Blameworthiness in Strategic Games

Pavel Naumov
Department of Mathematical Sciences
Claremont McKenna College
Claremont, California 91711
pgn2@cornell.edu

Jia Tao
Department of Computer Science
Lafayette College
Easton, Pennsylvania 18042
taoj@lafayette.edu

Abstract

There are multiple notions of coalitional responsibility. The focus of this paper is on the blameworthiness defined through the principle of alternative possibilities: a coalition is blamable for a statement if the statement is true, but the coalition had a strategy to prevent it. The main technical result is a sound and complete bimodal logical system that describes properties of blameworthiness in one-shot games.

Introduction

It was a little after 9am on Friday, July 20th 2018, when a four-year-old boy accidentally shot his two-year old cousin in the town of Muscoy in Southern California. The victim was taken to a hospital, where she died an hour later (Oreskes 2018). The police arrested Cesar Lopez, victim's grandfather, as a felon in possession of a firearm and for child endangerment (Juarez and Miracle 2018).

The first charge against Lopez, a previously convicted felon, is based on California Penal Code §29800 (a) (1) that prohibits firearm access to "any person who has been convicted of, or has an outstanding warrant for, a felony under the laws of the United States, the State of California, or any other state, government, or country...". We assume that Lopez knew that California state law bans him from owning a gun, but his actions guaranteed that he broke the law.

The second charge is different because Lopez clearly never intended for his granddaughter to be killed. He never took any actions that would force her death. Nevertheless, he is *blamed* for not taking an action (locking the gun) to prevent the tragedy. Blameworthiness is tightly connected to the legal liability for negligence (Goudkamp 2004).

We are interested in logical systems for reasoning about different forms of responsibility. Xu (1998) introduced a complete axiomatization of a modal logical system for reasoning about responsibility defined as taking actions that guarantee a certain outcome. In our example, by possessing a gun Lopez guaranteed that he was responsible for breaking California law. Broersen, Herzig, and Troquard (2009) extended Xu's work from individual responsibility to group responsibility. In this paper we propose a complete logical system for reasoning about another form of responsibility that we call blameworthiness: a coalition is blamable for an outcome φ if φ is true, but the coalition had a strategy to prevent φ. In our example, Lopez had a strategy to prevent the death by keeping the gun in a safe place.

Principle of Alternative Possibilities Throughout centuries, blameworthiness, especially in the context of free will and moral responsibility, has been at the focus of philosophical discussions (Singer and Eddon 2013). Modern works on this topic include (Fields 1994; Fischer and Ravizza 2000; Nichols and Knobe 2007; Mason 2015; Widerker 2017). Frankfurt (1969) acknowledges that a dominant role in these discussions has been played by what he calls a *principle of alternate possibilities*: "a person is morally responsible for what he has done only if he could have done otherwise". As with many general principles, this one has many limitations that Frankfurt discusses; for example, when a person is coerced into doing something. Following the established tradition (Widerker 2017), we refer to this principle as the principle of *alternative* possibilities. Cushman (2015) talks about *counterfactual possibility*: "a person could have prevented their harmful conduct, even though they did not."

Halpern and Pearl proposed several versions of a formal definition of causality as a relation between sets of variables (Halpern 2016). This definition uses the counterfactual requirement which formalizes the principle of alternative possibilities. Halpern and Kleiman-Weiner (2018) used a similar setting to define *degrees* of blameworthiness. Batusov and Soutchanski (2018) gave a counterfactual-based definition of causality in situation calculus. (Alechina, Halpern, and Logan 2017) applied Halpern and Pearl definition of causality to team plans.

Coalitional Power in Strategic Games Pauly (2001; 2002) introduced logics of coalitional power that can be used to describe group abilities to achieve a certain result. His approach has been widely studied in the literature (Goranko 2001; van der Hoek and Wooldridge 2005; Borgo 2007; Sauro et al. 2006; Ågotnes et al. 2010; Ågotnes, van der Hoek, and Wooldridge 2009; Belardinelli 2014; Goranko, Jamroga, and Turrini 2013; Alechina et al. 2011; Galimullin and Alechina 2017; Goranko and Enqvist 2018).

In this paper we use Marc Pauly's framework to define blameworthiness of coalitions of players in strategic (one-shot) games. We say that a coalition C could be blamed for an outcome φ if φ is true, but the coalition C had a strategy to prevent φ. Thus, just like Halpern and Pearl's formal definition of causality, our definition of blameworthiness is based on the principle of alternative possibilities. However, because Marc Pauly's framework separates agents and outcomes, the proposed definition of blameworthiness is different and, arguably, more succinct.

The main technical result of this paper is a sound and complete bimodal logical system describing the interplay between group blameworthiness modality and necessity (or universal truth) modality. Our system is significantly different from earlier mentioned axiomatizations (Xu 1998) and (Broersen, Herzig, and Troquard 2009) because our semantics incorporates the principle of alternative possibilities.

Paper Outline This paper is organized as follows. First, we introduce the formal syntax and semantics of our logical system. Next, we state and discuss its axioms. In the section that follows, we give examples of formal derivations in our system. In the next two sections we prove the soundness and the completeness. The last section concludes with a discussion of possible future work.

Syntax and Semantics

In this paper we assume a fixed set $\mathcal{A}$ of agents and a fixed set of propositional variables Prop. By a coalition we mean an arbitrary subset of set $\mathcal{A}$.

Definition 1 *Φ is the minimal set of formulae such that*

1. *$p \in \Phi$ for each variable $p \in \mathsf{Prop}$,*
2. *$\varphi \to \psi, \neg\varphi \in \Phi$ for all formulae $\varphi, \psi \in \Phi$,*
3. *$\mathsf{N}\varphi$, $\mathsf{B}_C\varphi \in \Phi$ for each coalition $C \subseteq \mathcal{A}$ and each formula $\varphi \in \Phi$.*

Formula $\mathsf{N}\varphi$ is read as "statement φ is true under each play" and formula $\mathsf{B}_C\varphi$ as "coalition C is blamable for φ".

Boolean connectives $\vee$, $\wedge$, and $\leftrightarrow$ as well as constants $\bot$ and $\top$ are defined in the standard way. By formula $\overline{\mathsf{N}}\varphi$ we mean $\neg\mathsf{N}\neg\varphi$. For the disjunction of multiple formulae, we assume that parentheses are nested to the left. That is, formula $\chi_1 \vee \chi_2 \vee \chi_3$ is a shorthand for $(\chi_1 \vee \chi_2) \vee \chi_3$. As usual, the empty disjunction is defined to be $\bot$. For any two sets X and Y, by X^Y we denote the set of all functions from Y to X.

The formal semantics of modalities N and B is defined in terms of models, which we call *games*.

Definition 2 *A game is a tuple (Δ, Ω, P, π), where*

1. *Δ is a nonempty set of "actions",*
2. *Ω is a set of "outcomes",*
3. *the set of "plays" P is an arbitrary set of pairs (δ, ω) such that $\delta \in \Delta^{\mathcal{A}}$ and $\omega \in \Omega$,*
4. *π is a function that maps Prop into subsets of P.*

The example from the introduction can be captured in our setting by assuming that Lopez is the only actor who has two possible actions: $hide$ and $expose$ the gun in the game with two outcomes $alive$ and $dead$. Although a complete action profile is a function from the set of all agents to the domain of actions, in a single agent case any such profile can be described by specifying just the action of the single player. Thus, by complete action profile $hide$ we mean action profile that maps agent Lopez into action $hide$. The set of possible plays of this game consists of pairs $\{(hide, alive), (expose, alive), (expose, dead)\}$.

The above definition of a game is very close but not identical to the definition of a game frame in Pauly (2001; 2002) and the definition of a concurrent game structure, the semantics of ATL (Alur, Henzinger, and Kupferman 2002). Unlike these works, here we assume that the domain of choices is the same for all states and all agents. This difference is insignificant because all domains of choices in a game frame/concurrent game structure could be replaced with their union. More importantly, we assume that the mechanism is a relation, not a function. Our approach is more general, as it allows us to talk about blameworthiness in nondeterministic games, it also results in fewer axioms. Also, we do *not* assume that for any complete action profile δ there is at least one outcome ω such that $(\delta, \omega) \in P$. Thus, we allow the system to terminate under some action profiles without reaching an outcome. Without this assumption, we would need to add one extra axiom: $\neg\mathsf{B}_C\top$ and to make minor changes in the proof of the completeness.

Finally, in this paper we assume that atomic propositions are interpreted as statements about plays, not just outcomes. For example, the meaning of an atomic proposition p could be statement "either Lopez locked his gun or his granddaughter is dead". This is a more general approach than the one used in the existing literature, where atomic propositions are usually interpreted as statements about just outcomes. This difference is formally captured in the above definition through the assumption that value of π is a set of plays, not just a set of outcomes. As a result of this more general approach, all other statements in our logical system are also statements about plays, not outcomes. This is why relation $\Vdash$ in Definition 3 has a play (not an outcome) on the left.

If s_1 and s_2 are action profiles of coalitions C_1 and C_2, respectively, and C is any coalition such that $C \subseteq C_1 \cap C_2$, then we write $s_1 =_C s_2$ to denote that $s_1(a) = s_2(a)$ for each agent $a \in C$.

Next is the key definition of this paper. Its item 5 formally specifies blameworthiness using the principle of alternative possibilities.

Definition 3 *For any play $(\delta, \omega) \in P$ of a game (Δ, Ω, P, π) and any formula $\varphi \in \Phi$, the satisfiability relation $(\delta, \omega) \Vdash \varphi$ is defined recursively as follows:*

1. *$(\delta, \omega) \Vdash p$ if $(\delta, \omega) \in \pi(p)$, where $p \in \mathsf{Prop}$,*
2. *$(\delta, \omega) \Vdash \neg\varphi$ if $(\delta, \omega) \nVdash \varphi$,*
3. *$(\delta, \omega) \Vdash \varphi \to \psi$ if $(\delta, \omega) \nVdash \varphi$ or $(\delta, \omega) \Vdash \psi$,*
4. *$(\delta, \omega) \Vdash \mathsf{N}\varphi$ if $(\delta', \omega') \Vdash \varphi$ for each play $(\delta', \omega') \in P$,*
5. *$(\delta, \omega) \Vdash \mathsf{B}_C\varphi$ if $(\delta, \omega) \Vdash \varphi$ and there is $s \in \Delta^C$ such that for each play $(\delta', \omega') \in P$, if $s =_C \delta'$, then $(\delta', \omega') \nVdash \varphi$.*

Note that in part 5 of the above definition we do not assume that coalition C is a minimal one that could have prevented the outcome. This is different from the definition of blameworthiness in (Halpern 2017). Our approach is consistent with how word "blame" is often used in English. For example, the sentence "Millennials being blamed for decline of American cheese" (Gant 2018) does not imply that no one in the millennial generation likes American cheese.

Axioms

In addition to the propositional tautologies in language Φ, our logical system contains the following axioms.

1. Truth: $\mathsf{N}\varphi \to \varphi$ and $\mathsf{B}_C\varphi \to \varphi$,
2. Distributivity: $\mathsf{N}(\varphi \to \psi) \to (\mathsf{N}\varphi \to \mathsf{N}\psi)$,
3. Negative Introspection: $\neg\mathsf{N}\varphi \to \mathsf{N}\neg\mathsf{N}\varphi$,
4. None to Blame: $\neg\mathsf{B}_\varnothing\varphi$,
5. Joint Responsibility: if $C \cap D = \varnothing$, then $\overline{\mathsf{N}}\mathsf{B}_C\varphi \wedge \overline{\mathsf{N}}\mathsf{B}_D\psi \to (\varphi \vee \psi \to \mathsf{B}_{C\cup D}(\varphi \vee \psi))$,
6. Blame for Cause: $\mathsf{N}(\varphi \to \psi) \to (\mathsf{B}_C\psi \to (\varphi \to \mathsf{B}_C\varphi))$,
7. Monotonicity: $\mathsf{B}_C\varphi \to \mathsf{B}_D\varphi$, where $C \subseteq D$,
8. Fairness: $\mathsf{B}_C\varphi \to \mathsf{N}(\varphi \to \mathsf{B}_C\varphi)$.

We write $\vdash \varphi$ if formula φ is provable from the axioms of our system using the Modus Ponens and the Necessitation inference rules:

$$\frac{\varphi, \varphi \to \psi}{\psi}, \frac{\varphi}{\mathsf{N}\varphi}.$$

We write $X \vdash \varphi$ if formula φ is provable from the theorems of our logical system and an additional set of axioms X using only the Modus Ponens inference rule.

The Truth axiom for modality N, the Distributivity axiom, and the Negative Introspection axiom together with the Necessitation inference rule capture the fact that modality N, per Definition 3, is an S5 modality and thus satisfies all standard S5 properties.

The Truth axiom for modality B states that any coalition can be blamed only for a statement which is true. The None to Blame axiom states that the empty coalition cannot be blamed for anything. Intuitively, this axiom is true because the empty coalition has no power to prevent anything.

The Joint Responsibility axiom states that if disjoint coalitions C and D can be blamed for statements φ and ψ, respectively, on *some other (possibly two different) plays of the game* and the disjunction $\varphi \vee \psi$ is true on the current play, then the union of the two coalitions can be blamed for this disjunction on the current play. This axiom remotely resembles Xu (1998) axiom for the independence of individual agents, which in our notations can be stated as

$$\overline{\mathsf{N}}\mathsf{B}_{a_1}\varphi_1 \wedge \dots \wedge \overline{\mathsf{N}}\mathsf{B}_{a_n}\varphi_n \to \overline{\mathsf{N}}(\mathsf{B}_{a_1}\varphi_1 \wedge \dots \wedge \mathsf{B}_{a_n}\varphi_n).$$

Broersen, Herzig, and Troquard (2009) captured the independence of disjoint coalitions C and D in their Lemma 17:

$$\overline{\mathsf{N}}\mathsf{B}_C\varphi \wedge \overline{\mathsf{N}}\mathsf{B}_D\psi \to \overline{\mathsf{N}}(\mathsf{B}_C\varphi \wedge \mathsf{B}_D\psi).$$

In spite of these similarities, the definition of responsibility used in (Xu 1998) and (Broersen, Herzig, and Troquard 2009) does not assume the principle of alternative possibilities. The Joint Responsibility axiom is also similar to Marc Pauly (2001; 2002) Cooperation axiom for logic of coalitional power:

$$\mathsf{S}_C\varphi \wedge \mathsf{S}_D\psi \to \mathsf{S}_{C\cup D}(\varphi \wedge \psi),$$

where coalitions C and D are disjoint and $\mathsf{S}_C\varphi$ stands for "coalition C has a strategy to achieve φ".

The Blame for Cause axiom states that if formula φ universally implies ψ (informally, φ is a "cause" of ψ), then any coalition blamable for ψ should also be blamable for the "cause" φ as long as φ is actually true. The Monotonicity axiom states that any coalition is blamed for anything that a subcoalition is blamed for. Finally, the Fairness axiom states that if a coalition C is blamed for φ, then it should be blamed for φ whenever φ is true.

Examples of Derivations

The soundness of the axioms of our logical system is established in the next section. In this section we give several examples of formal proofs in our system. Together with the Truth axiom, the first example shows that statements $\mathsf{B}_C\mathsf{B}_C\varphi$ and $\mathsf{B}_C\varphi$ are equivalent in our system. That is, coalition C can be blamed for being blamed for φ if and only if it can be blamed for φ.

Lemma 1 $\vdash \mathsf{B}_C\varphi \to \mathsf{B}_C\mathsf{B}_C\varphi$.

PROOF. Note that $\vdash \mathsf{B}_C\varphi \to \varphi$ by the Truth axiom. Thus, $\vdash \mathsf{N}(\mathsf{B}_C\varphi \to \varphi)$ by the Necessitation rule. At the same time,

$$\vdash \mathsf{N}(\mathsf{B}_C\varphi \to \varphi) \to (\mathsf{B}_C\varphi \to (\mathsf{B}_C\varphi \to \mathsf{B}_C\mathsf{B}_C\varphi))$$

is an instance of the Blame for Cause axiom. Then, $\vdash \mathsf{B}_C\varphi \to (\mathsf{B}_C\varphi \to \mathsf{B}_C\mathsf{B}_C\varphi)$ by the Modus Ponens inference rule. Therefore, $\vdash \mathsf{B}_C\varphi \to \mathsf{B}_C\mathsf{B}_C\varphi$ by the propositional reasoning. ☒

The rest of the examples in this section are used later in the proof of the completeness.

Lemma 2 $\vdash \overline{\mathsf{N}}\mathsf{B}_C\varphi \to (\varphi \to \mathsf{B}_C\varphi)$.

PROOF. Note that $\vdash \mathsf{B}_C\varphi \to \mathsf{N}(\varphi \to \mathsf{B}_C\varphi)$ by the Fairness axiom. Hence, $\vdash \neg\mathsf{N}(\varphi \to \mathsf{B}_C\varphi) \to \neg\mathsf{B}_C\varphi$, by the law of contrapositive. Thus, $\vdash \mathsf{N}(\neg\mathsf{N}(\varphi \to \mathsf{B}_C\varphi) \to \neg\mathsf{B}_C\varphi)$ by the Necessitation inference rule. Hence, by the Distributivity axiom and the Modus Ponens inference rule,

$$\vdash \mathsf{N}\neg\mathsf{N}(\varphi \to \mathsf{B}_C\varphi) \to \mathsf{N}\neg\mathsf{B}_C\varphi.$$

At the same time, by the Negative Introspection axiom:

$$\vdash \neg\mathsf{N}(\varphi \to \mathsf{B}_C\varphi) \to \mathsf{N}\neg\mathsf{N}(\varphi \to \mathsf{B}_C\varphi).$$

Thus, by the laws of propositional reasoning,

$$\vdash \neg\mathsf{N}(\varphi \to \mathsf{B}_C\varphi) \to \mathsf{N}\neg\mathsf{B}_C\varphi.$$

Hence, by the law of contrapositive,

$$\vdash \neg\mathsf{N}\neg\mathsf{B}_C\varphi \to \mathsf{N}(\varphi \to \mathsf{B}_C\varphi).$$

Note that $\mathsf{N}(\varphi \to \mathsf{B}_C\varphi) \to (\varphi \to \mathsf{B}_C\varphi)$ is an instance of the Truth axiom. Thus, by propositional reasoning,

$$\vdash \neg\mathsf{N}\neg\mathsf{B}_C\varphi \to (\varphi \to \mathsf{B}_C\varphi).$$

Hence, $\vdash \overline{\mathsf{N}}\mathsf{B}_C\varphi \to (\varphi \to \mathsf{B}_C\varphi)$ by the definition of $\overline{\mathsf{N}}$. ☒

Lemma 3 *If* $\vdash \varphi \leftrightarrow \psi$*, then* $\vdash \mathsf{B}_C\varphi \to \mathsf{B}_C\psi$.

PROOF. By the Blame for Cause axiom,

$$\vdash \mathsf{N}(\psi \to \varphi) \to (\mathsf{B}_C\varphi \to (\psi \to \mathsf{B}_C\psi)).$$

Assumption $\vdash \varphi \leftrightarrow \psi$ implies $\vdash \psi \to \varphi$ by the laws of propositional reasoning. Thus, $\vdash \mathsf{N}(\psi \to \varphi)$ by the Necessitation inference rule. Hence, by the Modus Ponens rule,

$$\vdash \mathsf{B}_C\varphi \to (\psi \to \mathsf{B}_C\psi).$$

Thus, by the laws of propositional reasoning,

$$\vdash (\mathsf{B}_C\varphi \to \psi) \to (\mathsf{B}_C\varphi \to \mathsf{B}_C\psi). \tag{1}$$

Note that $\vdash \mathsf{B}_C\varphi \to \varphi$ by the Truth axiom. At the same time, $\vdash \varphi \leftrightarrow \psi$ by the assumption of the lemma. Thus, by the laws of propositional reasoning, $\vdash \mathsf{B}_C\varphi \to \psi$. Therefore, $\vdash \mathsf{B}_C\varphi \to \mathsf{B}_C\psi$ by the Modus Ponens inference rule from statement (1). ☒

Lemma 4 $\varphi \vdash \overline{\mathsf{N}}\varphi$.

PROOF. By the Truth axioms, $\vdash \mathsf{N}\neg\varphi \to \neg\varphi$. Thus, by the law of contrapositive, $\vdash \varphi \to \neg\mathsf{N}\neg\varphi$. Hence, $\vdash \varphi \to \overline{\mathsf{N}}\varphi$ by the definition of the modality $\overline{\mathsf{N}}$. Therefore, $\varphi \vdash \overline{\mathsf{N}}\varphi$ by the Modus Ponens inference rule. ☒

The next lemma generalizes the Joint Responsibility axiom from two coalitions to multiple coalitions.

Lemma 5 *For any integer* $n \geq 0$ *and any pairwise disjoint sets* $D_1, \dots, D_n$,

$$\{\overline{\mathsf{N}}\mathsf{B}_{D_i}\chi_i\}_{i=1}^n, \chi_1 \vee \dots \vee \chi_n \vdash \mathsf{B}_{D_1\cup\dots\cup D_n}(\chi_1 \vee \dots \vee \chi_n).$$

PROOF. We prove the lemma by induction on n. If $n = 0$, then disjunction $\chi_1 \vee \dots \vee \chi_n$ is Boolean constant false $\bot$ by definition. Thus, the statement of the lemma is $\bot \vdash \mathsf{B}_\varnothing\bot$, which is provable in the propositional logic due to the assumption $\bot$ on the left-hand side of $\vdash$.

Next, suppose that $n = 1$. Then, from Lemma 2 it follows that $\overline{\mathsf{N}}\mathsf{B}_{D_1}\chi_1, \chi_1 \vdash \mathsf{B}_{D_1}\chi_1$.

Suppose that $n \geq 2$. By the Joint Responsibility axiom and the Modus Ponens inference rule,

$$\begin{aligned}&\overline{\mathsf{N}}\mathsf{B}_{D_1\cup\dots\cup D_{n-1}}(\chi_1 \vee \dots \vee \chi_{n-1}), \overline{\mathsf{N}}\mathsf{B}_{D_n}\chi_n,\\ &\chi_1 \vee \dots \vee \chi_{n-1} \vee \chi_n\\ &\vdash \mathsf{B}_{D_1\cup\dots\cup D_{n-1}\cup D_n}(\chi_1 \vee \dots \vee \chi_{n-1} \vee \chi_n).\end{aligned}$$

Thus, by Lemma 4,

$$\begin{aligned}&\mathsf{B}_{D_1\cup\dots\cup D_{n-1}}(\chi_1 \vee \dots \vee \chi_{n-1}), \overline{\mathsf{N}}\mathsf{B}_{D_n}\chi_n,\\ &\chi_1 \vee \dots \vee \chi_{n-1} \vee \chi_n\\ &\vdash \mathsf{B}_{D_1\cup\dots\cup D_{n-1}\cup D_n}(\chi_1 \vee \dots \vee \chi_{n-1} \vee \chi_n).\end{aligned}$$

At the same time, by the induction hypothesis,

$$\begin{aligned}&\{\overline{\mathsf{N}}\mathsf{B}_{D_i}\chi_i\}_{i=1}^{n-1}, \chi_1 \vee \dots \vee \chi_{n-1}\\ &\vdash \mathsf{B}_{D_1\cup\dots\cup D_{n-1}}(\chi_1 \vee \dots \vee \chi_{n-1}).\end{aligned}$$

Hence,

$$\begin{aligned}&\{\overline{\mathsf{N}}\mathsf{B}_{D_i}\chi_i\}_{i=1}^{n}, \chi_1 \vee \dots \vee \chi_{n-1}, \chi_1 \vee \dots \vee \chi_{n-1} \vee \chi_n\\ &\vdash \mathsf{B}_{D_1\cup\dots\cup D_{n-1}\cup D_n}(\chi_1 \vee \dots \vee \chi_{n-1} \vee \chi_n).\end{aligned}$$

Since $\chi_1 \vee \dots \vee \chi_{n-1} \vdash \chi_1 \vee \dots \vee \chi_{n-1} \vee \chi_n$ is provable in propositional logic,

$$\begin{aligned}&\{\overline{\mathsf{N}}\mathsf{B}_{D_i}\chi_i\}_{i=1}^{n}, \chi_1 \vee \dots \vee \chi_{n-1}\\ &\vdash \mathsf{B}_{D_1\cup\dots\cup D_{n-1}\cup D_n}(\chi_1 \vee \dots \vee \chi_{n-1} \vee \chi_n).\end{aligned} \tag{2}$$

Similarly, by the Joint Responsibility axiom and the Modus Ponens inference rule,

$$\begin{aligned}&\overline{\mathsf{N}}\mathsf{B}_{D_1}\chi_1, \overline{\mathsf{N}}\mathsf{B}_{D_2\cup\dots\cup D_n}(\chi_2 \vee \dots \vee \chi_n),\\ &\chi_1 \vee (\chi_2 \vee \dots \vee \chi_n)\\ &\vdash \mathsf{B}_{D_1\cup\dots\cup D_{n-1}\cup D_n}(\chi_1 \vee (\chi_2 \vee \dots \vee \chi_n)).\end{aligned}$$

Since formula $\chi_1 \vee (\chi_2 \vee \dots \vee \chi_n) \leftrightarrow \chi_1 \vee \chi_2 \vee \dots \vee \chi_n$ is provable in the propositional logic, by Lemma 3,

$$\begin{aligned}&\overline{\mathsf{N}}\mathsf{B}_{D_1}\chi_1, \overline{\mathsf{N}}\mathsf{B}_{D_2\cup\dots\cup D_n}(\chi_2 \vee \dots \vee \chi_n), \chi_1 \vee \chi_2 \vee \dots \vee \chi_n\\ &\vdash \mathsf{B}_{D_1\cup\dots\cup D_{n-1}\cup D_n}(\chi_1 \vee \chi_2 \vee \dots \vee \chi_n).\end{aligned}$$

Thus, by Lemma 4,

$$\begin{aligned}&\overline{\mathsf{N}}\mathsf{B}_{D_1}\chi_1, \mathsf{B}_{D_2\cup\dots\cup D_n}(\chi_2 \vee \dots \vee \chi_n), \chi_1 \vee \chi_2 \vee \dots \vee \chi_n\\ &\vdash \mathsf{B}_{D_1\cup\dots\cup D_{n-1}\cup D_n}(\chi_1 \vee \chi_2 \vee \dots \vee \chi_n).\end{aligned}$$

At the same time, by the induction hypothesis,

$$\{\overline{\mathsf{N}}\mathsf{B}_{D_i}\chi_i\}_{i=2}^n, \chi_2 \vee \dots \vee \chi_n \vdash \mathsf{B}_{D_2\cup\dots\cup D_n}(\chi_2 \vee \dots \vee \chi_n).$$

Hence,

$$\begin{aligned}&\{\overline{\mathsf{N}}\mathsf{B}_{D_i}\chi_i\}_{i=1}^n, \chi_2 \vee \dots \vee \chi_n, \chi_1 \vee \chi_2 \vee \dots \vee \chi_n\\ &\vdash \mathsf{B}_{D_1\cup D_2\cup\dots\cup D_n}(\chi_1 \vee \chi_2 \vee \dots \vee \chi_n).\end{aligned}$$

Since $\chi_2 \vee \dots \vee \chi_n \vdash \chi_1 \vee \dots \vee \chi_{n-1} \vee \chi_n$ is provable in propositional logic,

$$\begin{aligned}&\{\overline{\mathsf{N}}\mathsf{B}_{D_i}\chi_i\}_{i=1}^n, \chi_2 \vee \dots \vee \chi_n\\ &\vdash \mathsf{B}_{D_1\cup\dots\cup D_{n-1}\cup D_n}(\chi_1 \vee \chi_2 \vee \dots \vee \chi_n).\end{aligned} \tag{3}$$

Finally, note that the following statement is provable in the propositional logic for $n \geq 2$,

$$\vdash \chi_1 \vee \dots \vee \chi_n \to (\chi_1 \vee \dots \vee \chi_{n-1}) \vee (\chi_2 \vee \dots \vee \chi_n).$$

Therefore, from statement (2) and statement (3),

$$\{\overline{\mathsf{N}}\mathsf{B}_{D_i}\chi_i\}_{i=1}^n, \chi_1 \vee \dots \vee \chi_n \vdash \mathsf{B}_{D_1\cup\dots\cup D_n}(\chi_1 \vee \dots \vee \chi_n)$$

by the laws of propositional reasoning. ☒

Note that modality N satisfies all axioms of S5. The following two lemmas state well-known property of S5. Their proofs can be found, for example, in (Naumov and Tao 2018a).

Lemma 6 *If* $\varphi_1, \dots, \varphi_n \vdash \psi$*, then* $\mathsf{N}\varphi_1, \dots, \mathsf{N}\varphi_n \vdash \mathsf{N}\psi$. ☒

Lemma 7 $\vdash \mathsf{N}\varphi \to \mathsf{N}\mathsf{N}\varphi$. ☒

Lemma 8 *For any integer* $n \geq 0$ *and any disjoint sets* $D_1, \dots, D_n \subseteq C$,

$$\{\overline{\mathsf{N}}\mathsf{B}_{D_i}\chi_i\}_{i=1}^n, \mathsf{N}(\varphi \to \chi_1 \vee \dots \vee \chi_n) \vdash \mathsf{N}(\varphi \to \mathsf{B}_C\varphi).$$

PROOF. By Lemma 5,

$$\{\overline{\mathsf{N}}\mathsf{B}_{D_i}\chi_i\}_{i=1}^n, \chi_1 \vee \dots \vee \chi_n \vdash \mathsf{B}_{D_1\cup\dots\cup D_n}(\chi_1 \vee \dots \vee \chi_n).$$

Thus, by the Monotonicity axiom,

$$\{\overline{\mathsf{N}}\mathsf{B}_{D_i}\chi_i\}_{i=1}^n, \chi_1 \vee \dots \vee \chi_n \vdash \mathsf{B}_C(\chi_1 \vee \dots \vee \chi_n).$$

Hence, by the Modus Ponens inference rule

$$\{\overline{\mathsf{N}}\mathsf{B}_{D_i}\chi_i\}_{i=1}^n, \varphi, \varphi \to \chi_1 \vee \dots \vee \chi_n \vdash \mathsf{B}_C(\chi_1 \vee \dots \vee \chi_n).$$

By the Truth axiom and the Modus Ponens inference rule,

$$\{\overline{\mathsf{N}}\mathsf{B}_{D_i}\chi_i\}_{i=1}^n, \varphi, \mathsf{N}(\varphi \to \chi_1 \vee \dots \vee \chi_n) \vdash \mathsf{B}_C(\chi_1 \vee \dots \vee \chi_n).$$

Note that $\mathsf{N}(\varphi \to \chi_1 \vee \dots \vee \chi_n) \to (\mathsf{B}_C(\chi_1 \vee \dots \vee \chi_n) \to (\varphi \to \mathsf{B}_C\varphi))$ is an instance of the Blame for Cause axiom. Thus, by the Modus Ponens inference rule applied twice,

$$\{\overline{\mathsf{N}}\mathsf{B}_{D_i}\chi_i\}_{i=1}^n, \varphi, \mathsf{N}(\varphi \to \chi_1 \vee \dots \vee \chi_n) \vdash \varphi \to \mathsf{B}_C\varphi.$$

By the Modus Ponens inference rule,

$$\{\overline{\mathsf{N}}\mathsf{B}_{D_i}\chi_i\}_{i=1}^n, \varphi, \mathsf{N}(\varphi \to \chi_1 \vee \dots \vee \chi_n) \vdash \mathsf{B}_C\varphi.$$

By the deduction lemma,

$$\{\overline{\mathsf{N}}\mathsf{B}_{D_i}\chi_i\}_{i=1}^n, \mathsf{N}(\varphi \to \chi_1 \vee \dots \vee \chi_n) \vdash \varphi \to \mathsf{B}_C\varphi.$$

By Lemma 6,

$$\{\mathsf{N}\overline{\mathsf{N}}\mathsf{B}_{D_i}\chi_i\}_{i=1}^n, \mathsf{N}\mathsf{N}(\varphi \to \chi_1 \vee \dots \vee \chi_n) \vdash \mathsf{N}(\varphi \to \mathsf{B}_C\varphi).$$

By the definition of modality $\overline{\mathsf{N}}$, the Negative Introspection axiom, and the Modus Ponens inference rule,

$$\{\overline{\mathsf{N}}\mathsf{B}_{D_i}\chi_i\}_{i=1}^n, \mathsf{N}\mathsf{N}(\varphi \to \chi_1 \vee \dots \vee \chi_n) \vdash \mathsf{N}(\varphi \to \mathsf{B}_C\varphi)$$

Therefore, by Lemma 7 and the Modus Ponens inference rule, the statement of the lemma follows. $\boxtimes$

Soundness

In the following lemmas, $(\delta,\omega) \in P$ is a play of an arbitrary game (Δ, Ω, P, π) and $\varphi, \psi \in \Phi$ are arbitrary formulae.

Lemma 9 $(\delta,\omega) \nVdash \mathsf{B}_\varnothing\varphi$.

PROOF. Suppose that $(\delta,\omega) \Vdash \mathsf{B}_\varnothing\varphi$. Thus, by Definition 3, we have $(\delta,\omega) \Vdash \varphi$ and there is an action profile $s \in \Delta^\varnothing$ such that for each play $(\delta',\omega') \in P$, if $s =_\varnothing \delta'$, then $(\delta',\omega') \nVdash \varphi$.

Consider $\delta' = \delta$ and $\omega' = \omega$. Note that $s =_\varnothing \delta'$ is vacuously true. Hence, $(\delta',\omega') \nVdash \varphi$. In other words, $(\delta,\omega) \nVdash \varphi$, which leads to a contradiction. $\boxtimes$

Lemma 10 *For all sets $C, D \subseteq \mathcal{A}$ such that $C \cap D = \varnothing$, if $(\delta,\omega) \Vdash \overline{\mathsf{N}}\mathsf{B}_C\varphi$, $(\delta,\omega) \Vdash \overline{\mathsf{N}}\mathsf{B}_D\psi$, and $(\delta,\omega) \Vdash \varphi \vee \psi$, then $(\delta,\omega) \Vdash \mathsf{B}_{C\cup D}(\varphi \vee \psi)$.*

PROOF. Let $(\delta,\omega) \Vdash \overline{\mathsf{N}}\mathsf{B}_C\varphi$ and $(\delta,\omega) \Vdash \overline{\mathsf{N}}\mathsf{B}_D\psi$. Thus, by Definition 3 and the definition of modality $\overline{\mathsf{N}}$, there are plays $(\delta_1,\omega_1) \in P$ and $(\delta_2,\omega_2) \in P$ such that $(\delta_1,\omega_1) \Vdash \mathsf{B}_C\varphi$ and $(\delta_2,\omega_2) \Vdash \mathsf{B}_D\psi$.

By Definition 3, statement $(\delta_1,\omega_1) \Vdash \mathsf{B}_C\varphi$ implies that there is $s_1 \in \Delta^C$ such that for each play $(\delta',\omega') \in P$, if $s_1 =_C \delta'$, then $(\delta',\omega') \nVdash \varphi$.

Similarly, by Definition 3, statement $(\delta_2,\omega_2) \Vdash \mathsf{B}_D\psi$ implies that there is $s_2 \in \Delta^D$ such that for each play $(\delta',\omega') \in P$, if $s_2 =_D \delta'$, then $(\delta',\omega') \nVdash \psi$.

Consider an action profile s of coalition $C \cup D$ such that

$$s(a) = \begin{cases} s_1(a), & \text{if } a \in C, \\ s_2(a), & \text{if } a \in D. \end{cases}$$

Note that the action profile s is well-defined because sets C and D are disjoint by the assumption of the lemma.

The choice of action profiles s_1, s_2, and s implies that for each play $(\delta',\omega') \in P$, if $s =_{C\cup D} \delta'$, then $(\delta',\omega') \nVdash \varphi$ and $(\delta',\omega') \nVdash \psi$. Thus, for each play $(\delta',\omega') \in P$, if $s =_{C\cup D} \delta'$, then $(\delta',\omega') \nVdash \varphi \vee \psi$. Therefore, $(\delta,\omega) \Vdash \mathsf{B}_{C\cup D}(\varphi \vee \psi)$ by Definition 3 and due to the assumption $(\delta,\omega) \Vdash \varphi \vee \psi$ of the lemma. $\boxtimes$

Lemma 11 *If $(\delta,\omega) \Vdash \mathsf{N}(\varphi \to \psi)$, $(\delta,\omega) \Vdash \mathsf{B}_C\psi$, and $(\delta,\omega) \Vdash \varphi$, then $(\delta,\omega) \Vdash \mathsf{B}_C\varphi$.*

PROOF. By Definition 3, assumption $(\delta,\omega) \Vdash \mathsf{B}_C\psi$ implies that there is $s \in \Delta^C$ such that for each play $(\delta',\omega') \in P$, if $s =_C \delta'$, then $(\delta',\omega') \nVdash \psi$.

At the same time, $(\delta',\omega') \Vdash \varphi \to \psi$ for each play $(\delta',\omega') \in P$ by the assumption $(\delta,\omega) \Vdash \mathsf{N}(\varphi \to \psi)$ of the lemma and Definition 3.

Thus, $(\delta',\omega') \nVdash \varphi$ for each play $(\delta',\omega') \in P$ such that $s =_C \delta'$ by Definition 3. Hence, $(\delta,\omega) \Vdash \mathsf{B}_C\varphi$ by Definition 3 and the assumption $(\delta,\omega) \Vdash \varphi$ of the lemma. $\boxtimes$

Lemma 12 *For all sets $C, D \subseteq \mathcal{A}$ such that $C \subseteq D$, if $(\delta,\omega) \Vdash \mathsf{B}_C\varphi$, then $(\delta,\omega) \Vdash \mathsf{B}_D\varphi$.*

PROOF. By Definition 3, assumption $(\delta,\omega) \Vdash \mathsf{B}_C\varphi$ implies that $(\delta,\omega) \Vdash \varphi$ and there is $s \in \Delta^C$ such that for each play $(\delta',\omega') \in P$, if $s =_C \delta'$, then $(\delta',\omega') \nVdash \varphi$.

By Definition 2, set Δ is not empty. Let $d_0 \in \Delta$. Consider an action profile s' of coalition D such that

$$s'(a) = \begin{cases} s(a), & \text{if } a \in C, \\ d_0, & \text{if } a \in D \setminus C. \end{cases}$$

Then, by the choice of action profile s and because $C \subseteq D$, for each play $(\delta',\omega') \in P$, if $s' =_D \delta'$, then $(\delta',\omega') \nVdash \varphi$. Therefore, $(\delta,\omega) \Vdash \mathsf{B}_D\varphi$ by Definition 3 and because $(\delta,\omega) \Vdash \varphi$, as we have shown earlier. $\boxtimes$

Lemma 13 *If $(\delta,\omega) \Vdash \mathsf{B}_C\varphi$, then $(\delta,\omega) \Vdash \mathsf{N}(\varphi \to \mathsf{B}_C\varphi)$.*

PROOF. Consider any play $(\delta',\omega') \in P$. By Definition 3, it suffices to show that if $(\delta',\omega') \Vdash \varphi$, then $(\delta',\omega') \Vdash \mathsf{B}_C\varphi$. Thus, again by Definition 3, it suffices to prove there is $s \in \Delta^C$ such that for each play $(\delta'',\omega'') \in P$, if $s =_C \delta''$, then $(\delta'',\omega'') \nVdash \varphi$. The last statement follows from the assumption $(\delta,\omega) \Vdash \mathsf{B}_C\varphi$ and Definition 3. $\boxtimes$

Completeness

We start the proof of the completeness by defining the canonical game $G(\omega_0) = (\Delta, \Omega, P, \pi)$ for each maximal consistent set of formulae ω_0.

Definition 4 *The set of outcomes Ω is the set of all maximal consistent sets of formulae ω such that for each formula $\varphi \in \Phi$ if $\mathsf{N}\varphi \in \omega_0$, then $\varphi \in \omega$.*

Informally, an action of an agent in the canonical game is designed to "veto" a formula. The domain of choices of the canonical model consists of all formulae in set Φ. To veto a formula ψ, an agent must choose action ψ. The mechanism of the canonical game guarantees that if $\overline{\mathsf{N}}\mathsf{B}_C\psi \in \omega_0$ and all agents in the coalition C veto formula ψ, then $\neg\psi$ is satisfied in the outcome.

Definition 5 *The domain of actions Δ is set Φ.*

Definition 6 *The set $P \subseteq \Delta^{\mathcal{A}} \times \Omega$ consists of all pairs (δ, ω) such that for any formula $\overline{\mathsf{N}}\mathsf{B}_C\psi \in \omega_0$, if $\delta(a) = \psi$ for each agent $a \in C$, then $\neg\psi \in \omega$.*

Definition 7 $\pi(p) = \{(\delta, \omega) \in P \mid p \in \omega\}$.

This concludes the definition of the canonical game $G(\omega_0)$. The next four lemmas are auxiliary results leading to the proof of the completeness in Theorem 1.

Lemma 14 *For any play $(\delta, \omega) \in P$, any action profile $s \in \Delta^C$, and any formula $\neg(\varphi \to \mathsf{B}_C\varphi) \in \omega$, there is a play $(\delta', \omega') \in P$ such that $s =_C \delta'$ and $\varphi \in \omega'$.*

PROOF. Consider the following set of formulae:

$$\begin{aligned} X \;=\; & \{\varphi\} \cup \{\psi \mid \mathsf{N}\psi \in \omega_0\} \\ & \cup \{\neg\chi \mid \overline{\mathsf{N}}\mathsf{B}_D\chi \in \omega_0, D \subseteq C, \forall a \in D(s(a) = \chi)\}. \end{aligned}$$

Claim 1 *Set X is consistent.*

PROOF OF CLAIM. Suppose the opposite. Thus, there are

$$\begin{aligned} \text{formulae} \quad & \mathsf{N}\psi_1, \dots, \mathsf{N}\psi_m \in \omega_0, && (4)\\ \text{and formulae} \quad & \overline{\mathsf{N}}\mathsf{B}_{D_1}\chi_1, \dots, \overline{\mathsf{N}}\mathsf{B}_{D_n}\chi_n \in \omega_0, && (5)\\ \text{such that} \quad & D_1, \dots, D_n \subseteq C, && (6)\\ & s(a) = \chi_i \text{ for all } a \in D_i, i \leq n, && (7)\\ \text{and} \quad & \psi_1, \dots, \psi_m, \neg\chi_1, \dots, \neg\chi_n \vdash \neg\varphi. && (8) \end{aligned}$$

Without loss of generality, we can assume that formulae $\chi_1, \dots, \chi_n$ are distinct. Thus, assumption (7) implies that sets $D_1, \dots, D_n$ are pairwise disjoint.

By propositional reasoning, assumption (8) implies that

$$\psi_1, \dots, \psi_m \vdash \varphi \to \chi_1 \vee \dots \vee \chi_n.$$

Thus, $\mathsf{N}\psi_1, \dots, \mathsf{N}\psi_m \vdash \mathsf{N}(\varphi \to \chi_1 \vee \dots \vee \chi_n)$, by Lemma 6. Hence, $\omega_0 \vdash \mathsf{N}(\varphi \to \chi_1 \vee \dots \vee \chi_n)$, by assumption (4),

Thus, by Lemma 8, using assumptions (5) and the fact that sets $D_1, \dots, D_n$ are pairwise disjoint, $\omega_0 \vdash \mathsf{N}(\varphi \to \mathsf{B}_C\varphi)$. Hence $\mathsf{N}(\varphi \to \mathsf{B}_C\varphi) \in \omega_0$ because set ω_0 is maximal. Then, $\varphi \to \mathsf{B}_C\varphi \in \omega$ by Definition 4, which contradicts the assumption $\neg(\varphi \to \mathsf{B}_C\varphi) \in \omega$ of the lemma because set ω is consistent. Therefore, set X is consistent. $\boxtimes$

Let ω' be any maximal consistent extension of set X. Thus, $\varphi \in X \subseteq \omega'$ by the choice of sets X and ω'. Also, $\omega' \in \Omega$ by Definition 4 and the choice of sets X and ω'.

Let the complete action profile δ' be defined as follows:

$$\delta'(a) = \begin{cases} s(a), & \text{if } a \in C, \\ \bot, & \text{otherwise.} \end{cases} \tag{9}$$

Then, $s =_C \delta'$.

Claim 2 $(\delta', \omega') \in P$.

PROOF OF CLAIM. Consider any formula $\overline{\mathsf{N}}\mathsf{B}_D\chi \in \omega_0$ such that $\delta'(a) = \chi$ for each $a \in D$. By Definition 6, it suffices to show that $\neg\chi \in \omega'$.

Case I: $D \subseteq C$. Thus, $\neg\chi \in X$ by the definition of set X. Therefore, $\neg\chi \in \omega'$ by the choice of set ω'.

Case II: $D \nsubseteq C$. Consider any $d_0 \in D \setminus C$. Thus, $\delta'(d_0) = \bot$ by equation (9). Also, $\delta'(d_0) = \chi$ because $d_0 \in D$. Thus, $\chi \equiv \bot$ and formula $\neg\chi$ is a tautology. Hence, $\neg\chi \in \omega'$ by the maximality of set ω'. $\boxtimes$

This concludes the proof of the lemma. $\boxtimes$

Lemma 15 *For any outcome $\omega \in \Omega$, there is a complete action profile $\delta \in \Delta^{\mathcal{A}}$ such that $(\delta, \omega) \in P$.*

PROOF. Define a complete action profile δ such that $\delta(a) = \bot$ for each agent $a \in \mathcal{A}$. To prove $(\delta, \omega) \in P$, consider any formula $\overline{\mathsf{N}}\mathsf{B}_D\chi \in \omega_0$ such that $\delta(a) = \chi$ for each $a \in D$. By Definition 6, it suffices to show that $\neg\chi \in \omega$.

Case I: $D = \varnothing$. Thus, $\vdash \neg\mathsf{B}_D\chi$ by the None to Blame axiom. Hence, $\vdash \mathsf{N}\neg\mathsf{B}_D\chi$ by the Necessitation inference rule. Then, $\neg\mathsf{N}\neg\mathsf{B}_D\chi \notin \omega_0$ by the consistency of the set ω_0. Therefore, $\overline{\mathsf{N}}\mathsf{B}_D\chi \notin \omega_0$ by the definition of the modality $\overline{\mathsf{N}}$, which contradicts the choice of formula $\overline{\mathsf{N}}\mathsf{B}_D\chi$.

Case II: $D \neq \varnothing$. Thus, there is at least one agent $d_0 \in D$. Hence, $\chi = \delta(d_0) = \bot$ by the definition of the complete action profile δ. Then, $\neg\chi$ is a tautology. Thus, $\neg\chi \in \omega$ by the maximality of set ω. $\boxtimes$

Lemma 16 *For any play $(\delta, \omega) \in P$ and any formula $\neg\mathsf{N}\varphi \in \omega$, there is a play $(\delta', \omega') \in P$ such that $\neg\varphi \in \omega'$.*

PROOF. Consider the set $X = \{\neg\varphi\} \cup \{\psi \mid \mathsf{N}\psi \in \omega_0\}$. First, we show that set X is consistent. Suppose the opposite. Thus, there are formulae $\mathsf{N}\psi_1, \dots, \mathsf{N}\psi_n \in \omega_0$ such that $\psi_1, \dots, \psi_n \vdash \varphi$. Hence, $\mathsf{N}\psi_1, \dots, \mathsf{N}\psi_n \vdash \mathsf{N}\varphi$ by Lemma 6. Thus, $\omega_0 \vdash \mathsf{N}\varphi$ because $\mathsf{N}\psi_1, \dots, \mathsf{N}\psi_n \in \omega_0$. Hence, $\omega_0 \vdash \mathsf{N}\mathsf{N}\varphi$ by Lemma 7. Therefore, $\mathsf{N}\varphi \in \omega$ by assumption $\omega \in \Omega$ and Definition 4. Hence, $\neg\mathsf{N}\varphi \notin \omega$ by the consistency of set ω, which contradicts the assumption of the lemma. Thus, set X is consistent.

Let ω' be any maximal consistent extension of set X. Note that $\neg\varphi \in X \subseteq \omega'$ by the definition of set X. By Lemma 15, there is a complete action profile δ' such that $(\delta', \omega') \in P$. $\boxtimes$

Lemma 17 $(\delta, \omega) \Vdash \varphi$ *iff $\varphi \in \omega$ for each play $(\delta, \omega) \in P$ and each formula $\varphi \in \Phi$.*

PROOF. We prove the lemma by structural induction on formula φ. If φ is a propositional variable, then the required follows from Definition 3 and Definition 7. The cases when φ is an implication or a negation follow from the maximality and the consistency of set ω in the standard way.

Suppose that formula φ has the form $\mathsf{N}\psi$.

$(\Rightarrow)$: Let $\mathsf{N}\psi \notin \omega$. Thus, $\neg\mathsf{N}\psi \in \omega$ by the maximality of set ω. Hence, by Lemma 16, there is a play $(\delta', \omega') \in P$ such that $\neg\psi \in \omega'$. Then, $\psi \notin \omega'$ by the consistency of set ω'. Thus, $(\delta', \omega') \nVdash \psi$ by the induction hypothesis. Therefore, $(\delta, \omega) \nVdash \mathsf{N}\psi$ by Definition 3.

$(\Leftarrow)$: Let $\mathsf{N}\psi \in \omega$. Thus, $\neg\mathsf{N}\psi \notin \omega$ by the consistency of set ω. Hence, $\mathsf{N}\neg\mathsf{N}\psi \notin \omega_0$ by Definition 4. Then, $\omega_0 \nvdash \mathsf{N}\neg\mathsf{N}\psi$ by the maximality of set ω_0. Thus, $\omega_0 \nvdash \neg\mathsf{N}\psi$ by the Negative Introspection axiom. Hence, $\mathsf{N}\psi \in \omega_0$ by the maximality of set ω_0. Then, $\psi \in \omega'$ for each $\omega' \in \Omega$ by Definition 4. Thus, by the induction hypothesis, $(\delta', \omega') \Vdash \psi$ for each $(\delta', \omega') \in P$. Therefore, $(\delta, \omega) \Vdash \mathsf{N}\psi$ by Definition 3.

Suppose that formula φ has the form $\mathsf{B}_C\psi$.

$(\Rightarrow)$: Assume that $\mathsf{B}_C\psi \notin \omega$. First, we consider the case when $\psi \notin \omega$. Then, $(\delta, \omega) \nVdash \psi$ by the induction hypothesis. Hence, $(\delta, \omega) \nVdash \mathsf{B}_C\psi$ by Definition 3.

Next, assume that $\psi \in \omega$. Note that $\psi \to \mathsf{B}_C\psi \notin \omega$. Indeed, if $\psi \to \mathsf{B}_C\psi \in \omega$, then $\omega \vdash \mathsf{B}_C\psi$ by the Modus Ponens inference rule. Thus, $\mathsf{B}_C\psi \in \omega$ by the maximality of set ω, which contradicts the assumption above.

Because ω is a maximal set, statement $\psi \to \mathsf{B}_C\psi \notin \omega$ implies that $\neg(\psi \to \mathsf{B}_C\psi) \in \omega$. Thus, by Lemma 14, for any action profile $s \in \Delta^C$, there is a play (δ', ω') such that $s =_C \delta'$ $\psi \in \omega'$. Hence, by the induction hypothesis, for any action profile $s \in \Delta^C$ there is a play (δ', ω') such that $(\delta', \omega') \Vdash \psi$. Therefore, $(\delta, \omega) \nVdash \mathsf{B}_C\psi$ by Definition 3.

$(\Leftarrow)$: Suppose that $\mathsf{B}_C\psi \in \omega$. Thus, $\omega \vdash \psi$ by the Truth axiom. Hence, $\psi \in \omega$ by the maximality of the set ω. Thus, $(\delta, \omega) \Vdash \psi$ by the induction hypothesis.

Next, define an action profile $s \in \Delta^C$ to be such that $s(a) = \psi$ for each $a \in C$. Consider any play $(\delta', \omega') \in P$ such that $s =_C \delta'$. By Definition 3, it suffices to show that $(\delta', \omega') \nVdash \psi$.

Statement $\mathsf{B}_C\psi \in \omega$ implies that $\neg\mathsf{B}_C\psi \notin \omega$ because set ω is consistent. Thus, $\mathsf{N}\neg\mathsf{B}_C\psi \notin \omega_0$ by Definition 4 and because $\omega \in \Omega$. Hence, $\neg\mathsf{N}\neg\mathsf{B}_C\psi \in \omega_0$ due to the maximality of the set ω_0. Thus, $\overline{\mathsf{N}}\mathsf{B}_C\psi \in \omega_0$ by the definition of modality $\overline{\mathsf{N}}$. Also, $\delta'(a) = s(a) = \psi$ for each $a \in C$. Hence, $\neg\psi \in \omega'$ by Definition 6 and the assumption $(\delta', \omega') \in P$. Then, $\psi \notin \omega'$ by the consistency of set ω'. Therefore, $(\delta', \omega') \nVdash \psi$ by the induction hypothesis. ⊠

We are now ready to state and prove the strong completeness of our logical system.

Theorem 1 *If $X \nvdash \varphi$, then there is a game, a complete action profile δ, and an outcome ω of this game such that $(\delta, \omega) \Vdash \chi$ for each $\chi \in X$ and $(\delta, \omega) \nVdash \varphi$.*

PROOF. Suppose that $X \nvdash \varphi$. Thus, set $X \cup \{\neg\varphi\}$ is consistent. Let ω_0 be any maximal consistent extension of set $X \cup \{\neg\varphi\}$ and $G(\omega_0) = (\Delta, \Omega, P, \pi)$ be the canonical game defined above. Note that $\omega_0 \in \Omega$ by Definition 4 and the Truth axiom.

By Lemma 15, there exists a complete action profile $\delta \in \Delta^{\mathcal{A}}$ such that $(\delta, \omega_0) \in P$. Thus, $(\delta, \omega_0) \Vdash \chi$ for each $\chi \in X$ and $(\delta, \omega_0) \Vdash \neg\varphi$ by Lemma 17 and the choice of set ω_0. Therefore, $(\delta, \omega_0) \nVdash \varphi$ by Definition 3. ⊠

Conclusion

In this paper we defined a formal semantics of blameworthiness using the principle of alternative possibilities and Marc Pauly's framework for logics of coalitional power. Our main technical result is a sound and complete bimodal logical system that captures properties of blameworthiness in this setting. This work is meant to be a step towards formal reasoning about blameworthiness and responsibility.

Recently, there have been several works combining Marc Pauly's and epistemic logic frameworks to study the interplay between knowledge and know-how strategies (Ågotnes and Alechina 2012; 2016; Naumov and Tao 2017; 2018b; 2018c; 2018a) as well as a study of such strategies in a single-agent case (Fervari et al. 2017). Knowledge is clearly relevant to the study of blameworthiness. Indeed, one can hardly be blamed for not preventing an outcome if one had a strategy to prevent it but did not know what this strategy was. Furthermore, in the legal domain, responsibility is connected to knowledge. For example, US Model Penal Code specifies five types of responsibility based on what the responsible party knew or should have known (Institute 1985 Print). In the future, we plan to explore the interplay between knowledge and blameworthiness/responsibility by introducing epistemic component to the framework of this paper.

References

Ågotnes, T., and Alechina, N. 2012. Epistemic coalition logic: completeness and complexity. In *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems-Volume 2 (AAMAS)*, 1099–1106.

Ågotnes, T., and Alechina, N. 2016. Coalition logic with individual, distributed and common knowledge. *Journal of Logic and Computation*. exv085.

Ågotnes, T.; Balbiani, P.; van Ditmarsch, H.; and Seban, P. 2010. Group announcement logic. *Journal of Applied Logic* 8(1):62 – 81.

Ågotnes, T.; van der Hoek, W.; and Wooldridge, M. 2009. Reasoning about coalitional games. *Artificial Intelligence* 173(1):45 – 79.

Alechina, N.; Logan, B.; Nguyen, H. N.; and Rakib, A. 2011. Logic for coalitions with bounded resources. *Journal of Logic and Computation* 21(6):907–937.

Alechina, N.; Halpern, J. Y.; and Logan, B. 2017. Causality, responsibility and blame in team plans. In *Proceedings of the 16th Conference on Autonomous Agents and MultiAgent Systems*, 1091–1099. International Foundation for Autonomous Agents and Multiagent Systems.

Alur, R.; Henzinger, T. A.; and Kupferman, O. 2002. Alternating-time temporal logic. *Journal of the ACM* 49(5):672–713.

Batusov, V., and Soutchanski, M. 2018. Situation calculus semantics for actual causality. In *Proceedings of the 32nd AAAI Conference on Artificial Intelligence (AAAI-18)*.

Belardinelli, F. 2014. Reasoning about knowledge and strategies: Epistemic strategy logic. In *Proceedings 2nd International Workshop on Strategic Reasoning, Grenoble, France, April 5-6, 2014*, volume 146 of *EPTCS*, 27–33.

Borgo, S. 2007. Coalitions in action logic. In *20th International Joint Conference on Artificial Intelligence*, 1822–1827.

Broersen, J.; Herzig, A.; and Troquard, N. 2009. What groups do, can do, and know they can do: an analysis in normal modal logics. *Journal of Applied Non-Classical Logics* 19(3):261–289.

Cushman, F. 2015. Deconstructing intent to reconstruct morality. *Current Opinion in Psychology* 6:97–103.

Fervari, R.; Herzig, A.; Li, Y.; and Wang, Y. 2017. Strategically knowing how. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI-17*, 1031–1038.

Fields, L. 1994. Moral beliefs and blameworthiness: Introduction. *Philosophy* 69(270):397–415.

Fischer, J. M., and Ravizza, M. 2000. *Responsibility and control: A theory of moral responsibility*. Cambridge University Press.

Frankfurt, H. G. 1969. Alternate possibilities and moral responsibility. *The Journal of Philosophy* 66(23):829–839.

Galimullin, R., and Alechina, N. 2017. Coalition and group announcement logic. In *Proceedings Sixteenth Conference on Theoretical Aspects of Rationality and Knowledge (TARK) 2017, Liverpool, UK, 24-26 July 2017*, 207–220.

Gant, M. 2018. Millennials being blamed for decline of American cheese. *Fox News*. www.foxnews.com/food-drink/millennials-kraft-american-cheese-sales-decline.amp.

Goranko, V., and Enqvist, S. 2018. Socially friendly and group protecting coalition logics. In *Proceedings of the 17th International Conference on Autonomous Agents and Multi-Agent Systems*, 372–380. International Foundation for Autonomous Agents and Multiagent Systems.

Goranko, V.; Jamroga, W.; and Turrini, P. 2013. Strategic games and truly playable effectivity functions. *Autonomous Agents and Multi-Agent Systems* 26(2):288–314.

Goranko, V. 2001. Coalition games and alternating temporal logics. In *Proceedings of the 8th conference on Theoretical aspects of rationality and knowledge*, 259–272. Morgan Kaufmann Publishers Inc.

Goudkamp, J. 2004. The spurious relationship between moral blameworthiness and liability for negligence. *Melb. UL Rev.* 28:343.

Halpern, J. Y., and Kleiman-Weiner, M. 2018. Towards formal definitions of blameworthiness, intention, and moral responsibility. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence (AAAI-18)*.

Halpern, J. Y. 2016. *Actual causality*. MIT Press.

Halpern, J. Y. 2017. *Reasoning about uncertainty*. MIT press.

Institute, A. L. 1985 Print. *Model Penal Code: Official Draft and Explanatory Notes. Complete Text of Model Penal Code as Adopted at the 1962 Annual Meeting of the American Law Institute at Washington, D.C., May 24, 1962*. The Institute.

Juarez, L., and Miracle, V. 2018. Toddler dies in Muscoy shooting after 4-year-old cousin gets hold of gun; grandfather arrested. *KABC*. http://abc7.com/4-year-old-shoots-kills-toddler-cousin-in-ie;-grandpa-arrested/3794943/.

Mason, E. 2015. Moral ignorance and blameworthiness. *Philosophical Studies* 172(11):3037–3057.

Naumov, P., and Tao, J. 2017. Coalition power in epistemic transition systems. In *Proceedings of the 2017 International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, 723–731.

Naumov, P., and Tao, J. 2018a. Second-order know-how strategies. In *Proceedings of the 2018 International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, 390–398.

Naumov, P., and Tao, J. 2018b. Strategic coalitions with perfect recall. In *Proceedings of Thirty-Second AAAI Conference on Artificial Intelligence*.

Naumov, P., and Tao, J. 2018c. Together we know how to achieve: An epistemic logic of know-how. *Artificial Intelligence* 262:279 – 300.

Nichols, S., and Knobe, J. 2007. Moral responsibility and determinism: The cognitive science of folk intuitions. *Nous* 41(4):663–685.

Oreskes, B. 2018. 4-year-old accidentally shoots and kills toddler cousin in San Bernardino County. *Los Angeles Times*. http://www.latimes.com/local/lanow/la-me-ln-muscoy-toddler-shooting-20180720-story.html.

Pauly, M. 2001. *Logic for Social Software*. Ph.D. Dissertation, Institute for Logic, Language, and Computation.

Pauly, M. 2002. A modal logic for coalitional power in games. *Journal of Logic and Computation* 12(1):149–166.

Sauro, L.; Gerbrandy, J.; van der Hoek, W.; and Wooldridge, M. 2006. Reasoning about action and cooperation. In *Proceedings of the Fifth International Joint Conference on Autonomous Agents and Multiagent Systems*, AAMAS '06, 185–192. New York, NY, USA: ACM.

Singer, P., and Eddon, M. 2013. Moral responsibility, problem of. *Encyclopædia Britannica*. https://www.britannica.com/topic/problem-of-moral-responsibility.

van der Hoek, W., and Wooldridge, M. 2005. On the logic of cooperation and propositional control. *Artificial Intelligence* 164(1):81 – 119.

Widerker, D. 2017. *Moral responsibility and alternative possibilities: Essays on the importance of alternative possibilities*. Routledge.

Xu, M. 1998. Axioms for deliberative stit. *Journal of Philosophical Logic* 27(5):505–552.