, w \models \circledcirc\varphi$ and $M, w \models \circledcirc\chi$. Without loss of generality, we assume that $\forall_{w',w'' \in W}((R_1ww' \Rightarrow M, w' \models \varphi)$ and $(R_2ww'' \Rightarrow M, w'' \models \neg\varphi))$. Since $\forall_{w' \in W}(R_1ww' \Rightarrow M, w' \models \varphi)$, $\models_s \varphi \to \psi$ and $\models_s \psi \to \chi$ we get $\forall_{w' \in W}(R_1ww' \Rightarrow M, w' \models \psi)$ and $\forall_{w' \in W}(R_1ww' \Rightarrow M, w' \models \chi)$. Hence, since M is based on a serial frame, it is not the case that $\forall_{w' \in W}(R_1ww' \Rightarrow M, w' \models \neg\chi)$. Then, since $M, w \models \circledcirc\chi$, $\forall_{w'' \in W}(R_2ww'' \Rightarrow M, w'' \models \neg\chi)$ must be true. From $\forall_{w'' \in W}(R_2ww'' \Rightarrow M, w'' \models \neg\chi)$ and $\models_s \psi \to \chi$ we get $\forall_{w'' \in W}(R_2ww'' \Rightarrow M, w'' \models \neg\psi)$. Hence, we have $\forall_{w' \in W}(R_1ww' \Rightarrow M, w' \models \psi)$ and $\forall_{w'' \in W}(R_2ww'' \Rightarrow M, w'' \models \neg\psi)$. Therefore, $M, w \models \circledcirc\psi$. ■

Theorem 2.2. *In SD we have:*

(1) $\vdash \varphi \leftrightarrow \psi \Rightarrow \vdash \circledcirc\varphi \leftrightarrow \circledcirc\psi$

(2) $\vdash \circledcirc\varphi \leftrightarrow \circledcirc\neg\varphi$

(3) $\vdash \circledcirc(\varphi \wedge \psi) \wedge \circledcirc(\varphi \vee \chi) \rightarrow \circledcirc\varphi$

(4) $\vdash \circledcirc\varphi \wedge \circledcirc(\varphi \wedge \psi_0) \wedge \cdots \wedge \circledcirc(\varphi \wedge \psi_n) \rightarrow \circledcirc(\varphi \wedge \psi_0 \cdots \wedge \psi_n)$

(5) $\vdash \varphi \Rightarrow \vdash \neg \circledcirc \varphi$

(6) $\vdash \varphi \rightarrow \psi \Rightarrow \vdash \circledcirc\varphi \leftrightarrow \circledcirc(\varphi \wedge \psi)$

We do not provide the poof of Theorem 2.2 in this paper.

Notably, if we drop Ax.4 in SD and add $\vdash \varphi \Rightarrow \vdash \neg \circledcirc \varphi$ [Theorem 2.2, (5)] as an inference rule, the resulting system is deductively equivalent to SD.

To demonstrate the completeness of system SD, we use the canonical model method, which requires us first to define what is meant by a maximally consistent set of formulae (with respect to this system).

Given a logic system S, a set of formulae Γ is maximally S-consistent iff Γ is S-consistent, and for any set of formulae Λ, if $\Gamma \subset \Lambda$, then Λ is S-inconsistent.

In the following, we focus on deriving a maximally SD-consistent set of formulae. Lindenbaum's lemma holds for SD, and it indicates that every SD-consistent set can be extended to a maximally SD-consistent set. According to the properties of maximally consistent sets and Theorem 2.2, we have the following:

LEMMA 2.3. *Let Γ be a maximally SD-consistent set. Therefore, for any formula φ, ψ:*

(1) $\circledcirc\varphi \in \Gamma$ *iff* $\circledcirc\neg\varphi \in \Gamma$, *and*

(2) *if* $\vdash \varphi \leftrightarrow \psi$, *then* $\circledcirc\varphi \in \Gamma$ *iff* $\circledcirc\psi \in \Gamma$.

DEFINITION 2.4. Let Γ be a maximally SD-consistent set. If there exists no formula α such that $\circledcirc\alpha \in \Gamma$, then we define $S(\Gamma)$ as the empty set, otherwise we enumerate the elements of $\{\alpha | \circledcirc \alpha \in \Gamma\}$ and define $S(\Gamma)$ as $\{\varphi | \circledcirc \varphi, \circledcirc(\theta \wedge \varphi) \in \Gamma\}$,where θ is the first element in the enumeration of $\{\alpha | \circledcirc \alpha \in \Gamma\}$.

LEMMA 2.4. *Let Γ be a maximal SD-consistent set. Therefore,*

(1) $S(\Gamma)$ *is SD-consistent,*

(2) $\varphi \in S(\Gamma)$, $\circledcirc(\varphi \wedge \psi) \in \Gamma \Rightarrow \varphi \wedge \psi \in S(\Gamma)$,

(3) $\circledcirc\varphi \in \Gamma \Rightarrow \varphi \in S(\Gamma)$ *or* $\neg\varphi \in S(\Gamma)$,

(4) $\varphi \in S(\Gamma)$, $\vdash \varphi \leftrightarrow \psi \Rightarrow \psi \in S(\Gamma)$,

(5) $\varphi \in S(\Gamma)$, $\psi \in S(\Gamma) \Rightarrow \varphi \wedge \psi \in S(\Gamma)$,

(6) $\varphi \in S(\Gamma)$, $\psi \in S(\Gamma) \Rightarrow \varphi \vee \psi \in S(\Gamma)$.

PROOF. If $\{\alpha | \circledcirc \alpha \in \Gamma\} = \emptyset$, then $S(\Gamma) = \emptyset$, and (1)–(6) are trivially true. Next, we treat the case of $\{\alpha | \circledcirc \alpha \in \Gamma\} \neq \emptyset$.

Let Γ be a maximally SD-consistent set with $\{\alpha | \circledcirc \alpha \in \Gamma\} \neq \emptyset$. According to Definition 2.4, $S(\Gamma)$ is $\{\varphi | \circledcirc \varphi, \circledcirc(\theta \wedge \varphi) \in \Gamma\}$, where θ is the first element in the enumeration of $\{\alpha | \circledcirc \alpha \in \Gamma\}$.

For (1), suppose that $S(\Gamma)$ is SD-inconsistent. Therefore, there exist $\varphi_0, \ldots, \varphi_n \in S(\Gamma)$ such that $\vdash \neg(\varphi_0 \wedge \cdots \wedge \varphi_n)$. In view of Theorem 2.2 (5) and the fact that $\vdash \neg(\varphi_0 \wedge \cdots \wedge \varphi_n)$, we conclude that $\vdash \neg\circledcirc\neg(\theta \wedge \varphi_0 \wedge \cdots \wedge \varphi_n)$. Hence, according to Theorem 2.2 (2), $\vdash \circledcirc(\theta \wedge \varphi_0 \wedge \cdots \wedge \varphi_n) \leftrightarrow \circledcirc\neg(\theta \wedge \varphi_0 \wedge \cdots \wedge \varphi_n))$, we obtain $\vdash \neg\circledcirc(\theta \wedge \varphi_0 \wedge \cdots \wedge \varphi_n)$. This means that $\neg\circledcirc(\theta \wedge \varphi_0 \wedge \cdots \wedge \varphi_n)$ must be contained in the maximally SD-consistent set Γ. Hence, we have (a): $\circledcirc(\theta \wedge \varphi_0 \wedge \cdots \wedge \varphi_n) \notin \Gamma$. Moreover, since $\varphi_0, \ldots, \varphi_n \in S(\Gamma)$, we obtain $\circledcirc(\theta \wedge \varphi_0), \ldots, \circledcirc(\theta \wedge \varphi_n) \in \Gamma$; therefore, $\circledcirc\theta \wedge \circledcirc(\theta \wedge \varphi_0) \wedge \cdots \wedge \circledcirc(\theta \wedge \varphi_n) \in \Gamma$. By (4) in Theorem 2.2, $\vdash \circledcirc\theta \wedge \circledcirc(\theta \wedge \varphi_0) \wedge \cdots \wedge \circledcirc(\theta \wedge \varphi_n) \to \circledcirc(\theta \wedge \varphi_0 \wedge \cdots \wedge \varphi_n)$ holds. Because Γ is maximally SD-consistent, $\circledcirc(\theta \wedge \varphi_0 \wedge \cdots \wedge \varphi_n) \in \Gamma$, which contradicts (a).

For (2), suppose that $\varphi \in S(\Gamma)$ and $\circledcirc(\varphi \wedge \psi) \in \Gamma$. Through $\varphi \in S(\Gamma)$ and the definition of $S(\Gamma)$, we have $\circledcirc\varphi$, $\circledcirc(\theta \wedge \varphi) \in \Gamma$. Next, according to (2) in Lemma 2.3, $\circledcirc(\varphi \wedge \theta) \in \Gamma$. Hence, $\circledcirc\varphi$, $\circledcirc(\varphi \wedge \theta)$, $\circledcirc(\varphi \wedge \psi) \in \Gamma$ and consequently, (*) $\circledcirc\varphi \wedge \circledcirc(\varphi \wedge \theta) \wedge \circledcirc(\varphi \wedge \psi) \in \Gamma$. Through (*) and Ax.3 (i.e., $\circledcirc\varphi \wedge \circledcirc(\varphi \wedge \theta) \wedge \circledcirc(\varphi \wedge \psi) \to \circledcirc(\varphi \wedge \theta \wedge \psi)))$, we have $\circledcirc(\varphi \wedge \theta \wedge \psi) \in \Gamma$. Moreover, according to (2) in Lemma 2.3, $\circledcirc(\theta \wedge \varphi \wedge \psi) \in \Gamma$. Because $\circledcirc(\varphi \wedge \psi) \in \Gamma$ and $\circledcirc(\theta \wedge \varphi \wedge \psi) \in \Gamma$, as desired, $\varphi \wedge \psi \in S(\Gamma)$.

For (3), suppose that $\circledcirc\varphi \in \Gamma$. Through $\circledcirc\theta \in \Gamma$, $\circledcirc\varphi \in \Gamma$, and Ax.2 (i.e., $\circledcirc\theta \wedge \circledcirc\varphi \to \circledcirc(\theta \wedge \varphi) \vee \circledcirc(\theta \wedge \neg\varphi)$), we have $\circledcirc(\theta \wedge \varphi) \vee \circledcirc(\theta \wedge \neg\varphi) \in \Gamma$, implying that either $\circledcirc(\theta \wedge \varphi) \in \Gamma$ or $\circledcirc(\theta \wedge \neg\varphi) \in \Gamma$. According to $\circledcirc\varphi \in \Gamma$ and (1) in Lemma 2.3, $\circledcirc\neg\varphi \in \Gamma$. Therefore, from $\circledcirc\varphi \in \Gamma$ and $\circledcirc(\theta \wedge \varphi) \in \Gamma$, $\varphi \in S(\Gamma)$, and from $\circledcirc\neg\varphi \in \Gamma$ and $\circledcirc(\theta \wedge \neg\varphi) \in \Gamma$, $\neg\varphi \in S(\Gamma)$. Hence, either $\varphi \in S(\Gamma)$ or $\neg\varphi \in S(\Gamma)$.

For (4), suppose that $\varphi \in S(\Gamma)$ and $\vdash \varphi \leftrightarrow \psi$. Through $\varphi \in S(\Gamma)$, we have $\circledcirc\varphi \in \Gamma$ and $\circledcirc(\theta \wedge \varphi) \in \Gamma$. Therefore, by Theoreme 2.2 (1), $\circledcirc\psi \in \Gamma$ and $\circledcirc(\theta \wedge \psi) \in \Gamma$, indicating that $\psi \in S(\Gamma)$, as desired.

For (5), suppose that $\varphi \in S(\Gamma)$ and $\psi \in S(\Gamma)$. Consequently, through the definition of $S(\Gamma)$, we derive $\circledcirc\varphi \in \Gamma$ and $\circledcirc\psi \in \Gamma$; hence, $\circledcirc\varphi \wedge \circledcirc\psi \in \Gamma$. Further, by Ax.2 (and MP), $\circledcirc(\varphi \wedge \psi) \vee \circledcirc(\varphi \wedge \neg\psi) \in \Gamma$. Therefore, either $\circledcirc(\varphi \wedge \psi) \in \Gamma$ or $\circledcirc(\varphi \wedge \neg\psi) \in \Gamma$. Nevertheless, if $\circledcirc(\varphi \wedge \neg\psi) \in \Gamma$, then by (2) in Lemma 2.4, which we just proved, $\varphi \wedge \neg\psi \in S(\Gamma)$ and this, together with $\psi \in S(\Gamma)$, renders $S(\Gamma)$ SD-inconsistent, thus contradicting (1) in Lemma 2.4. Consequently, $\circledcirc(\varphi \wedge \psi) \in \Gamma$, which, together with (2) in Lemma 2.4 and $\varphi \in S(\Gamma)$, implies $\varphi \wedge \psi \in S(\Gamma)$.

For (6), suppose that $\varphi \in S(\Gamma)$ and $\psi \in S(\Gamma)$. Consequently, through the definition of $S(\Gamma)$, we have $\circledcirc\varphi \in \Gamma$ and $\circledcirc\psi \in \Gamma$, which, according to (1) in Lemma 2.3, implies that $\circledcirc\neg\varphi \in \Gamma$ and $\circledcirc\neg\psi \in \Gamma$. Through $\circledcirc\neg\varphi \in \Gamma$, $\circledcirc\neg\psi \in \Gamma$ and Ax.2, we obtain $\circledcirc(\neg\varphi \wedge \neg\psi) \vee \circledcirc(\neg\varphi \wedge \neg\neg\psi) \in \Gamma$, implying that either $\circledcirc(\neg\varphi \wedge \neg\psi) \in \Gamma$ or $\circledcirc(\neg\varphi \wedge \neg\neg\psi) \in \Gamma$. However, we claim that $\circledcirc(\neg\varphi \wedge \neg\neg\psi) \notin \Gamma$. Otherwise, through Lemma 2.3 (2) (from $\circledcirc(\neg\varphi \wedge \neg\neg\psi) \in \Gamma$), we derive $\circledcirc(\psi \wedge \neg\varphi) \in \Gamma$, and this, together with (2) in Lemma 2.4, results in $\psi \wedge \neg\varphi \in S(\Gamma)$, implying that $S(\Gamma)$ is SD-inconsistent, which contradicts the conclusion of (1) in Lemma 2.4. Hence, we derive $\circledcirc(\neg\varphi \wedge \neg\psi) \in \Gamma$. By Lemma 2.4 (3), either $\neg\varphi \wedge \neg\psi \in S(\Gamma)$ or $\neg(\neg\varphi \wedge \neg\psi) \in S(\Gamma)$. However, $\neg\varphi \wedge \neg\psi \notin S(\Gamma)$, otherwise $S(\Gamma)$ is SD-inconsistent; consequently, $\neg(\neg\varphi \wedge \neg\psi) \in S(\Gamma)$. By (4) in Lemma 2.4, because $\neg(\neg\varphi \wedge \neg\psi) \in S(\Gamma)$ and $\vdash \neg(\neg\varphi \wedge \neg\psi) \leftrightarrow (\varphi \vee \psi)$, we obtain $\varphi \vee \psi \in S(\Gamma)$, as desired. ■

Note that (5) and (6) in Lemma 2.4 mean that $S(\Gamma)$ is closed under conjunction and disjunction.

Definition 2.5. Let Γ be a maximally SD-consistent set. $S(\Gamma)^-$ is defined as $\{\varphi | \neg\varphi \in S(\Gamma)\}$.

Lemma 2.5. *Let Γ be a maximally SD-consistent set. Therefore,*

(1) *for all φ, $\varphi \in S(\Gamma) \Leftrightarrow \neg\varphi \in S(\Gamma)^-$, and $\neg\varphi \in S(\Gamma) \Leftrightarrow \varphi \in S(\Gamma)^-$, and*

(2) *$S(\Gamma)^-$ is SD-consistent.*

Proof. Lemma 2.5 (1) obviously holds.

For (2), suppose that $S(\Gamma)^-$ is not SD-consistent. Hence, there exist $\varphi_0, \ldots, \varphi_n \in S(\Gamma)^-$ such that $\vdash \neg\varphi_0 \vee \cdots \vee \neg\varphi_n$. Through (5) in Theorem 2.2 and $\vdash \neg\varphi_0 \vee \cdots \vee \neg\varphi_n$, we derive $\vdash \neg \circledcirc (\neg\varphi_0 \vee \cdots \vee \neg\varphi_n)$. Because Γ is maximally SD-consistent, $\neg \circledcirc (\neg\varphi_0 \vee \cdots \vee \neg\varphi_n) \in \Gamma$. Therefore, $\circledcirc(\neg\varphi_0 \vee \cdots \vee \neg\varphi_n) \notin \Gamma$, implying that (*) $\neg\varphi_0 \vee \cdots \vee \neg\varphi_n \notin S(\Gamma)$. However, according to $\varphi_0, \ldots, \varphi_n \in S(\Gamma)^-$ and Lemma 2.5 (1), $\neg\varphi_0, \ldots, \neg\varphi_n \in S(\Gamma)$. Consequently, since $S(\Gamma)$ is closed under disjunction, $\neg\varphi_0 \vee \cdots \vee \neg\varphi_n \in S(\Gamma)$, which contradicts (*).

Definition 2.6. (*Canonical model*) The canonical model M^C for SD is a 4-tuple $< W^C, R_1^C, R_2^C, V^C >$, where:

$W^C = \{w | w$ is a maximally SD-consistent set of formulae$\}$,
$R_1^C ww'$ iff $S(w) \subseteq w'$,
$R_2^C ww'$ iff $S(w)^- \subseteq w'$, and
$V^C(p) = \{w \in W^C | p \in w\}$.

Theorem 2.6. *M^C is based on a serial frame.*

PROOF. According to (1) in Lemma 2.4 and (2) in Lemma 2.5, for any $w \in W^C$, both $S(w)$ and $S(w)^-$ are SD-consistent. Hence, there exist w', $w'' \in W^C$ such that $S(w) \subseteq w'$ and $S(w)^- \subseteq w''$. Therefore, according to Definition 2.6, for any $w \in W^C$ there exist w', $w'' \in W^C$ such that $R_1^C ww'$ and $R_2^C ww''$, thus demonstrating that M^C is actually based on a serial frame. ■

LEMMA 2.7. (Truth lemma) *Let $M^C =< W^C, R_1^C, R_2^C, V^C >$ be the canonical model for system SD. For any formula α and any $w \in W^C$, $M^C, w \models \alpha$ iff $\alpha \in w$.*

PROOF. We prove the lemma through induction on the structure of α. Here, we prove only the case of $\circledcirc\varphi$.

From right to left: Assume that $\circledcirc\varphi \in w$. According to (3) in Lemma 2.4, two possible cases exist: $\varphi \in S(w)$ or $\neg\varphi \in S(w)$. We demonstrate that $M^C, w \models \circledcirc\varphi$ in the two possibilities.

(1) $\varphi \in S(w)$. According to (1) in Lemma 2.5, $\neg\varphi \in S(w)^-$. Therefore, $\forall_{w',w'' \in W^C}((S(w) \subseteq w' \Rightarrow \varphi \in w')$ and $(S(w)^- \subseteq w'' \Rightarrow \neg\varphi \in w''))$, and consequently, $\forall_{w',w'' \in W^C}((R_1^C ww' \Rightarrow \varphi \in w')$ and $(R_2^C ww'' \Rightarrow \neg\varphi \in w''))$. Through the induction hypothesis, we determine that $\forall_{w',w'' \in W^C}$ $((R_1^C ww' \Rightarrow M^C, w' \models \varphi)$ and $(R_2^C ww'' \Rightarrow M^C, w'' \models \neg\varphi))$, implying that $M^C, w \models \circledcirc\varphi$, as desired.

(2) $\neg\varphi \in S(w)$. Similar to (1), we determine that $\forall_{w',w'' \in W^C}((R_1^C ww' \Rightarrow M^C, w' \models \neg\varphi)$ and $(R_2^C ww'' \Rightarrow M^C, w'' \models \varphi))$, also implying that $M^C, w \models \circledcirc\varphi$.

From left to right: Assume that $M^C, w \models \circledcirc\varphi$. According to Definition 2.1, two possible cases also exist.

(1) $\forall_{w',w'' \in W^C}((R_1^C ww' \Rightarrow M^C \models \varphi)$ and $(R_2^C ww'' \Rightarrow M^C, w'' \models \neg\varphi))$. Therefore, $\forall_{w',w'' \in W^C}((S(w) \subseteq w' \Rightarrow M^C, w' \models \varphi)$ and $(S(w)^- \subseteq w'' \Rightarrow M^C, w'' \models \neg\varphi))$. Through induction hypothesis, we derive $\forall_{w' \in W^C}(S(w) \subseteq w' \Rightarrow \varphi \in w')$ and $\forall_{w' \in W^C}(S(w)^- \subseteq w' \Rightarrow \neg\varphi \in w')$. Hence, there is no such a w that is a maximally SD-consistent set of formulae and contains $S(w) \cup \{\neg\varphi\}$ or $S(w)^- \cup \{\varphi\}$. According to Lindenbaum's lemma, both $S(w) \cup \{\neg\varphi\}$ and $S(w)^- \cup \{\varphi\}$ are SD-inconsistent. Because $S(w) \cup \{\neg\varphi\}$ is SD-inconsistent, there exist $\psi_0, \ldots, \psi_n \in S(w)$ such that $\vdash \psi_0 \wedge \cdots \psi_n \rightarrow \varphi$. According to (6) in Theorem 2.2, $\vdash \circledcirc(\psi_0 \wedge \cdots \psi_n) \rightarrow \circledcirc(\psi_0 \wedge \cdots \psi_n \wedge \varphi)$. Moreover, through $\psi_0, \ldots, \psi_n \in S(w)$, we derive $\psi_0 \wedge \cdots \wedge \psi_n \in S(w)$, implying that $\circledcirc(\psi_0 \wedge \cdots \psi_n) \in w$. Therefore, $\circledcirc(\psi_0 \wedge \cdots \psi_n \wedge \varphi) \in w$. According to (2) in Lemma 2.3, $\circledcirc(\varphi \wedge \psi_0 \wedge \cdots \psi_n) \in w$. In addition, because $S(w)^- \cup \{\varphi\}$ is SD-inconsistent, there exist $\chi_0, \ldots, \chi_m \in S(w)^-$ such that $\vdash \chi_0 \wedge \cdots \wedge \chi_m \rightarrow \neg\varphi$. Through (6) in Theorem 2.2, we determine that

$\vdash \circledcirc(\chi_0 \wedge \cdots \wedge \chi_m) \to \circledcirc(\chi_0 \wedge \cdots \wedge \chi_m \wedge \neg\varphi)$. According to Lemma 2.5 (1) and $\chi_0, \ldots, \chi_m \in S(w)^-$, we obtain $\neg\chi_0, \ldots, \neg\chi_m \in S(w)$, implying $\neg\chi_0 \vee \cdots \vee \neg\chi_m \in S(w)$; hence, $\circledcirc(\neg\chi_0 \vee \cdots \vee \neg\chi_m) \in w$. On the basis of $\circledcirc(\neg\chi_0 \vee \cdots \vee \neg\chi_m) \in w$, we can easily conclude that $\circledcirc(\chi_0 \wedge \cdots \wedge \chi_m) \in w$, which, together with $\vdash \circledcirc(\chi_0 \wedge \cdots \wedge \chi_m) \to \circledcirc(\chi_0 \wedge \cdots \wedge \chi_m \wedge \neg\varphi)$, implies that $\circledcirc(\chi_0 \wedge \cdots \wedge \chi_m \wedge \neg\varphi) \in w$. Consequently, $\circledcirc\neg(\chi_0 \wedge \cdots \wedge \chi_m \wedge \neg\varphi) \in w$, which, in addition to $\vdash \neg(\chi_0 \wedge \cdots \wedge \chi_m \wedge \neg\varphi) \leftrightarrow (\varphi \vee \neg\chi_0 \vee \cdots \vee \neg\chi_m)$ and (2) in Lemma 2.3, implies that $\circledcirc(\varphi \vee \neg\chi_0 \vee \cdots \vee \neg\chi_m) \in w$. According to $\circledcirc(\varphi \wedge \psi_0 \wedge \cdots \psi_n) \in w$, $\circledcirc(\varphi \vee \neg\chi_0 \vee \cdots \vee \neg\chi_m) \in w$ and Theorem 2.2 (3) (i.e., $\vdash \circledcirc(\varphi \wedge \psi_0 \wedge \cdots \psi_n) \wedge \circledcirc(\varphi \vee \neg\chi_0 \vee \cdots \vee \neg\chi_m) \to \circledcirc\varphi$), it immediately follows that $\circledcirc\varphi \in w$, as desired.

(2) $\forall_{w',w'' \in W^C}((R_1^C ww' \Rightarrow M^C \models \neg\varphi)$ and $(R_2^C ww'' \Rightarrow M^C, w'' \models \varphi))$. Hence, similar to (1), we obtain $\circledcirc\neg\varphi \in w$. Therefore, through (1) in Lemma 2.3, we also obtain $\circledcirc\varphi \in w$. ∎

THEOREM 2.8. (Completeness) *SD is complete with respect to the class of serial frames.*

PROOF. We demonstrate that every SD-consistent set is satisfiable in the class of serial frames. Let Γ be a SD-consistent set. Therefore, through Lindenbaum's lemma, we can extend Γ to a maximally SD-consistent set (e.g., s). According to Lemma 2.7, for any formula φ, $M^C, s \models \varphi$ iff $\varphi \in s$. Therefore, Γ is satisfied in M^C, which, according to Lemma 2.6, is based on a serial frame. ∎

3. Remarks

We presented a logical system for strong belief-disagreement between agents which describes rational relations on strong belief-disagreement, and demonstrated its soundness and completeness. We didn't explore how a strong belief-disagreement is eliminated or maintained by agents. We have five additional remarks on our work.

First, strong belief-disagreement as well as moderate belief-disagreement and weak belief-disagreement is an epistemic notion describing an epistemic state between agents. When belief-disagreement as an epistemic state occurs between agents, it is a nonphysical but objective state, which can be known by the agents involved and by other agents. And we didn't presuppose that each of the agents involved must know the belief-disagreement. An agent can reflect and know his or her own beliefs; however, if he or she wants to know another agent's beliefs, he or she must engage in observation or communication. An agent knows a belief-disagreement between him or her and another agent only if he or she knows the other agent's beliefs. Therefore, a belief-disagreement can exist either implicitly or explicitly, and if it exists explicitly, it can exist in the form of shared knowledge or belief, or common knowledge or belief.

Second, although strong belief-disagreement can be expressed by introducing the definition $\circledcirc p =_{def} (B_1 p \wedge B_2 \neg p) \vee (B_1 \neg p \wedge B_2 p)$ to doxastic logics of two agents, the system SD focuses on those valid formulas of strong belief-disagreement, and 'ignores' those doxastic formulas which is not related to strong belief-disagreement. The method we followed to show SD's soundness and completeness is provided by logicians who dealt with such nonstandard modal notions as noncontingency and accident [3–5, 7–9, 11–13].

Third, the system SD is different from logics of noncontingency not because noncontingency is a metaphysical notion while belief-disagreement is an epistemic one. Rather,the system SD is semantically different from logics of noncontingecy. Two accessibility relations are involved in the semantic definition of belief-disagreement while only one accessibility relation is involved in the definition of noncontingency. This means that we cannot establish the logic(s) of belief-disagreement by translating logics of noncontingency.

Forth, the notion of belief-disagreement is helpful in understanding the act of speech and trade. People typically aspire to persuade others, and belief-disagreements can be dissolved sometimes, though not always, through communication. It is natural to assume that each party to a trade believes that he or she will benefit by it. Hence, every trade is always accompanied by belief-agreement or belief-disagreement between agents. A win-win trade originates from belief-agreement between agents, whereas a zero-sum trade originates from belief-disagreement between agents. Each trader in a win-win trade believes that both benefit from the win-win trade, while in a zero-sum trade, no matter whether the trade would be completed, each trader believes that he or she would benefit from the trade and the other's interests would be impaired in the trade.

Fifth, the SD logical system is applied for only two agents. This paper mentions but does not explain the general definitions of strong belief-disagreement for instances involving more than two agents (i.e., a group). Strong belief-disagreement associated with a group is direct in that one agent believes p and another believes not-p. Indirectly strong belief-disagreement is that in which a certain proposition p and its negation are not the beliefs of two agents in a group; however, if we collect the beliefs of each agent in a group, we can logically infer a contradiction from the collection.

Consider, for example, A believes p, B believes that p implies q, and C believes not-q. Although there is no directly strong belief-disagreement on p, q, or other complex propositions among A, B, and C, we can, according to the beliefs of any two agents, infer a contradiction with the belief of the third agent. Therefore, indirectly belief-disagreement exists among A, B, and C.

Acknowledgements. The authors are grateful to the anonymous referee(s) for detailed and careful remarks, as well as useful suggestions, that greatly helped us to improve the paper.

References

[1] Aumann, R. J., Agreeing to disagree, *The Annals of Statistics* 6: 1236–1239, 1976.

[2] Fagin, R., Y. Moses, J. Y. Halpern and M. Y. Vardi, *Reasoning about knowledge*, MIT Press, Cambridge, Mass., and London, England, 2003.

[3] Humberstone, I. L., The logic of non-contingency, *Notre Dame Journal of Formal Logic* 36: 214–229, 1995.

[4] Humberstone, I. L., The modal logic of agreement and noncontingency, *Notre Dame Journal of Formal Logic* 43: 95–127, 2002.

[5] Kuhn, S. T., Minimal non-contingency logic, *Notre Dame Journal of Formal Logic* 36: 230–234, 1995.

[6] Lewis, D., *Convention: A Philosophical Study*, Harvard University Press, Cambridge, Mass., 1969.

[7] Marcos, J., Logics of essence and accident, *Bulletin of the Section of Logic* 34: 43–56, 2005.

[8] Montgomery, H., and R. Routley, Contingency and non-contingency bases for normal modal logics, *Logique et Analyse* 9: 318–328, 1966.

[9] Montgomery, H., and R. Routley, Modalities in a sequence of normal non-contingency modal systems, *Logique et Analyse* 12: 225–227, 1969.

[10] Pan, T., On logic of belief-disagreement among agents, in H. P. van Ditmarsch, J. Lang, S. Ju (eds.), *Logic, Rationality, and Interaction: Third International Workshop, LORI 2011, Guangzhou, China, October 10–13, 2011. Proceedings*, Springer, Berlin, Heidelberg, 2011, pp. 392–393.

[11] Steinsvold, C., Completeness for various logics of essence and accident, *Bulletin of the Section of Logic* 37: 93–101, 2008.

[12] Zolin, E., Completeness and definability in the logic of non-contingency, *Notre Dame Journal of Formal Logic* 40: 533–547, 1999.

[13] Zolin, E., Sequential reflexive logics with a non-contingency operator, *Mathematical Notes* 72: 784–798, 2002.

J. Chen, T. Pan
Department of Philosophy
Nanjing University
Nanjing
China
tqpan@126.com

J. Chen
chenjay509@hotmail.com