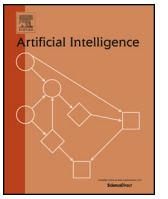
Artificial Intelligence 283 (2020) 103269

Contents lists available at ScienceDirect

Artificial Intelligence

www.elsevier.com/locate/artint



An epistemic logic of blameworthiness

Pavel Naumov [a], Jia Tao [b,*]

[a] Tulane University, New Orleans, LA, USA
[b] Lafayette College, Easton, PA, USA

ARTICLE INFO

Article history:
Received 25 March 2019
Received in revised form 5 February 2020
Accepted 21 March 2020
Available online 26 March 2020

Keywords:
Logic
Blameworthiness
Responsibility
Knowledge
Strategies
Know-how
Axiomatization
Completeness

ABSTRACT

Blameworthiness of an agent or a coalition of agents can be defined in terms of the principle of alternative possibilities: for the coalition to be responsible for an outcome, the outcome must take place and the coalition should be a minimal one that had a strategy to prevent the outcome. In this article we argue that in the settings with imperfect information, not only should the coalition have had a strategy, but it also should be the minimal one that knew that it had a strategy and what the strategy was.

The main technical result of the article is a sound and complete bimodal logic that describes the interplay between knowledge and blameworthiness in strategic games with imperfect information.

1. Introduction

In this article we study blameworthiness of agents and their coalitions in multiagent systems. Throughout centuries, blameworthiness, especially in the context of free will and moral responsibility, has been at the focus of philosophical discussions [1]. These discussions continue in the modern time [2–6]. Frankfurt 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" [7]. 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 [6], we refer to this principle as the principle of *alternative* possibilities.

The principle of alternative possibilities, sometimes referred to as "counterfactual possibility" [8], is also used to define causality [9–11]. Halpern and Kleiman-Weiner used a similar setting to define *degrees* of blameworthiness [12]. Alechina, Halpern, and Logan applied counterfactual definition of causality to team plans [13]. In [14], we proposed a logical system that describes properties of coalition blameworthiness in strategic games as a modal operator whose semantics is also based on the principle of alternative possibilities.

Although the principle of alternative possibilities makes sense in the settings with perfect information, it needs to be adjusted for settings with imperfect information. Indeed, consider a traffic situation depicted in Fig. 1. A self-driving truck t and a regular car c are approaching an intersection at which truck t must stop to yield to car c. The truck is experiencing

* Corresponding author.
E-mail addresses: pgn2@cornell.edu (P. Naumov), taoj@lafayette.edu (J. Tao).

https://doi.org/10.1016/j.artint.2020.103269
0004-3702/

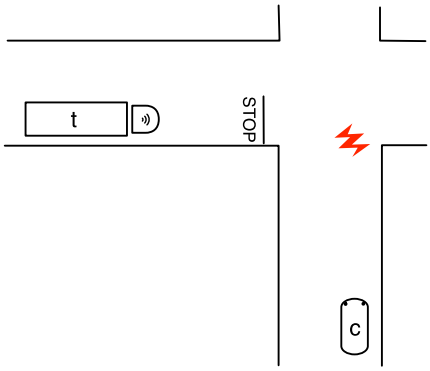


Fig. 1. A traffic situation.

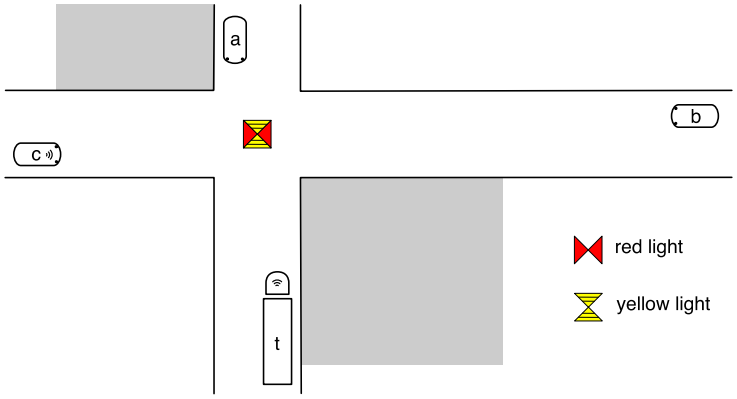


Fig. 2. A more involved traffic situation. (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

a sudden brake failure and it cannot stop, nor can it slow down at the intersection. The truck turns on flashing lights and sends distress signals to other self-driving cars by radio. The driver of car c can see the flashing lights, but she does not receive the radio signal. She can also observe that the truck does not slow down. The driver of car c has two potential strategies to avoid a collision with the truck: to slow down or to accelerate. The driver understands that one of these two strategies will succeed, but since she does not know the exact speed of the truck, she does not know which of the two strategies will succeed. Suppose that the collision could be avoided if the car accelerates, but the car driver decides to slow down. The vehicles collide. According to the principle of alternative possibilities, the driver of the car is responsible for the collision because she had a strategy to avoid the collision but did not use it.

It is not likely, however, that a court will find the driver of car c responsible for the accident. For example, US Model Penal Code [15] distinguishes different forms of legal liability as different combinations of "guilty actions" and "guilty mind". The situation in our example falls under strict liability (pure "guilty actions" without an accompanied "guilty mind"). In many situations, strict liability does not lead to legal liability.

In this article we propose a formal semantics of blameworthiness in strategic games with imperfect information. According to this semantics, an agent is blamable for φ if φ is true and the agent *knew how* to prevent φ. In our example, since the driver of the car does not know that she must accelerate in order to avoid the collision, she cannot be blamed for the collision. We write this as: $\neg \mathsf{B}_c$("Vehicles collided."). Now, consider a similar traffic situation in which car c is a self-driving vehicle. The car receives the distress signal from truck t, which contains the truck's exact speed. From this information, car c determines that it can avoid the collision if it accelerates. However, if the car slows down, then the vehicles collide and the self-driving car c is blameable for the collision: B_c("Vehicles collided.").

As another example, consider the road situation depicted in Fig. 2. Here, regular cars a and b as well as self-driving car c and self-driving truck t are approaching the intersection. The light is red for cars c and b. The light just turned yellow for car a and truck t. Under normal circumstances, car a and truck t would drive through the intersection on yellow light while car c would stop and wait for the light to turn green. Car b which is sufficiently far away from the intersection would maintain a constant speed. Thus, in the normal situation vehicles would cross the intersection in the following order: $(a, t, *, b, c)$, where symbol $*$ represents the change of traffic lights.

Imagine now that car c is experiencing a sudden brake failure and truck t does not have enough time to stop before the intersection. Thus, car c risks to collide with the truck t. However, through cooperation by radio, self-driving vehicles

c and t could devise the following strategy to avoid the collision: car c will accelerate and truck t will slow-down so that car c could pass the intersection before truck t. Under this strategy, vehicles would cross the intersection in the order: $(a, c, *, t, b)$. Note that according to this plan, both car c and truck t will cross the intersection on red light. By doing this, they may collide with other vehicles crossing the intersection. In our case, car c might collide with car a and truck t might collide with car b. To avoid these collisions, car c must not accelerate *too much* so that it will let car a pass before it. Also, truck t should not slow-down *too much* so that it can pass the intersection before car b. Hence, to execute the plan safely, car c must not only be aware of the existence of car a, but also know a's precise speed. Similarly, truck t must be aware of car b and also know b's precise speed. Neither of vehicles c and t individually has the necessary information because their views are obstructed by buildings marked by gray rectangles in Fig. 2. Thus, neither of them has a know-how strategy to prevent the collisions. However, car c can see car b and use a radar to measure b's speed. Similarly, truck t can see car a and use a radar to measure a's speed. Thus, car c and truck t have a *distributed* knowledge of how to prevent the collisions. If a collision happens, the coalition consisting of car c and truck t should be blamed for it:

$$\mathsf{B}_{c,t}(\text{“A collision happened”}).$$

In addition to replacing “could have prevented” with “knew how they could have prevented”, the current work adds one more significant refinement to the definition of blameworthiness. Namely, we require that to be blamable, the coalition must be a *minimal* one that knew how it could have prevented. For example, the coalition of four vehicles $\{a, b, c, t\}$ also has a distributively known strategy to prevent collisions in Fig. 2, but we do not blame it

$$\neg\mathsf{B}_{a,b,c,t}(\text{“A collision happened”}),$$

because it is not a minimal such coalition.

The main technical result of this article is a bimodal logical system that describes the interplay between knowledge and blameworthiness of coalitions in strategic games with imperfect information.

The article is organized as follows. Section 2 presents the formal syntax and semantics of our logical system. In Section 3 we discuss our formal semantics in relation to the existing literature. Section 4 introduces our axioms, compares them with those in the related works, and proves basic properties of our logical system. In Section 5 we prove the soundness of our system. The proof of the completeness is divided into two steps. First, in Section 6, we define a counterfactual modality “coalition C *had* a strategy to prevent φ” through the knowledge and the blameworthiness modalities. In the same section we also prove a long list of properties of the counterfactual modality. In Section 7 we use these properties to show the completeness of our system. Section 8 concludes.

2. Syntax and semantics

In this article we assume a fixed finite set $\mathcal{A}$ of agents and a fixed set of propositional variables. 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 propositional variable p,
2. $\varphi \to \psi, \neg\varphi \in \Phi$ for all formulae $\varphi, \psi \in \Phi$,
3. $\mathsf{K}_C\varphi$, $\mathsf{B}_C\varphi \in \Phi$ for each coalition $C \subseteq \mathcal{A}$ and each $\varphi \in \Phi$.

In other words, language Φ is defined by grammar:

$$\varphi := p \mid \neg\varphi \mid \varphi \to \varphi \mid \mathsf{K}_C\varphi \mid \mathsf{B}_C\varphi.$$

Formula $\mathsf{K}_C\varphi$ is read as “coalition C distributively knew before the actions were taken that statement φ would be true” 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 $\overline{\mathsf{K}}_C\varphi$ we mean $\neg\mathsf{K}_C\neg\varphi$. 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 K and B is defined in terms of models, which we call *games*. These are one-shot strategic games with imperfect information. We specify the set of actions by all agents, or a *complete action profile*, as a function $\delta \in \Delta^{\mathcal{A}}$ from the set of all agents $\mathcal{A}$ to the set of all actions Δ.

Definition 2. A game is a tuple $(I, \{\sim_a\}_{a\in\mathcal{A}}, \Delta, \Omega, P, \pi)$, where

1. I is a set of “initial states”,
2. $\sim_a$ is an “indistinguishability” equivalence relation on set I,

3. Δ is a nonempty set of "actions",
4. Ω is a set of "outcomes",
5. the set of "plays" P is an arbitrary set of tuples $(\alpha, \delta, \omega) \in I \times \Delta^{\mathcal{A}} \times \Omega$ where for each initial state $\alpha \in I$ and each complete action profile $\delta \in \Delta^{\mathcal{A}}$, there is at least one outcome $\omega \in \Omega$ such that $(\alpha, \delta, \omega) \in P$,
6. π is a function that maps propositional variables into subsets of P.

In the introductory example, the set I has two states *high* and *low*, corresponding to the truck going at a high or low speed, respectively. The driver of the regular car c cannot distinguish these two states while these states can be distinguished by a self-driving version of car c. For the sake of simplicity, assume that there are two actions that car c can take: $\Delta = \{\textit{slow-down}, \textit{speed-up}\}$ and two possible outcomes: $\Omega = \{\textit{collision}, \textit{no collision}\}$. Vehicles collide if either the truck goes with a low speed and the car decides to slow-down or the truck goes with a high speed and the car decides to accelerate. In our case there is only one agent (car c), so the complete action profile can be described by giving just the action of this agent. We refer to the two complete action profiles in this situation simply as profile *slow-down* and profile *speed-up*. The list of all possible scenarios (or "plays") is given by the set

$$P = \{(\textit{high}, \textit{speed-up}, \textit{collision}), (\textit{high}, \textit{slow-down}, \textit{no collision}),$$
$$\{(\textit{low}, \textit{speed-up}, \textit{no collision}), (\textit{low}, \textit{slow-down}, \textit{collision})\}.$$

Note that in our example an initial state and an action profile uniquely determine the outcome. In general, just like in [14], we allow nondeterministic games where this does not have to be true.

Whether statement $\mathsf{B}_C\varphi$ is true or false depends not only on the outcome but also on the initial state of the game. Indeed, coalition C might have known how to prevent φ in one initial state but not in the other. For this reason, we assume that all statements are true or false for a particular play of the game. For example, propositional variable p can stand for "car c slowed down and collided with truck t going at a high speed". As a result, function π in the definition above maps p into subsets of P rather than subsets of Ω.

By an action profile of a coalition C we mean an arbitrary function $s \in \Delta^C$ that assigns an action to each member of the coalition. 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 $a \in C$. We write $\alpha \sim_C \alpha'$ if $\alpha \sim_a \alpha'$ for each $a \in C$. In particular, it means that $\alpha \sim_\varnothing \alpha'$ for any two initial states $\alpha, \alpha' \in I$.

Next is the key definition of this article. Its item 5 formally specifies blameworthiness using the principle of alternative possibilities. In order for a coalition to be blamable for φ, not only must φ be true and the coalition should have had a strategy to prevent φ, but this strategy should work in all initial states that the coalition cannot distinguish from the current state. In other words, the coalition should have known the strategy. Furthermore, we require coalition C to be a minimal such coalition.

Definition 3. For any game $(I, \{\sim_a\}_{a\in\mathcal{A}}, \Delta, \Omega, P, \pi)$, any formula $\varphi \in \Phi$, and any play $(\alpha, \delta, \omega) \in P$, the satisfiability relation $(\alpha, \delta, \omega) \Vdash \varphi$ is defined recursively as follows:

1. $(\alpha, \delta, \omega) \Vdash p$ if $(\alpha, \delta, \omega) \in \pi(p)$, where p is a propositional variable,
2. $(\alpha, \delta, \omega) \Vdash \neg\varphi$ if $(\alpha, \delta, \omega) \nVdash \varphi$,
3. $(\alpha, \delta, \omega) \Vdash \varphi \to \psi$ if $(\alpha, \delta, \omega) \nVdash \varphi$ or $(\alpha, \delta, \omega) \Vdash \psi$,
4. $(\alpha, \delta, \omega) \Vdash \mathsf{K}_C\varphi$ if $(\alpha', \delta', \omega') \Vdash \varphi$ for each play $(\alpha', \delta', \omega') \in P$ such that $\alpha \sim_C \alpha'$,
5. $(\alpha, \delta, \omega) \Vdash \mathsf{B}_C\varphi$ if all of the following conditions hold
   (a) $(\alpha, \delta, \omega) \Vdash \varphi$,
   (b) there is an action profile $s \in \Delta^C$ of coalition C such that for each play $(\alpha', \delta', \omega') \in P$, if $\alpha \sim_C \alpha'$ and $s =_C \delta'$, then $(\alpha', \delta', \omega') \nVdash \varphi$,
   (c) for each proper subset $D \subsetneq C$ and each action profile $s \in \Delta^D$ of coalition D, there is a play $(\alpha', \delta', \omega') \in P$ such that $\alpha \sim_D \alpha'$, $s =_D \delta'$, and $(\alpha', \delta', \omega') \Vdash \varphi$.

3. Discussion

In this section we discuss our formal semantics of blameworthiness and related notions of coalition power, know-how strategy, responsibility, regret, intention, subgame blameworthiness, and counterfactual.

3.1. Coalition power and know-how strategies

The definition of coalition blameworthiness is closely related to Marc Pauly [16,17] coalition power modality and the more recent works on know-how modalities. We write $\alpha \Vdash \mathsf{S}_C\varphi$ if coalition C has a strategy to achieve φ from state α:

$\alpha \Vdash \mathsf{S}_C\varphi$ when there is an action profile $s \in \Delta^C$ of coalition C such that for each play $(\alpha, \delta, \omega) \in P$, if $s =_C \delta$, then $\omega \Vdash \varphi$.

Marc Pauly gave a sound and complete axiomatization of this coalition power modality [16,17]. Note that if the coalition cannot distinguish state α from a state α' and the strategies to achieve φ from these two states, s and s', are different, then coalition C has a strategy to achieve φ in α but does not *know* what this strategy is. If the same strategy could be used to achieve φ in all states indistinguishable from α, then we say that the coalition has a *know-how* strategy to achieve φ:

> $\alpha \Vdash \mathsf{H}_C\varphi$ when there is an action profile $s \in \Delta^C$ of coalition C such that for each play $(\alpha', \delta, \omega) \in P$, if $\alpha \sim_C \alpha'$ and $s =_C \delta$, then $\omega \Vdash \varphi$.

The properties of know-how as a modality have been axiomatized in different settings. Ågotnes and Alechina introduced a complete axiomatization of an interplay between single-agent knowledge and coalition know-how modalities to achieve a goal in one step [18]. A modal logic that combines the distributed knowledge modality with the coalition know-how modality to maintain a goal was axiomatized by us in [19]. A sound and complete logical system in a single-agent setting for know-how strategies to achieve a goal in multiple steps rather than to maintain a goal is developed by Fervari, Herzig, Li, and Wang [20]. In [21,22], we developed a trimodal logical system that describes an interplay between the coalition power modality, the coalition know-how modality, and the distributed knowledge modality. In [23], we proposed a logical system that combines the coalition know-how modality with the distributed knowledge modality in the perfect recall setting. In [24], we introduced a logical system for the second-order know-how. In [25] we proposed a related know-how modality with degrees of uncertainty. Its semantics is using metric spaces instead of indistinguishability equivalence relations. Wang proposed a complete axiomatization of "knowing how" as a binary modality [26,27], but his logical system does not include the knowledge modality.

3.2. States vs. plays

The key difference that sets apart blameworthiness modality from other forms of coalition power modality is its simultaneous reference to two distinct moments in time: a coalition is blamable for a statement if the statement *is* true and the coalition *had* a strategy to prevent it. As a result, it cannot be expressed, for example, through modality $\mathsf{S}_C\varphi$. Indeed, formula $\varphi \wedge \mathsf{S}_C\neg\varphi$, for instance, means that "statement φ is true now and coalition C has a strategy to prevent φ in the future". The blameworthiness modality could be expressed through modality S in combination with a "past" modality P as $\varphi \wedge \mathsf{PS}_C\neg\varphi$. However, the modality P would require a significant change to Marc Pauly semantics because the satisfiability of a formula that uses modality P could only be defined with respect to the whole history of the previous states, not just the current state as in the case of modality S. In other words, $h \Vdash \varphi$ will have to be a relation between a list of all previous states h and a formula φ.

In this article we take an alternative approach of defining $\Vdash$ as a relation between a *transition* and a formula. A transition is formally specified by an initial state $\alpha \in I$, a complete action profile $\delta \in \Delta^{\mathcal{A}}$, and an outcome $\omega \in \Omega$. We define relation $\Vdash$ only for triples (α, δ, ω) that are valid transitions, or *plays*, of our game. Thus, formally, $\Vdash$ is a relation between plays and formulae. In particular, this means that propositional variables are also statements about plays. Such statements could refer to the initial state, the complete action profile, or the outcome. In this article we have chosen to distinguish the set of initial states I from the set of outcomes Ω to keep our presentation more elegant. Alternatively, one can assume that there is no distinction between these two types of states.

3.3. Knowledge and regret

In Definition 2 we assume the existence of an indistinguishability relation on initial states. This relation is used in item 4 of Definition 3 to define the semantics of the knowledge modality K_C. Thus, statement $(\alpha, \delta, \omega) \Vdash \mathsf{K}_C\varphi$ means that coalition C knows φ in initial state α, *before the transition from α to ω takes place*. We call such knowledge *knowledge ex ante*. One could also add an indistinguishability relation $\approx_a$ on the set of outcomes Ω in Definition 2 and then define *knowledge ex post* in Definition 3 as:

6. $(\alpha, \delta, \omega) \Vdash \mathsf{K}_C^{\mathsf{post}}\varphi$ if $(\alpha', \delta', \omega') \Vdash \varphi$ for each play $(\alpha', \delta', \omega') \in P$ such that $\omega \approx_C \omega'$.

The knowledge indirectly incorporated into item 5 of Definition 3 is also *knowledge ex ante* because it uses relation $\sim_C$. Indeed, one could be blamed for φ only if she knew how to prevent φ before the transition took place. On the other hand, the *knowledge ex post* could be used to capture *regret*: an agent regrets φ if she knows that φ is true and she has learned how she could have prevented it. Formally,

7. $(\alpha, \delta, \omega) \Vdash \mathsf{Reg}_C\varphi$ if the following conditions hold
   (a) $(\alpha, \delta, \omega) \Vdash \mathsf{K}_C^{\mathsf{post}}\varphi$,
   (b) there is an action profile $s \in \Delta^C$ of coalition C such that for any play $(\alpha', \delta', \omega') \in P$ and any play $(\alpha', \delta'', \omega'') \in P$, if $\omega \approx_C \omega'$ and $s =_C \delta''$, then $(\alpha', \delta'', \omega'') \nVdash \varphi$.

In the above definition of regret we assumed that the coalition must *learn how it could have* prevented φ. Alternatively, one can impose a weaker requirement that the coalition only needs to *learn that it could have* prevented φ:

7′. $(\alpha,\delta,\omega)\Vdash \mathsf{Reg}^{\mathsf{LS}}_C\varphi$ if the following conditions hold
   (a) $(\alpha,\delta,\omega)\Vdash \mathsf{K}^{\mathsf{post}}_C\varphi$,
   (b) for any play $(\alpha',\delta',\omega')\in P$ there is an action profile $s\in\Delta^C$ of coalition C such that for any play $(\alpha',\delta'',\omega'')\in P$, if $\omega\approx_C\omega'$ and $s=_C\delta''$, then $(\alpha',\delta'',\omega'')\nVdash\varphi$.

The former (our) definition of regret requires to learn the strategy. The latter, proposed by Lorini and Schwarzentruber [28], only requires to learn the existence of a strategy.

Finally, item 4 of Definition 3 defines K_C as *distributed* knowledge modality. The knowledge indirectly incorporated into item 5 of Definition 3 is also distributed. In other words, we assume that if a coalition is blameable for φ, then it should have had distributed knowledge of how to prevent φ, as in our example in Fig. 2.

3.4. Responsibility and blameworthiness

If a coalition had a way to prevent an outcome, but it did not know how to prevent, then the coalition is *responsible, but not blamable* for the outcome. The responsibility modality could be defined by modifying item 5 of Definition 3:

8. $(\alpha,\delta,\omega)\Vdash \mathsf{Resp}_C\varphi$ if all of the following conditions hold
   (a) $(\alpha,\delta,\omega)\Vdash\varphi$,
   (b) there is an action profile $s\in\Delta^C$ of coalition C such that for each play $(\alpha,\delta',\omega')\in P$, if $s=_C\delta'$, then $(\alpha,\delta',\omega')\nVdash\varphi$,
   (c) for each proper subset $D\subsetneq C$ and each action profile $s\in\Delta^D$ of coalition D, there is a play $(\alpha,\delta',\omega')\in P$ such that $s=_D\delta'$ and $(\alpha,\delta',\omega')\Vdash\varphi$.

The difference between responsibility modality Resp and blameworthiness modality B is similar to the difference between Marc Pauly coalition power modality S and know-how modality H. In [22] we studied the interplay between modalities S, H, and K and proposed a new axiom that connects all three modalities: $\mathsf{K}_C\mathsf{S}_\varnothing(\varphi\to\psi)\to(\mathsf{H}_C\varphi\to\mathsf{H}_C\psi)$. Although in the current article we only consider modalities B and K, it might be interesting to combine blameworthiness B, responsibility Resp, and distributed knowledge modality K in a single logical system.

In item 5 of Definition 3, as well as in item 8 above, we require set C to be minimal, which extends our original definition in [14] that does not have the minimality requirement. This requirement is natural in many settings, including the legal one. At the same time, the word "blame" sometimes is used in English without the assumption of minimality. For example, the sentence "Millennials being blamed for decline of American cheese" [29] does not imply that no one in the millennial generation likes American cheese. The initial version of this article included the requirement for the strategy to prevent to be a know-how strategy, but did not include the minimality requirement [30]. Yazdanpanah, Dastani, Jamroga, Alechina, and Logan proposed a definition of the blameworthiness (that they call "backward group responsibility") that combines know-how strategy requirement and the minimality requirement [31]. They do not treat blameworthiness as a modality and do not prove any completeness results.

3.5. Blame for doing intentionally

Xu proposed a complete logical system for reasoning about modality "agent a took an action that forces outcome φ" [32]. Broersen, Herzig, and Troquard extended his system to coalitions [33]. One can add knowledge *ex ante* and a counterfactual to this modality to define modality "coalition C is blamable for *intentionally* achieving φ":

9. $(\alpha,\delta,\omega)\Vdash \mathsf{Int}_C\varphi$ if the following conditions hold
   (a) for each play $(\alpha',\delta',\omega')\in P$, if $\alpha\sim_C\alpha'$ and $\delta=_C\delta'$, then $(\alpha',\delta',\omega')\Vdash\varphi$,
   (b) there is a play $(\alpha',\delta',\omega')\in P$ such that $\alpha\sim_C\alpha'$ and $(\alpha',\delta',\omega')\nVdash\varphi$.

That is, coalition C is blamable for intentionally achieving φ if the coalition thought that $\neg\varphi$ might happen, but it knowingly acted to force φ.

3.6. Subgame blameworthiness

We say that a coalition C is blamable for statement φ in a strategic game if φ is true and C is a minimal coalition that had a know-how strategy to prevent φ. This definition will need to be further refined in the case of extensive games. Indeed, consider the game depicted in Fig. 3. Suppose that first agent a uses action 1, then agent b uses action 1, and finally agent a uses action 1 again. As a result, the game terminates with an outcome in which statement p is true. Note that agent a does not have a strategy to prevent p in this game, but it has such a strategy in the subgame marked by the dashed line in Fig. 3. In other words, agent a was given an opportunity to prevent p in the middle of the game. Since agent a did not

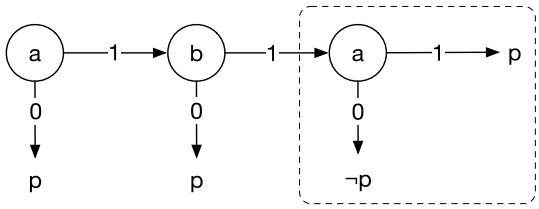


Fig. 3. An Extensive Form Game.

use this opportunity, she should be blamed for statement p being true in the outcome of the game. In [34] we use this adjusted definition of blameworthiness in the spacial case of security games and give a sound and complete axiomatization of all properties of blameworthiness in those games.

3.7. Counterfactual know-how modality

One can modify item 5(b) of Definition 3 to define modality "coalition C *had* a know-how strategy to achieve φ":

10. $(\alpha, \delta, \omega) \Vdash \Box_C \varphi$ when there is an action profile $s \in \Delta^C$ of coalition C such that for each play $(\alpha', \delta', \omega') \in P$, if $\alpha \sim_C \alpha'$ and $s =_C \delta'$, then $(\alpha', \delta', \omega') \Vdash \varphi$.

This is a counterfactual know-how modality because it talks about a past ability. Modality $\Box$ has different properties from know-how modality H. For example, one can show that formula $\mathsf{K}_C\varphi \to \Box_C\varphi$ is universally true under the play-based semantics of Definition 3, while formula $\mathsf{K}_C\varphi \to \mathsf{H}_C\varphi$ is not true for the state-based semantics of know-how modality [22]. The blameworthiness modality B_C could be defined through the counterfactual know-how modality:

$$\mathsf{B}_C\varphi \equiv \varphi \wedge \Box_C\neg\varphi \wedge \bigwedge_{D \subsetneq C} \neg\Box_D\neg\varphi.$$

Perhaps unexpectedly, counterfactual know-how modality could be defined through the blameworthiness and the distributed knowledge modalities:

$$\Box_C\varphi \equiv \mathsf{K}_C\varphi \vee \bigvee_{D \subseteq C} \overline{\mathsf{K}}_D\mathsf{B}_D\neg\varphi. \tag{1}$$

In this article we have chosen B over $\Box$ as our primitive modality because it captures the actual notion, blameworthiness, that we study. However, in Section 7 we use the counterfactual know-how modality $\Box$ as specified by equation (1) to facilitate the proof of the completeness.

4. Axioms

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

1. Truth: $\mathsf{K}_C\varphi \to \varphi$ and $\mathsf{B}_C\varphi \to \varphi$,
2. Distributivity: $\mathsf{K}_C(\varphi \to \psi) \to (\mathsf{K}_C\varphi \to \mathsf{K}_C\psi)$,
3. Negative Introspection: $\neg\mathsf{K}_C\varphi \to \mathsf{K}_C\neg\mathsf{K}_C\varphi$,
4. Monotonicity: $\mathsf{K}_C\varphi \to \mathsf{K}_D\varphi$, where $C \subseteq D$,
5. Minimality: $\mathsf{B}_C\varphi \to \mathsf{K}_C\neg\mathsf{B}_D\varphi$, where $C \subsetneq D$,
6. None to Blame: $\neg\mathsf{B}_\varnothing\varphi$,
7. Joint Blameworthiness:

$$\overline{\mathsf{K}}_C\mathsf{B}_C\varphi \wedge \overline{\mathsf{K}}_D\mathsf{B}_D\psi \to \left(\varphi \vee \psi \to \bigvee_{E \subseteq C \cup D} \mathsf{B}_E(\varphi \vee \psi)\right),$$

   where $C \cap D = \varnothing$,
8. Knowledge and Blameworthiness:

$$\mathsf{K}_C(\varphi \to \psi) \wedge \overline{\mathsf{K}}_D\mathsf{B}_D\psi \to \left(\varphi \to \bigvee_{E \subseteq C \cup D} \mathsf{B}_E\varphi\right).$$

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

$$\frac{\varphi, \varphi \to \psi}{\psi}, \qquad \frac{\varphi}{\mathsf{K}_C\varphi}, \qquad \frac{\varphi \leftrightarrow \psi}{\mathsf{B}_C\varphi \to \mathsf{B}_C\psi}.$$

We write $X \vdash \varphi$ if formula $\varphi \in \Phi$ is provable from the theorems of our logical system and an additional set of axioms X using only the Modus Ponens inference rule. Note that if set X is empty, then statement $X \vdash \varphi$ is equivalent to $\vdash \varphi$. We say that set X is consistent if $X \nvdash \bot$.

The Truth, the Distributivity, the Negative Introspection, and the Monotonicity axioms for epistemic modality K are the standard S5 axioms from the logic of distributed knowledge. The Truth axiom for blameworthiness modality B states that a coalition could only be blamed for something true. The Minimality axiom captures the minimality condition 5(c) of Definition 3. The None to Blame axiom says that an empty coalition can be blamed for nothing. The remaining three axioms describe the interplay between knowledge and blameworthiness modalities. The Joint Blameworthiness axiom says that if a coalition C cannot exclude a possibility of being blamable for φ, a coalition D cannot exclude a possibility of being blamable for ψ, and the disjunction $\varphi \vee \psi$ is true, then a subset of the joint coalition $C \cup D$ is blamable for the disjunction.

Note that if $\mathsf{K}_C(\varphi \to \psi)$ is true and coalition C has a know-how strategy to prevent ψ, then the same strategy is also a know-how strategy to prevent φ. However, it is not true that the coalition C should be blamed for φ if it can be blamed for ψ because (a) φ might not be true (b) C might not be the minimal coalition that has a know-how strategy to prevent φ. The Knowledge and Blameworthiness axiom states that if $\mathsf{K}_C(\varphi \to \psi)$ is true, coalition C is blamable for ψ, and φ is true, then a *subcoalition* of C is blamable for φ.

The next lemma is an example of a proof in our logical system. This lemma will later be used in the proof of the completeness.

Lemma 1. $\vdash \mathsf{B}_D\varphi \to \mathsf{K}_C\neg\mathsf{B}_C\varphi$, *where* $C \subsetneq D$.

Proof. Note that $\mathsf{K}_C\neg\mathsf{B}_D\varphi \to \neg\mathsf{B}_D\varphi$ is an instance of the Truth axiom. Thus, $\vdash \mathsf{B}_D\varphi \to \neg\mathsf{K}_C\neg\mathsf{B}_D\varphi$ by the contraposition. Hence, by the Negative Introspection axiom and propositional reasoning,

$$\vdash \mathsf{B}_D\varphi \to \mathsf{K}_C\neg\mathsf{K}_C\neg\mathsf{B}_D\varphi. \tag{2}$$

At the same time, $\mathsf{B}_C\varphi \to \mathsf{K}_C\neg\mathsf{B}_D\varphi$ is an instance of the Minimality axiom. Thus, $\vdash \neg\mathsf{K}_C\neg\mathsf{B}_D\varphi \to \neg\mathsf{B}_C\varphi$ by the contraposition. Then, by the Necessitation inference rule $\vdash \mathsf{K}_C(\neg\mathsf{K}_C\neg\mathsf{B}_D\varphi \to \neg\mathsf{B}_C\varphi)$. Hence, by the Distributivity axiom and the Modus Ponens rule, $\vdash \mathsf{K}_C\neg\mathsf{K}_C\neg\mathsf{B}_D\varphi \to \mathsf{K}_C\neg\mathsf{B}_C\varphi$. Therefore, $\vdash \mathsf{B}_D\varphi \to \mathsf{K}_C\neg\mathsf{B}_C\varphi$ by proposition reasoning using statement (2). ⊠

The following lemma states a well-known positive introspection principle.

Lemma 2. $\vdash \mathsf{K}_C\varphi \to \mathsf{K}_C\mathsf{K}_C\varphi$.

Proof. Formula $\mathsf{K}_C\neg\mathsf{K}_C\varphi \to \neg\mathsf{K}_C\varphi$ is an instance of the Truth axiom. Thus, $\vdash \mathsf{K}_C\varphi \to \neg\mathsf{K}_C\neg\mathsf{K}_C\varphi$ by contraposition. Hence, taking into account the following instance of the Negative Introspection axiom: $\neg\mathsf{K}_C\neg\mathsf{K}_C\varphi \to \mathsf{K}_C\neg\mathsf{K}_C\neg\mathsf{K}_C\varphi$, we have

$$\vdash \mathsf{K}_C\varphi \to \mathsf{K}_C\neg\mathsf{K}_C\neg\mathsf{K}_C\varphi. \tag{3}$$

At the same time, $\neg\mathsf{K}_C\varphi \to \mathsf{K}_C\neg\mathsf{K}_C\varphi$ is an instance of the Negative Introspection axiom. Thus, $\vdash \neg\mathsf{K}_C\neg\mathsf{K}_C\varphi \to \mathsf{K}_C\varphi$ by the law of contrapositive in the propositional logic. Hence, by the Necessitation inference rule, $\vdash \mathsf{K}_C(\neg\mathsf{K}_C\neg\mathsf{K}_C\varphi \to \mathsf{K}_C\varphi)$. Thus, by the Distributivity axiom and the Modus Ponens inference rule, $\vdash \mathsf{K}_C\neg\mathsf{K}_C\neg\mathsf{K}_C\varphi \to \mathsf{K}_C\mathsf{K}_C\varphi$. The latter, together with statement (3), implies the statement of the lemma by propositional reasoning. ⊠

Next, we state the deduction and Lindenbaum lemmas for our logical system. These lemmas are used later in the proof of the completeness.

Lemma 3 *(deduction). If* $X, \varphi \vdash \psi$, *then* $X \vdash \varphi \to \psi$.

Proof. Suppose that sequence $\psi_1, \dots, \psi_n$ is a proof from set $X \cup \{\varphi\}$ and the theorems of our logical system that uses the Modus Ponens inference rule only. In other words, for each $k \le n$, either

1. $\vdash \psi_k$, or
2. $\psi_k \in X$, or
3. ψ_k is equal to φ, or
4. there are $i, j < k$ such that formula ψ_j is equal to $\psi_i \to \psi_k$.

It suffices to show that $X \vdash \varphi \to \psi_k$ for each $k \leq n$. We prove this by induction on k through considering the four cases above separately.

Case I: $\vdash \psi_k$. Note that $\psi_k \to (\varphi \to \psi_k)$ is a propositional tautology, and thus, is an axiom of our logical system. Hence, $\vdash \varphi \to \psi_k$ by the Modus Ponens inference rule. Therefore, $X \vdash \varphi \to \psi_k$.

Case II: $\psi_k \in X$. Note again that $\psi_k \to (\varphi \to \psi_k)$ is a propositional tautology, and thus, is an axiom of our logical system. Therefore, by the Modus Ponens inference rule, $X \vdash \varphi \to \psi_k$.

Case III: formula ψ_k is equal to φ. Thus, $\varphi \to \psi_k$ is a propositional tautology. Therefore, $X \vdash \varphi \to \psi_k$.

Case IV: formula ψ_j is equal to $\psi_i \to \psi_k$ for some $i, j < k$. Thus, by the induction hypothesis, $X \vdash \varphi \to \psi_i$ and $X \vdash \varphi \to (\psi_i \to \psi_k)$. Note that formula $(\varphi \to \psi_i) \to ((\varphi \to (\psi_i \to \psi_k)) \to (\varphi \to \psi_k))$ is a propositional tautology. Therefore, $X \vdash \varphi \to \psi_k$ by applying the Modus Ponens inference rule twice. ⊠

Note that it is important for the above proof that $X \vdash \varphi$ stands for derivability only using the Modus Ponens inference rule. For example, if the Necessitation inference rule is allowed, then the proof will have to include one more case where ψ_k is formula $\mathsf{K}_C\psi_i$ for some coalition $C \subseteq \mathcal{A}$, and some integer $i < k$. In this case we will need to prove that if $X \vdash \varphi \to \psi_i$, then $X \vdash \varphi \to \mathsf{K}_C\psi_i$, which, in general, is not true.

Lemma 4. *If $\varphi_1, \dots, \varphi_n \vdash \psi$, then $\mathsf{K}_C\varphi_1, \dots, \mathsf{K}_C\varphi_n \vdash \mathsf{K}_C\psi$.*

Proof. By Lemma 3 applied n times, assumption $\varphi_1, \dots, \varphi_n \vdash \psi$ implies that $\vdash \varphi_1 \to (\varphi_2 \to \dots (\varphi_n \to \psi) \dots)$. Thus, by the Necessitation inference rule,

$$\vdash \mathsf{K}_C(\varphi_1 \to (\varphi_2 \to \dots (\varphi_n \to \psi) \dots)).$$

Hence, by the Distributivity axiom and the Modus Ponens rule,

$$\vdash \mathsf{K}_C\varphi_1 \to \mathsf{K}_C(\varphi_2 \to \dots (\varphi_n \to \psi) \dots).$$

Then, again by the Modus Ponens rule,

$$\mathsf{K}_C\varphi_1 \vdash \mathsf{K}_C(\varphi_2 \to \dots (\varphi_n \to \psi) \dots).$$

Therefore, $\mathsf{K}_C\varphi_1, \dots, \mathsf{K}_C\varphi_n \vdash \mathsf{K}_C\psi$ by applying the previous steps $(n-1)$ more times. ⊠

Lemma 5 *(Lindenbaum). Any consistent set of formulae can be extended to a maximal consistent set of formulae.*

Proof. The standard proof of Lindenbaum's lemma applies here [35, Proposition 2.14]. ⊠

5. Soundness

The epistemic part of the Truth axiom as well as the Distributivity, the Negative Introspection, and the Monotonicity axioms are the standard axioms of epistemic logic S5 for distributed knowledge. Their soundness follows in the standard way [36] from the assumption that $\sim_a$ is an equivalence relation and the fact that the intersection of equivalence relations is also an equivalence relation. The soundness of the blameworthiness part of the Truth axiom immediately follows from item 5(a) of Definition 3. In this section, we prove the soundness of each of the remaining axioms as a separate lemma. In these lemmas, $C, D \subseteq \mathcal{A}$ are coalitions, $\varphi, \psi \in \Phi$ are formulae, and $(\alpha, \delta, \omega) \in P$ is a play of a game $(I, \{\sim_a\}_{a\in\mathcal{A}}, \Delta, \Omega, P, \pi)$.

Lemma 6. *If $(\alpha, \delta, \omega) \Vdash \mathsf{B}_D\varphi$, then $(\alpha, \delta, \omega) \Vdash \mathsf{K}_D\neg\mathsf{B}_C\varphi$, where $D \subsetneq C$.*

Proof. Consider any play $(\alpha', \delta', \omega') \in P$ such that $\alpha \sim_D \alpha'$. By item 4 of Definition 3, it suffices to show that $(\alpha', \delta', \omega') \Vdash \neg\mathsf{B}_C\varphi$.

By item 5(b) of Definition 3, assumption $(\alpha, \delta, \omega) \Vdash \mathsf{B}_D\varphi$ implies that there is an action profile $s \in \Delta^D$ of coalition D such that for each play $(\alpha'', \delta'', \omega'') \in P$, if $\alpha \sim_D \alpha''$ and $s =_D \delta''$, then $(\alpha'', \delta'', \omega'') \nVdash \varphi$. Thus, because $\alpha \sim_D \alpha'$, for each play $(\alpha'', \delta'', \omega'') \in P$, if $\alpha' \sim_D \alpha''$ and $s =_D \delta''$, then $(\alpha'', \delta'', \omega'') \nVdash \varphi$. Therefore, $(\alpha', \delta', \omega') \Vdash \neg\mathsf{B}_C\varphi$ by item 5(c) of Definition 3 and the assumption $D \subsetneq C$ of the lemma. ⊠

Lemma 7. $(\alpha, \delta, \omega) \nVdash \mathsf{B}_\varnothing\varphi$.

Proof. Assume that $(\alpha,\delta,\omega)\Vdash \mathsf{B}_\varnothing\varphi$. Hence, by item 5(a) of Definition 3, we have $(\alpha,\delta,\omega)\Vdash\varphi$ and by item 5(b) of Definition 3, there is an action profile $s\in\Delta^\varnothing$ such that for each play $(\alpha',\delta',\omega')\in P$, if $\alpha\sim_\varnothing\alpha'$ and $s=_\varnothing\delta'$, then $(\alpha',\delta',\omega')\nVdash\varphi$.

Let $\alpha'=\alpha$, $\delta'=\delta$, and $\omega'=\omega$. Since $\alpha\sim_\varnothing\alpha'$ and $s=_\varnothing\delta'$, by the choice of action profile s we have $(\alpha',\delta',\omega')\nVdash\varphi$. Then, $(\alpha,\delta,\omega)\nVdash\varphi$, which leads to a contradiction. ⊠

Lemma 8. *If $C\cap D=\varnothing$, $(\alpha,\delta,\omega)\Vdash\overline{\mathsf{K}}_C\mathsf{B}_C\varphi$, $(\alpha,\delta,\omega)\Vdash\overline{\mathsf{K}}_D\mathsf{B}_D\psi$, and $(\alpha,\delta,\omega)\Vdash\varphi\vee\psi$, then there is a set $E\subseteq C\cup D$ such that $(\alpha,\delta,\omega)\Vdash\mathsf{B}_E(\varphi\vee\psi)$.*

Proof. Suppose that $(\alpha,\delta,\omega)\Vdash\overline{\mathsf{K}}_C\mathsf{B}_C\varphi$ and $(\alpha,\delta,\omega)\Vdash\overline{\mathsf{K}}_D\mathsf{B}_D\psi$. Hence, by item 2 and item 4 of Definition 3 and the definition of modality $\overline{\mathsf{K}}$, there are plays $(\alpha_1,\delta_1,\omega_1)\in P$ and $(\alpha_2,\delta_2,\omega_2)\in P$ such that $\alpha\sim_C\alpha_1$, $\alpha\sim_D\alpha_2$, $(\alpha_1,\delta_1,\omega_1)\Vdash\mathsf{B}_C\varphi$ and $(\alpha_2,\delta_2,\omega_2)\Vdash\mathsf{B}_D\psi$.

Statement $(\alpha_1,\delta_1,\omega_1)\Vdash\mathsf{B}_C\varphi$, by item 5(b) of Definition 3, implies that there is a profile $s_1\in\Delta^C$ such that for each play $(\alpha',\delta',\omega')\in P$, if $\alpha_1\sim_C\alpha'$ and $s_1=_C\delta'$, then $(\alpha',\delta',\omega')\nVdash\varphi$.

Similarly, statement $(\alpha_2,\delta_2,\omega_2)\Vdash\mathsf{B}_D\psi$, by item 5(b) of Definition 3, implies that there is an action profile $s_2\in\Delta^D$ such that for each play $(\alpha',\delta',\omega')\in P$, if $\alpha_2\sim_D\alpha'$ and $s_2=_D\delta'$, then $(\alpha',\delta',\omega')\nVdash\psi$.

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

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

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 $(\alpha',\delta',\omega')\in P$, if $\alpha\sim_{C\cup D}\alpha'$ and $s=_{C\cup D}\delta'$, then $(\alpha',\delta',\omega')\nVdash\varphi$ and $(\alpha',\delta',\omega')\nVdash\psi$. Thus, if $\alpha\sim_{C\cup D}\alpha'$ and $s=_{C\cup D}\delta'$, then $(\alpha',\delta',\omega')\nVdash\varphi\vee\psi$, for each play $(\alpha',\delta',\omega')\in P$. Let E be a minimal subset of $C\cup D$ such that if $\alpha\sim_E\alpha'$ and $s=_E\delta'$, then $(\alpha',\delta',\omega')\nVdash\varphi\vee\psi$, for each play $(\alpha',\delta',\omega')\in P$. Such a subset E exists because the set of all agents $\mathcal{A}$ is finite. Therefore, $(\alpha,\delta,\omega)\Vdash\mathsf{B}_E(\varphi\vee\psi)$ by item 5 of Definition 3 and the assumption $(\alpha,\delta,\omega)\Vdash\varphi\vee\psi$ of the lemma. ⊠

Lemma 9. *If $(\alpha,\delta,\omega)\Vdash\mathsf{K}_C(\varphi\to\psi)$, $(\alpha,\delta,\omega)\Vdash\overline{\mathsf{K}}_D\mathsf{B}_D\psi$, and $(\alpha,\delta,\omega)\Vdash\varphi$, then there is a set $E\subseteq C\cup D$ such that $(\alpha,\delta,\omega)\Vdash\mathsf{B}_E\varphi$.*

Proof. Suppose $(\alpha,\delta,\omega)\Vdash\overline{\mathsf{K}}_D\mathsf{B}_D\psi$. Then, by definition of modality $\overline{\mathsf{K}}$ and items 2 and 4 of Definition 3 there is a play $(\alpha',\beta',\omega')\in P$ such that $\alpha\sim_D\alpha'$ and $(\alpha',\beta',\omega')\Vdash\mathsf{B}_D\psi$.

Thus, by item 5(c) of Definition 3, there is an action profile $s\in\Delta^D$ of coalition D such that for each play $(\alpha'',\delta'',\omega'')\in P$, if $\alpha'\sim_D\alpha''$ and $s=_C\delta''$, then $(\alpha'',\delta'',\omega'')\nVdash\psi$. Recall that $\alpha\sim_D\alpha'$. Hence, for each play $(\alpha'',\delta'',\omega'')\in P$, if $\alpha\sim_D\alpha''$ and $s=_C\delta''$, then $(\alpha'',\delta'',\omega'')\nVdash\psi$. At the same time, assumption $(\alpha,\delta,\omega)\Vdash\mathsf{K}_C(\varphi\to\psi)$ implies that for each play $(\alpha'',\delta'',\omega'')\in P$, if $\alpha\sim_C\alpha''$, then $(\alpha'',\delta'',\omega'')\Vdash\varphi\to\psi$. Then, from the last two statements, for each play $(\alpha'',\delta'',\omega'')\in P$, if $\alpha\sim_{C\cup D}\alpha''$ and $s=_{C\cup D}\delta''$, then $(\alpha'',\delta'',\omega'')\nVdash\psi$ and $(\alpha'',\delta'',\omega'')\Vdash\varphi\to\psi$. Hence, by item 3 of Definition 3, for each play $(\alpha'',\delta'',\omega'')\in P$, if $\alpha\sim_{C\cup D}\alpha''$ and $s=_{C\cup D}\delta''$, then $(\alpha'',\delta'',\omega'')\nVdash\varphi$. Let set E be a minimal subset of $C\cup D$ such that for each play $(\alpha'',\delta'',\omega'')\in P$, if $\alpha\sim_E\alpha''$ and $s=_E\delta''$, then $(\alpha'',\delta'',\omega'')\nVdash\varphi$. Such a subset exists due to the assumption that the set of all agents $\mathcal{A}$ is finite. Therefore, $(\alpha,\delta,\omega)\Vdash\mathsf{B}_E\varphi$ by item 5 of Definition 3 and the assumption $(\alpha,\delta,\omega)\Vdash\varphi$ of the lemma. ⊠

6. Counterfactual know-how modality

As discussed in Section 3.7, for any coalition C and any formula φ, by $\Box_C\varphi$ we mean formula

$$\mathsf{K}_C\varphi\vee\bigvee_{D\subseteq C}\overline{\mathsf{K}}_D\mathsf{B}_D\neg\varphi.$$

Informally, formula $\Box_C\varphi$ means that either formula φ was known to coalition C to be true, or there is a subset D of coalition C that was not able to exclude the possibility that it will be blamed for $\neg\varphi$. Although this is not required for our proof of the completeness, it is relatively easy to prove the following:

Lemma 10. *$(\alpha,\delta,\omega)\Vdash\Box_C\varphi$ iff there is an action profile $s\in\Delta^C$ such that for each play $(\alpha',\delta',\omega')$, if $\alpha\sim_C\alpha'$ and $s=_C\delta'$, then $(\alpha',\delta',\omega')\Vdash\varphi$.*

Thus, modality $\Box_C\varphi$ states that coalition C had a know-how strategy to achieve φ. In Lemma 16 we will see that this modality satisfies Marc Pauly's Cooperation axiom, which is also true for know-how modality [22]. Nevertheless, as discussed in Section 3.7, this modality is different from the know-how modality. Because formula $\Box_C\varphi$ refers to a coalition

ability in the past, we call it *counterfactual know-how modality*. In the rest of this section we state and prove various properties of this modality. These properties are used later in the proof of the completeness. We start with an auxiliary property of the distributed knowledge modality that will be used to prove properties of the counterfactual know-how modality.

Lemma 11. *For any coalitions $C, D \subseteq \mathcal{A}$, any formulae $\varphi, \psi, \chi \in \Phi$, and any family of formulae $\{\theta_E \mid E \subseteq C \cup D\}$, if $\vdash \overline{\mathsf{K}}_C\varphi \wedge \overline{\mathsf{K}}_D\psi \to \chi \vee \bigvee_{E\subseteq C\cup D} \theta_E$, then $\vdash \overline{\mathsf{K}}_C\varphi \wedge \overline{\mathsf{K}}_D\psi \to \mathsf{K}_{C\cup D}\chi \vee \bigvee_{E\subseteq C\cup D} \overline{\mathsf{K}}_E\theta_E$.*

Proof. Suppose that $\vdash \overline{\mathsf{K}}_C\varphi \wedge \overline{\mathsf{K}}_D\psi \to \chi \vee \bigvee_{E\subseteq C\cup D} \theta_E$. Then, by propositional reasoning,

$$\vdash \left(\overline{\mathsf{K}}_C\varphi \wedge \overline{\mathsf{K}}_D\psi \wedge \bigwedge_{E\subseteq C\cup D} \neg\theta_E\right) \to \chi.$$

Thus, again by propositional reasoning,

$$\overline{\mathsf{K}}_C\varphi, \overline{\mathsf{K}}_D\psi, \{\neg\theta_E \mid E \subseteq C \cup D\} \vdash \chi.$$

Hence, by Lemma 4,

$$\mathsf{K}_{C\cup D}\overline{\mathsf{K}}_C\varphi, \mathsf{K}_{C\cup D}\overline{\mathsf{K}}_D\psi, \{\mathsf{K}_{C\cup D}\neg\theta_E \mid E \subseteq C \cup D\} \vdash \mathsf{K}_{C\cup D}\chi.$$

Then, by the Monotonicity axiom and the Modus Ponens inference rule,

$$\mathsf{K}_C\overline{\mathsf{K}}_C\varphi, \mathsf{K}_D\overline{\mathsf{K}}_D\psi, \{\mathsf{K}_E\neg\theta_E \mid E \subseteq C \cup D\} \vdash \mathsf{K}_{C\cup D}\chi.$$

Thus, by the Negative Introspection axiom, the Modus Ponens inference rule, and the definition of modality $\overline{\mathsf{K}}$,

$$\overline{\mathsf{K}}_C\varphi, \overline{\mathsf{K}}_D\psi, \{\mathsf{K}_E\neg\theta_E \mid E \subseteq C \cup D\} \vdash \mathsf{K}_{C\cup D}\chi.$$

Hence, by Lemma 3 and propositional reasoning,

$$\vdash \left(\overline{\mathsf{K}}_C\varphi \wedge \overline{\mathsf{K}}_D\psi \wedge \bigwedge_{E\subseteq C\cup D} \mathsf{K}_E\neg\theta_E\right) \to \mathsf{K}_{C\cup D}\chi.$$

Then, by the propositional reasoning,

$$\vdash \overline{\mathsf{K}}_C\varphi \wedge \overline{\mathsf{K}}_D\psi \to \mathsf{K}_{C\cup D}\chi \vee \bigvee_{E\subseteq C\cup D} \neg\mathsf{K}_E\neg\theta_E.$$

Finally,

$$\vdash \overline{\mathsf{K}}_C\varphi \wedge \overline{\mathsf{K}}_D\psi \to \mathsf{K}_{C\cup D}\chi \vee \bigvee_{E\subseteq C\cup D} \overline{\mathsf{K}}_E\theta_E$$

by the definition of modality $\overline{\mathsf{K}}$. ⊠

Lemma 12. $\vdash \mathsf{K}_C\varphi \wedge \mathsf{K}_D\psi \to \Box_{C\cup D}(\varphi \wedge \psi)$.

Proof. Tautology $\varphi \to (\psi \to (\varphi \wedge \psi))$, by the Necessitation inference rule, implies that $\vdash \mathsf{K}_{C\cup D}(\varphi \to (\psi \to (\varphi \wedge \psi)))$. Thus, by the Distributivity axiom and the Modus Ponens inference rule,

$$\vdash \mathsf{K}_{C\cup D}\varphi \to \mathsf{K}_{C\cup D}(\psi \to (\varphi \wedge \psi)).$$

Hence, by the Distributivity axiom and propositional reasoning,

$$\vdash \mathsf{K}_{C\cup D}\varphi \to (\mathsf{K}_{C\cup D}\psi \to \mathsf{K}_{C\cup D}(\varphi \wedge \psi)).$$

Then, by the laws of propositional reasoning,

$$\vdash \mathsf{K}_{C\cup D}\varphi \wedge \mathsf{K}_{C\cup D}\psi \to \mathsf{K}_{C\cup D}(\varphi \wedge \psi).$$

Thus, by the Monotonicity axiom and propositional reasoning,

$$\vdash \mathsf{K}_C\varphi \wedge \mathsf{K}_D\psi \to \mathsf{K}_{C\cup D}(\varphi \wedge \psi).$$

Hence, by propositional reasoning,

$$\vdash \mathsf{K}_C\varphi \wedge \mathsf{K}_D\psi \to \mathsf{K}_{C\cup D}(\varphi\wedge\psi) \vee \bigvee_{E\subseteq C\cup D} \overline{\mathsf{K}}_E\mathsf{B}_E\neg(\varphi\wedge\psi).$$

Therefore,

$$\vdash \mathsf{K}_C\varphi \wedge \mathsf{K}_D\psi \to \Box_{C\cup D}(\varphi\wedge\psi)$$

by the definition of modality $\Box$. ⊠

Lemma 13. *$\vdash \mathsf{K}_C\varphi \wedge \overline{\mathsf{K}}_F\mathsf{B}_F\neg\psi \to \Box_{C\cup D}(\varphi\wedge\psi)$, where $F\subseteq D$.*

Proof. Statement $\vdash \mathsf{K}_C(\varphi \to (\neg(\varphi\wedge\psi)\to\neg\psi))$ follows from tautology $\varphi \to (\neg(\varphi\wedge\psi)\to\neg\psi)$ by the Necessitation inference rule. Thus, by the Distributivity axiom and the Modus Ponens inference rule,

$$\vdash \mathsf{K}_C\varphi \to \mathsf{K}_C(\neg(\varphi\wedge\psi)\to\neg\psi).$$

Note that the following is an instance of the Knowledge and Blameworthiness axiom:

$$\vdash \mathsf{K}_C(\neg(\varphi\wedge\psi)\to\neg\psi) \wedge \overline{\mathsf{K}}_F\mathsf{B}_F\neg\psi \to \left(\neg(\varphi\wedge\psi) \to \bigvee_{E\subseteq C\cup F} \mathsf{B}_E\neg(\varphi\wedge\psi)\right).$$

Hence, by propositional reasoning from the last two formulae,

$$\vdash \mathsf{K}_C\varphi \wedge \overline{\mathsf{K}}_F\mathsf{B}_F\neg\psi \to \left(\neg(\varphi\wedge\psi) \to \bigvee_{E\subseteq C\cup F} \mathsf{B}_E\neg(\varphi\wedge\psi)\right).$$

Thus, by propositional reasoning,

$$\vdash \mathsf{K}_C\varphi \wedge \overline{\mathsf{K}}_F\mathsf{B}_F\neg\psi \to (\varphi\wedge\psi) \vee \bigvee_{E\subseteq C\cup F} \mathsf{B}_E\neg(\varphi\wedge\psi).$$

Note that $\vdash \overline{\mathsf{K}}_C\mathsf{K}_C\varphi \to \mathsf{K}_C\varphi$ by the contraposition of the Negative Introspection axiom and the definition of modality $\overline{\mathsf{K}}$. Then, by propositional reasoning,

$$\vdash \overline{\mathsf{K}}_C\mathsf{K}_C\varphi \wedge \overline{\mathsf{K}}_F\mathsf{B}_F\neg\psi \to (\varphi\wedge\psi) \vee \bigvee_{E\subseteq C\cup F} \mathsf{B}_E\neg(\varphi\wedge\psi).$$

Thus, by Lemma 11,

$$\vdash \overline{\mathsf{K}}_C\mathsf{K}_C\varphi \wedge \overline{\mathsf{K}}_F\mathsf{B}_F\neg\psi \to \mathsf{K}_{C\cup F}(\varphi\wedge\psi) \vee \bigvee_{E\subseteq C\cup F} \overline{\mathsf{K}}_E\mathsf{B}_E\neg(\varphi\wedge\psi).$$

Then, by the Monotonicity axiom and propositional reasoning using assumption $F\subseteq D$,

$$\vdash \overline{\mathsf{K}}_C\mathsf{K}_C\varphi \wedge \overline{\mathsf{K}}_F\mathsf{B}_F\neg\psi \to \mathsf{K}_{C\cup D}(\varphi\wedge\psi) \vee \bigvee_{E\subseteq C\cup F} \overline{\mathsf{K}}_E\mathsf{B}_E\neg(\varphi\wedge\psi).$$

Hence, just by propositional reasoning using assumption $F\subseteq D$,

$$\vdash \overline{\mathsf{K}}_C\mathsf{K}_C\varphi \wedge \overline{\mathsf{K}}_F\mathsf{B}_F\neg\psi \to \mathsf{K}_{C\cup D}(\varphi\wedge\psi) \vee \bigvee_{E\subseteq C\cup D} \overline{\mathsf{K}}_E\mathsf{B}_E\neg(\varphi\wedge\psi).$$

Note that $\vdash \mathsf{K}_C\varphi \to \overline{\mathsf{K}}_C\mathsf{K}_C\varphi$ by the contraposition of the Truth axiom and the definition of modality $\overline{\mathsf{K}}$. Thus, by propositional reasoning,

$$\vdash \mathsf{K}_C\varphi \wedge \overline{\mathsf{K}}_F\mathsf{B}_F\neg\psi \to \mathsf{K}_{C\cup D}(\varphi\wedge\psi) \vee \bigvee_{E\subseteq C\cup D} \overline{\mathsf{K}}_E\mathsf{B}_E\neg(\varphi\wedge\psi).$$

Therefore,

$$\vdash \mathsf{K}_C\varphi \wedge \overline{\mathsf{K}}_F\mathsf{B}_F\neg\psi \to \Box_{C\cup D}(\varphi\wedge\psi)$$

by the definition of modality $\Box$. ⊠

Lemma 14. *Inference rule* $\dfrac{\varphi \leftrightarrow \psi}{\Box_C \varphi \to \Box_C \psi}$ *is derivable in our system.*

Proof. Suppose that $\vdash \varphi \leftrightarrow \psi$. Thus, $\vdash \neg\varphi \leftrightarrow \neg\psi$ by the laws of propositional reasoning. Hence, $\vdash \mathsf{B}_D\neg\varphi \to \mathsf{B}_D\neg\psi$ for any coalition D by the Substitution inference rule. Then, $\vdash \neg\mathsf{B}_D\neg\psi \to \neg\mathsf{B}_D\neg\varphi$ by the contraposition. Thus, $\vdash \mathsf{K}_D(\neg\mathsf{B}_D\neg\psi \to \neg\mathsf{B}_D\neg\varphi)$ by the Necessitation inference rule. Hence, $\vdash \mathsf{K}_D\neg\mathsf{B}_D\neg\psi \to \mathsf{K}_D\neg\mathsf{B}_D\neg\varphi$ by the Distributivity axiom and the Modus Ponens inference rule. Then, $\vdash \neg\mathsf{K}_D\neg\mathsf{B}_D\neg\varphi \to \neg\mathsf{K}_D\neg\mathsf{B}_D\neg\psi$ by the contraposition. Thus, $\vdash \overline{\mathsf{K}}_D\mathsf{B}_D\neg\varphi \to \overline{\mathsf{K}}_D\mathsf{B}_D\neg\psi$ by the definition of modality $\overline{\mathsf{K}}$ for any coalition D. Hence, by propositional reasoning,

$$\vdash \bigvee_{D\subseteq C} \overline{\mathsf{K}}_D\mathsf{B}_D\neg\varphi \to \bigvee_{D\subseteq C} \overline{\mathsf{K}}_D\mathsf{B}_D\neg\psi.$$

At the same time, assumption $\vdash \varphi \leftrightarrow \psi$ implies that $\vdash \varphi \to \psi$. Hence, $\vdash \mathsf{K}_C(\varphi \to \psi)$ by the Necessitation inference rule. Then, $\vdash \mathsf{K}_C\varphi \to \mathsf{K}_C\psi$ by the Distributivity axiom and the Modus Ponens inference rule.

Thus, by propositional reasoning,

$$\vdash \mathsf{K}_C\varphi \vee \bigvee_{D\subseteq C} \overline{\mathsf{K}}_D\mathsf{B}_D\neg\varphi \to \mathsf{K}_C\psi \vee \bigvee_{D\subseteq C} \overline{\mathsf{K}}_D\mathsf{B}_D\neg\psi.$$

Therefore, $\vdash \Box_C\varphi \to \Box_C\psi$ by the definition of the modality $\Box$. ⊠

Lemma 15. $\vdash \mathsf{K}_C\psi \wedge \overline{\mathsf{K}}_F\mathsf{B}_F\neg\varphi \to \Box_{C\cup D}(\varphi \wedge \psi)$, *where* $F \subseteq D$.

Proof. Note that $\psi \wedge \varphi \leftrightarrow \varphi \wedge \psi$ is a propositional tautology. Thus, $\vdash \Box_C(\psi \wedge \varphi) \to \Box_C(\varphi \wedge \psi)$ by Lemma 14. Therefore,

$$\vdash \mathsf{K}_C\psi \wedge \overline{\mathsf{K}}_F\mathsf{B}_F\neg\varphi \to \Box_{C\cup D}(\varphi \wedge \psi)$$

by Lemma 13. ⊠

Lemma 16. $\vdash \Box_C\varphi \wedge \Box_D\psi \to \Box_{C\cup D}(\varphi \wedge \psi)$, *where* $C \cap D = \varnothing$.

Proof. Consider an arbitrary set $E \subseteq C$ and an arbitrary set $F \subseteq D$. By the Joint Blameworthiness axiom,

$$\vdash \overline{\mathsf{K}}_E\mathsf{B}_E\neg\varphi \wedge \overline{\mathsf{K}}_F\mathsf{B}_F\neg\psi \to \left((\neg\varphi \vee \neg\psi) \to \bigvee_{G\subseteq E\cup F} \mathsf{B}_G(\neg\varphi \vee \neg\psi)\right).$$

Note that $\neg\varphi \vee \neg\psi \leftrightarrow \neg(\varphi \wedge \psi)$ is a propositional tautology. Thus, by the Substitution inference rule, $\vdash \mathsf{B}_G(\neg\varphi \vee \neg\psi) \to \mathsf{B}_G\neg(\varphi \wedge \psi)$ for each coalition G. Then, by the laws of propositional reasoning,

$$\vdash \overline{\mathsf{K}}_E\mathsf{B}_E\neg\varphi \wedge \overline{\mathsf{K}}_F\mathsf{B}_F\neg\psi \to \left((\neg\varphi \vee \neg\psi) \to \bigvee_{G\subseteq E\cup F} \mathsf{B}_G\neg(\varphi \wedge \psi)\right).$$

Hence, again by the laws of propositional reasoning,

$$\vdash \overline{\mathsf{K}}_E\mathsf{B}_E\neg\varphi \wedge \overline{\mathsf{K}}_F\mathsf{B}_F\neg\psi \to (\varphi \wedge \psi) \vee \bigvee_{G\subseteq E\cup F} \mathsf{B}_G\neg(\varphi \wedge \psi).$$

Thus, by Lemma 11,

$$\vdash \overline{\mathsf{K}}_E\mathsf{B}_E\neg\varphi \wedge \overline{\mathsf{K}}_F\mathsf{B}_F\neg\psi \to \mathsf{K}_{E\cup F}(\varphi \wedge \psi) \vee \bigvee_{G\subseteq E\cup F} \overline{\mathsf{K}}_G\mathsf{B}_G\neg(\varphi \wedge \psi).$$

Note that $E \cup F \subseteq C \cup D$ due to the assumptions $E \subseteq C$ and $F \subseteq D$. Hence, by the Monotonicity axiom and the laws of propositional reasoning,

$$\vdash \overline{\mathsf{K}}_E\mathsf{B}_E\neg\varphi \wedge \overline{\mathsf{K}}_F\mathsf{B}_F\neg\psi \to \mathsf{K}_{C\cup D}(\varphi \wedge \psi) \vee \bigvee_{G\subseteq E\cup F} \overline{\mathsf{K}}_G\mathsf{B}_G\neg(\varphi \wedge \psi).$$

Then, because $E \cup F \subseteq C \cup D$, just by the laws of propositional reasoning,

$$\vdash \overline{\mathsf{K}}_E \mathsf{B}_E \neg\varphi \wedge \overline{\mathsf{K}}_F \mathsf{B}_F \neg\psi \to \mathsf{K}_{C\cup D}(\varphi\wedge\psi) \vee \bigvee_{G\subseteq C\cup D} \overline{\mathsf{K}}_G \mathsf{B}_G \neg(\varphi\wedge\psi).$$

Thus, by the definition of modality $\Box$,

$$\vdash \overline{\mathsf{K}}_E \mathsf{B}_E \neg\varphi \wedge \overline{\mathsf{K}}_F \mathsf{B}_F \neg\psi \to \Box_{C\cup D}(\varphi\wedge\psi),$$

for all sets E and F such that $E\subseteq C$ and $F\subseteq D$. Hence, by the laws of propositional reasoning,

$$\vdash \left(\bigvee_{E\subseteq C} \overline{\mathsf{K}}_E \mathsf{B}_E \neg\varphi\right) \wedge \left(\bigvee_{F\subseteq D} \overline{\mathsf{K}}_F \mathsf{B}_F \neg\psi\right) \to \Box_{C\cup D}(\varphi\wedge\psi).$$

Then, using Lemma 12, Lemma 13, and Lemma 15, by the laws of propositional reasoning,

$$\vdash \left(\mathsf{K}_C\varphi \vee \bigvee_{E\subseteq C} \overline{\mathsf{K}}_E \mathsf{B}_E \neg\varphi\right) \wedge \left(\mathsf{K}_D\psi \vee \bigvee_{F\subseteq D} \overline{\mathsf{K}}_F \mathsf{B}_F \neg\psi\right) \to \Box_{C\cup D}(\varphi\wedge\psi).$$

Therefore,

$$\vdash \Box_C\varphi \wedge \Box_D\psi \to \Box_{C\cup D}(\varphi\wedge\psi)$$

by the definition of modality $\Box$. ⊠

Lemma 17. *$\vdash \mathsf{K}_C\varphi \wedge \Box_D\psi \to \Box_C(\varphi\wedge\psi)$, where $D\subseteq C$.*

Proof. Consider any set $F\subseteq D$. Hence, $\vdash \mathsf{K}_C\varphi \wedge \overline{\mathsf{K}}_F \mathsf{B}_F\neg\psi \to \Box_C(\varphi\wedge\psi)$ by Lemma 13 and assumption $D\subseteq C$ of the lemma. Thus, by the laws of propositional reasoning,

$$\vdash \mathsf{K}_C\varphi \wedge \left(\bigvee_{F\subseteq D} \overline{\mathsf{K}}_F \mathsf{B}_F \neg\psi\right) \to \Box_C(\varphi\wedge\psi).$$

Then, by Lemma 12 and propositional reasoning,

$$\vdash \mathsf{K}_C\varphi \wedge \left(\mathsf{K}_D\psi \vee \bigvee_{F\subseteq D} \overline{\mathsf{K}}_F \mathsf{B}_F \neg\psi\right) \to \Box_C(\varphi\wedge\psi).$$

Therefore, $\vdash \mathsf{K}_C\varphi \wedge \Box_D\psi \to \Box_C(\varphi\wedge\psi)$ by the definition of modality $\Box$. ⊠

Lemma 18. *If $\vdash \varphi\to\psi$, then $\vdash \mathsf{K}_C\varphi \to \Box_C\psi$.*

Proof. Assumption $\vdash \varphi\to\psi$ implies $\vdash \mathsf{K}_C(\varphi\to\psi)$ by the Necessitation inference rule. Thus, $\vdash \mathsf{K}_C\varphi \to \mathsf{K}_C\psi$ by the Distributivity axiom and the Modus Ponens inference rule. Hence, by the laws of propositional reasoning,

$$\vdash \mathsf{K}_C\varphi \to \mathsf{K}_C\psi \vee \bigvee_{D\subseteq C} \overline{\mathsf{K}}_D \mathsf{B}_D \neg\psi.$$

Therefore, $\vdash \mathsf{K}_C\varphi \to \Box_C\psi$ by the definition of modality $\Box$. ⊠

Lemma 19. *Inference rule $\dfrac{\varphi\to\psi}{\Box_C\varphi\to\Box_C\psi}$ is derivable in our logical system.*

Proof. Consider an arbitrary set D such that $D\subseteq C$. Then, by the Knowledge and Blameworthiness axiom,

$$\vdash \mathsf{K}_C(\neg\psi\to\neg\varphi) \wedge \overline{\mathsf{K}}_D \mathsf{B}_D \neg\varphi \to \left(\neg\psi \to \bigvee_{E\subseteq C\cup D} \mathsf{B}_E\neg\psi\right).$$

At the same time, propositional tautology $\vdash (\varphi \to \psi) \to (\neg\psi \to \neg\varphi)$ implies $\vdash \mathsf{K}_C((\varphi \to \psi) \to (\neg\psi \to \neg\varphi))$ by the Necessitation inference rule. Then, $\vdash \mathsf{K}_C(\varphi \to \psi) \to \mathsf{K}_C(\neg\psi \to \neg\varphi)$ by the Distributivity axiom and the Modus Ponens inference rule. Thus, by the laws of propositional reasoning,

$$\vdash \mathsf{K}_C(\varphi \to \psi) \wedge \overline{\mathsf{K}}_D \mathsf{B}_D \neg\varphi \to \left(\neg\psi \to \bigvee_{E \subseteq C \cup D} \mathsf{B}_E \neg\psi\right).$$

Hence, again by propositional reasoning,

$$\vdash \mathsf{K}_C(\varphi \to \psi) \wedge \overline{\mathsf{K}}_D \mathsf{B}_D \neg\varphi \to \psi \vee \bigvee_{E \subseteq C \cup D} \mathsf{B}_E \neg\psi.$$

Then, by assumption $D \subseteq C$,

$$\vdash \mathsf{K}_C(\varphi \to \psi) \wedge \overline{\mathsf{K}}_D \mathsf{B}_D \neg\varphi \to \psi \vee \bigvee_{E \subseteq C} \mathsf{B}_E \neg\psi.$$

Note that $\vdash \overline{\mathsf{K}}_C \mathsf{K}_C(\varphi \to \psi) \to \mathsf{K}_C(\varphi \to \psi)$ by the contraposition of the Negative Introspection axiom and the definition of modality $\overline{\mathsf{K}}$. Thus,

$$\vdash \overline{\mathsf{K}}_C \mathsf{K}_C(\varphi \to \psi) \wedge \overline{\mathsf{K}}_D \mathsf{B}_D \neg\varphi \to \psi \vee \bigvee_{E \subseteq C} \mathsf{B}_E \neg\psi.$$

Hence, by Lemma 11 and the assumption $D \subseteq C$,

$$\vdash \overline{\mathsf{K}}_C \mathsf{K}_C(\varphi \to \psi) \wedge \overline{\mathsf{K}}_D \mathsf{B}_D \neg\varphi \to \mathsf{K}_C\psi \vee \bigvee_{E \subseteq C} \overline{\mathsf{K}}_E \mathsf{B}_E \neg\psi.$$

Observe that $\mathsf{K}_C \neg \mathsf{K}_C(\varphi \to \psi) \to \neg \mathsf{K}_C(\varphi \to \psi)$ is an instance of the Truth axiom. Thus, $\vdash \mathsf{K}_C(\varphi \to \psi) \to \neg \mathsf{K}_C \neg \mathsf{K}_C(\varphi \to \psi)$ by the law of contraposition. Also the assumption $\vdash \varphi \to \psi$ of the lemma implies $\vdash \mathsf{K}_C(\varphi \to \psi)$ by the Necessitation inference rule. Then, $\vdash \neg \mathsf{K}_C \neg \mathsf{K}_C(\varphi \to \psi)$ by the Modus Ponens inference rule. Hence, $\vdash \overline{\mathsf{K}}_C \mathsf{K}_C(\varphi \to \psi)$ by the definition of modality $\overline{\mathsf{K}}$. Thus,

$$\vdash \overline{\mathsf{K}}_D \mathsf{B}_D \neg\varphi \to \mathsf{K}_C\psi \vee \bigvee_{E \subseteq C} \overline{\mathsf{K}}_E \mathsf{B}_E \neg\psi.$$

Then, by the definition of modality $\Box$, for each subset $D \subseteq C$,

$$\vdash \overline{\mathsf{K}}_D \mathsf{B}_D \neg\varphi \to \Box_C\psi.$$

Hence, by the laws of propositional reasoning,

$$\vdash \bigvee_{D \subseteq C} \overline{\mathsf{K}}_D \mathsf{B}_D \neg\varphi \to \Box_C\psi.$$

Thus, by Lemma 18, assumption $\vdash \varphi \to \psi$, and propositional reasoning,

$$\vdash \mathsf{K}_C\varphi \vee \bigvee_{D \subseteq C} \overline{\mathsf{K}}_D \mathsf{B}_D \neg\varphi \to \Box_C\psi.$$

Therefore, $\vdash \Box_C\varphi \to \Box_C\psi$ by the definition of modality $\Box$. ⊠

Lemma 20. *For any disjoint $D_1, \dots, D_n \subseteq C$, if $\psi_1, \dots, \psi_m, \chi_1, \dots, \chi_n \vdash \varphi$, then $\mathsf{K}_C\psi_1, \dots, \mathsf{K}_C\psi_m, \Box_{D_1}\chi_1, \dots, \Box_{D_n}\chi_n \vdash \Box_C\varphi$.*

Proof. Note that $\top$ is a propositional tautology. Thus, $\vdash \mathsf{K}_\varnothing \top$ by the Necessitation inference rule. Hence, by propositional reasoning,

$$\vdash \mathsf{K}_\varnothing \top \vee \bigvee_{D \subseteq \varnothing} \overline{\mathsf{K}}_D \mathsf{B}_D \neg\top.$$

Thus, $\vdash \Box_\varnothing \top$ by the definition of modality $\Box$. Hence,

$$\Box_{D_1}\chi_1, \dots, \Box_{D_n}\chi_n \vdash \Box_{D_1 \cup \dots \cup D_n}(\chi_1 \wedge \dots \wedge \chi_n \wedge \top)$$

by Lemma 16 applied n times using the assumption that sets $D_1, \dots, D_n$ are disjoint. Then,

$$\mathsf{K}_C\psi_1, \dots, \mathsf{K}_C\psi_m, \Box_{D_1}\chi_1, \dots, \Box_{D_n}\chi_n \vdash \Box_C(\psi_1 \wedge \dots \wedge \psi_m \wedge \chi_1 \wedge \dots \wedge \chi_n \wedge \top)$$

by Lemma 17 applied m times and the assumption $D_1, \dots, D_n \subseteq C$.

Note now that assumption $\psi_1, \dots, \psi_m, \chi_1, \dots, \chi_n \vdash \varphi$ of the lemma implies $\vdash \psi_1 \wedge \dots \wedge \psi_m \wedge \chi_1 \wedge \dots \wedge \chi_n \wedge \top \to \varphi$ by Lemma 3 and the laws of propositional reasoning. Thus, $\vdash \Box_C(\psi_1 \wedge \dots \wedge \psi_m \wedge \chi_1 \wedge \dots \wedge \chi_n \wedge \top) \to \Box_C\varphi$ by Lemma 19. Therefore, $\mathsf{K}_C\psi_1, \dots, \mathsf{K}_C\psi_m, \Box_{D_1}\chi_1, \dots, \Box_{D_n}\chi_n \vdash \Box_C\varphi$ by the Modus Ponens inference rule. ⊠

Lemma 21. $\vdash \neg\mathsf{B}_C\varphi \to \neg\varphi \vee \neg\Box_C\neg\varphi \vee \bigvee_{D\subsetneq C} \Box_D\neg\varphi$.

Proof. $\mathsf{B}_D\varphi \to \mathsf{K}_D\neg\mathsf{B}_C\varphi$ is an instance of the Minimality axiom for each set $D \subsetneq C$. Hence, $\mathsf{B}_D\varphi \to \mathsf{K}_C\neg\mathsf{B}_C\varphi$ by the Monotonicity axiom and propositional reasoning. Thus, $\vdash \neg\mathsf{K}_C\neg\mathsf{B}_C\varphi \to \neg\mathsf{B}_D\varphi$ by the law of contraposition. Hence, $\vdash \overline{\mathsf{K}}_C\mathsf{B}_C\varphi \to \neg\mathsf{B}_D\varphi$ by the definition of modality $\overline{\mathsf{K}}$ for each set $D \subsetneq C$. Then, by the laws of propositional reasoning,

$$\vdash \overline{\mathsf{K}}_C\mathsf{B}_C\varphi \to \bigwedge_{D\subsetneq C} \neg\mathsf{B}_D\varphi.$$

At the same time, by the Knowledge and Blameworthiness axiom,

$$\vdash \mathsf{K}_C(\varphi \to \varphi) \wedge \overline{\mathsf{K}}_C\mathsf{B}_C\varphi \to \left(\varphi \to \bigvee_{D\subseteq C} \mathsf{B}_D\varphi\right).$$

Thus, by propositional reasoning from the previous two formulae,

$$\vdash \mathsf{K}_C(\varphi \to \varphi) \wedge \overline{\mathsf{K}}_C\mathsf{B}_C\varphi \to (\varphi \to \mathsf{B}_C\varphi).$$

Note that $\vdash \mathsf{K}_C(\varphi \to \varphi)$ by the Necessitation inference rule because $\varphi \to \varphi$ is a tautology. Thus, by propositional reasoning,

$$\vdash \overline{\mathsf{K}}_C\mathsf{B}_C\varphi \to (\varphi \to \mathsf{B}_C\varphi). \tag{4}$$

At the same time, $\neg\neg\varphi \leftrightarrow \varphi$ is a propositional tautology. Then, by the Substitution inference rule, $\vdash \mathsf{B}_C\neg\neg\varphi \to \mathsf{B}_C\varphi$. Thus, $\vdash \neg\mathsf{B}_C\varphi \to \neg\mathsf{B}_C\neg\neg\varphi$ by the contraposition. Hence, $\vdash \mathsf{K}_C(\neg\mathsf{B}_C\varphi \to \neg\mathsf{B}_C\neg\neg\varphi)$ by the Necessitation inference rule. Then, $\vdash \mathsf{K}_C\neg\mathsf{B}_C\varphi \to \mathsf{K}_C\neg\mathsf{B}_C\neg\neg\varphi$ by the Distributivity axiom and the Modus Ponens inference rule. Thus, $\vdash \neg\mathsf{K}_C\neg\mathsf{B}_C\neg\neg\varphi \to \neg\mathsf{K}_C\neg\mathsf{B}_C\varphi$ by the contraposition. Hence, $\vdash \overline{\mathsf{K}}_C\mathsf{B}_C\neg\neg\varphi \to \overline{\mathsf{K}}_C\mathsf{B}_C\varphi$ by the definition of modality $\overline{\mathsf{K}}$. Then, by the laws of proposition reasoning using statement (4),

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

By propositional reasoning,

$$\vdash \varphi \wedge \overline{\mathsf{K}}_C\mathsf{B}_C\neg\neg\varphi \to \mathsf{B}_C\varphi. \tag{5}$$

Note that the following formula is a propositional tautology:

$$\left(\bigvee_{D\subseteq C} \overline{\mathsf{K}}_D\mathsf{B}_D\neg\neg\varphi \wedge \bigwedge_{F\subsetneq C} \neg\overline{\mathsf{K}}_F\mathsf{B}_F\neg\neg\varphi\right) \to \overline{\mathsf{K}}_C\mathsf{B}_C\neg\neg\varphi.$$

Thus, by propositional reasoning, using statement (5),

$$\vdash \varphi \wedge \bigvee_{D\subseteq C} \overline{\mathsf{K}}_D\mathsf{B}_D\neg\neg\varphi \wedge \bigwedge_{F\subsetneq C} \neg\overline{\mathsf{K}}_F\mathsf{B}_F\neg\neg\varphi \to \mathsf{B}_C\varphi.$$

By propositional reasoning, we can introduce an extra assumption,

$$\vdash \varphi \wedge \bigvee_{D\subseteq C} \overline{\mathsf{K}}_D\mathsf{B}_D\neg\neg\varphi \wedge \bigwedge_{D\subsetneq C} \neg\mathsf{K}_D\neg\varphi \wedge \bigwedge_{F\subsetneq C} \neg\overline{\mathsf{K}}_F\mathsf{B}_F\neg\neg\varphi \to \mathsf{B}_C\varphi.$$

Again by the laws of propositional reasoning,

$$\vdash \varphi \wedge \bigvee_{D\subseteq C} \overline{\mathsf{K}}_D\mathsf{B}_D\neg\neg\varphi \wedge \bigwedge_{D\subsetneq C} \left(\neg\mathsf{K}_D\neg\varphi \wedge \bigwedge_{F\subseteq D} \neg\overline{\mathsf{K}}_F\mathsf{B}_F\neg\neg\varphi\right) \to \mathsf{B}_C\varphi.$$

Using De Morgan's laws and propositional reasoning,

$$\vdash \varphi \wedge \bigvee_{D\subseteq C} \overline{\mathsf{K}}_D \mathsf{B}_D \neg\neg\varphi \wedge \bigwedge_{D\subsetneq C} \neg\left(\mathsf{K}_D\neg\varphi \vee \bigvee_{F\subseteq D} \overline{\mathsf{K}}_F \mathsf{B}_F \neg\neg\varphi\right) \to \mathsf{B}_C\varphi.$$

Then, using the definition of modality $\Box$,

$$\vdash \varphi \wedge \bigvee_{D\subseteq C} \overline{\mathsf{K}}_D \mathsf{B}_D \neg\neg\varphi \wedge \bigwedge_{D\subsetneq C} \neg\Box_D\neg\varphi \to \mathsf{B}_C\varphi.$$

Note that $\mathsf{K}_C\neg\varphi \to \neg\varphi$ is an instance of the Truth axiom. Thus, by the law of contrapositive, $\vdash \varphi \to \neg\mathsf{K}_C\neg\varphi$. Hence, by propositional reasoning,

$$\vdash \varphi \wedge \left(\mathsf{K}_C\neg\varphi \vee \bigvee_{D\subseteq C} \overline{\mathsf{K}}_D \mathsf{B}_D \neg\neg\varphi\right) \wedge \bigwedge_{D\subsetneq C} \neg\Box_D\neg\varphi \to \mathsf{B}_C\varphi.$$

Then, by the definition of modality $\Box$,

$$\vdash \varphi \wedge \Box_C\neg\varphi \wedge \bigwedge_{D\subsetneq C} \neg\Box_D\neg\varphi \to \mathsf{B}_C\varphi.$$

Therefore,

$$\vdash \neg\mathsf{B}_C\varphi \to \neg\varphi \vee \neg\Box_C\neg\varphi \vee \bigvee_{D\subsetneq C} \Box_D\neg\varphi,$$

by the laws of propositional reasoning. ⊠

Lemma 22. $\vdash \mathsf{B}_C\varphi \to \Box_C\neg\varphi$.

Proof. Tautology $\varphi \leftrightarrow \neg\neg\varphi$, by the Substitution inference rule, implies that $\vdash \mathsf{B}_C\varphi \to \mathsf{B}_C\neg\neg\varphi$. Hence, $\vdash \neg\mathsf{B}_C\neg\neg\varphi \to \neg\mathsf{B}_C\varphi$ by the law of contraposition. Thus, $\vdash \mathsf{K}_C\neg\mathsf{B}_C\neg\neg\varphi \to \neg\mathsf{B}_C\varphi$ by the Truth axiom and propositional reasoning. Then, $\vdash \mathsf{B}_C\varphi \to \neg\mathsf{K}_C\neg\mathsf{B}_C\neg\neg\varphi$ by the contraposition. Hence, $\vdash \mathsf{B}_C\varphi \to \overline{\mathsf{K}}_C\mathsf{B}_C\neg\neg\varphi$ by the definition of modality $\overline{\mathsf{K}}$. Then, by propositional reasoning, we can add a disjunct to the conclusion,

$$\vdash \mathsf{B}_C\varphi \to \mathsf{K}_C\neg\varphi \vee \overline{\mathsf{K}}_C\mathsf{B}_C\neg\neg\varphi.$$

And even more disjuncts,

$$\vdash \mathsf{B}_C\varphi \to \mathsf{K}_C\neg\varphi \vee \bigvee_{D\subseteq C} \overline{\mathsf{K}}_D\mathsf{B}_D\neg\neg\varphi.$$

Therefore, $\vdash \mathsf{B}_C\varphi \to \Box_C\neg\varphi$, by the definition of modality $\Box$. ⊠

Lemma 23. $\vdash \mathsf{B}_C\varphi \to \neg\Box_D\neg\varphi$ *for each* $D \subsetneq C$.

Proof. By the Substitution inference rule, tautology $\varphi \leftrightarrow \neg\neg\varphi$ implies that $\vdash \mathsf{B}_C\varphi \to \mathsf{B}_C\neg\neg\varphi$. Consider any set $E \subseteq D$. Then, $E \subseteq D \subsetneq C$. Thus, $\vdash \mathsf{B}_C\varphi \to \mathsf{K}_E\neg\mathsf{B}_E\neg\neg\varphi$ by Lemma 1. Then, $\vdash \mathsf{B}_C\varphi \to \neg\neg\mathsf{K}_E\neg\mathsf{B}_E\neg\neg\varphi$ by propositional reasoning. Hence, $\vdash \mathsf{B}_C\varphi \to \neg\overline{\mathsf{K}}_E\mathsf{B}_E\neg\neg\varphi$ by the definition of modality $\overline{\mathsf{K}}$ for each set $E \subseteq D$. Thus, by propositional reasoning,

$$\vdash \mathsf{B}_C\varphi \to \bigwedge_{E\subseteq D} \neg\overline{\mathsf{K}}_E\mathsf{B}_E\neg\neg\varphi. \tag{6}$$

At the same time, $\mathsf{K}_D\neg\varphi \to \neg\varphi$ is an instance of the Truth axiom. Then, $\vdash \varphi \to \neg\mathsf{K}_D\neg\varphi$ by the contraposition. Hence, $\vdash \mathsf{B}_C\varphi \to \neg\mathsf{K}_D\neg\varphi$ by propositional reasoning using the instance $\mathsf{B}_C\varphi \to \varphi$ of the Truth axiom. Thus, by propositional reasoning using statement (6),

$$\vdash \mathsf{B}_C\varphi \to \neg\mathsf{K}_D\neg\varphi \wedge \bigwedge_{E\subseteq D} \neg\overline{\mathsf{K}}_E\mathsf{B}_E\neg\neg\varphi.$$

Then, by De Morgan's laws and propositional reasoning,

$$\vdash \mathsf{B}_C\varphi \to \neg\left(\mathsf{K}_D\neg\varphi \vee \bigvee_{E\subseteq D} \overline{\mathsf{K}}_E\mathsf{B}_E\neg\neg\varphi\right).$$

Therefore, $\vdash \mathsf{B}_C\varphi \to \neg\Box_D\neg\varphi$ by the definition of modality $\Box$. ⊠

Lemma 24. $\vdash \mathsf{K}_C\varphi \vee \mathsf{K}_C\psi \to \mathsf{K}_C(\varphi \vee \psi)$.

Proof. Note that $\varphi \to \varphi \vee \psi$ is a proposition tautology. Thus, by the Necessitation inference rule, $\vdash \mathsf{K}_C(\varphi \to \varphi \vee \psi)$. Hence, $\vdash \mathsf{K}_C\varphi \to \mathsf{K}_C(\varphi \vee \psi)$ by the Distributivity axiom and the Modus Ponens inference rule. Similarly, $\vdash \mathsf{K}_C\psi \to \mathsf{K}_C(\varphi \vee \psi)$. Therefore, $\vdash \mathsf{K}_C\varphi \vee \mathsf{K}_C\psi \to \mathsf{K}_C(\varphi \vee \psi)$ by propositional reasoning. ⊠

Lemma 25. $\vdash \Box_C\varphi \to \mathsf{K}_C\Box_C\varphi$.

Proof. Formula $\Box_C\varphi \to \Box_C\varphi$ is propositional tautology. Thus, by the definition of modality $\Box$,

$$\vdash \Box_C\varphi \to \left(\mathsf{K}_C\varphi \vee \bigvee_{D\subseteq C} \overline{\mathsf{K}}_D\mathsf{B}_D\neg\varphi\right).$$

Hence, by Lemma 2 and propositional reasoning,

$$\vdash \Box_C\varphi \to \left(\mathsf{K}_C\mathsf{K}_C\varphi \vee \bigvee_{D\subseteq C} \overline{\mathsf{K}}_D\mathsf{B}_D\neg\varphi\right).$$

Then, by the Negative Introspection axiom, the definition of modality $\overline{\mathsf{K}}$ and the laws of propositional reasoning,

$$\vdash \Box_C\varphi \to \left(\mathsf{K}_C\mathsf{K}_C\varphi \vee \bigvee_{D\subseteq C} \mathsf{K}_D\overline{\mathsf{K}}_D\mathsf{B}_D\neg\varphi\right).$$

Thus, by the Monotonicity axiom and propositional reasoning,

$$\vdash \Box_C\varphi \to \left(\mathsf{K}_C\mathsf{K}_C\varphi \vee \bigvee_{D\subseteq C} \mathsf{K}_C\overline{\mathsf{K}}_D\mathsf{B}_D\neg\varphi\right).$$

Hence, by propositional reasoning using Lemma 24,

$$\vdash \Box_C\varphi \to \mathsf{K}_C\left(\mathsf{K}_C\varphi \vee \bigvee_{D\subseteq C} \overline{\mathsf{K}}_D\mathsf{B}_D\neg\varphi\right).$$

Therefore, $\vdash \Box_C\varphi \to \mathsf{K}_C\Box_C\varphi$ by the definition of modality $\Box$. ⊠

Lemma 26. $\vdash \Box_\varnothing\varphi \to \varphi$.

Proof. Formula $\neg\mathsf{B}_\varnothing\neg\varphi$ is an instance of the None to Blame axiom. Thus, $\vdash \mathsf{K}_\varnothing\neg\mathsf{B}_\varnothing\neg\varphi$ by the Necessitation inference rule. Hence, $\vdash \neg\overline{\mathsf{K}}_\varnothing\mathsf{B}_\varnothing\neg\varphi$ by propositional reasoning using the definition of modality $\overline{\mathsf{K}}$. At the same time, $\mathsf{K}_\varnothing\varphi \to \varphi$ is an instance of the Truth axiom. Then, by propositional reasoning,

$$\vdash \mathsf{K}_\varnothing\varphi \vee \overline{\mathsf{K}}_\varnothing\mathsf{B}_\varnothing\neg\varphi \to \varphi.$$

Note that empty set $\varnothing$ has only one subset. Thus,

$$\vdash \mathsf{K}_\varnothing\varphi \vee \bigvee_{D\subseteq\varnothing} \overline{\mathsf{K}}_D\mathsf{B}_D\neg\varphi \to \varphi.$$

Therefore, $\vdash \Box_\varnothing\varphi \to \varphi$ by the definition of modality $\Box$. ⊠

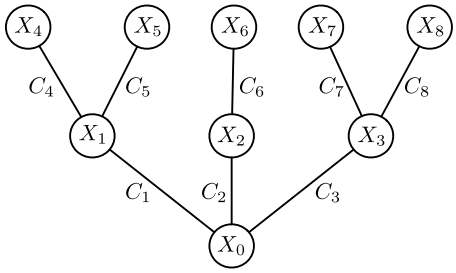


Fig. 4. A fragment of tree.

7. Completeness

In this section we prove the completeness of our logical system. The completeness theorem is stated in the end of this section as Theorem 1.

The standard completeness proof for epistemic logic of individual knowledge defines states as maximal consistent sets. Similarly, we defined outcomes of the game as maximal consistent sets in [14]. In the case of the epistemic logic of individual knowledge, two states are usually defined to be indistinguishable by an agent a if these two states have the same K_a formulae. Unfortunately, this approach does not work for distributed knowledge. Indeed, two maximal consistent sets that have the same K_a and K_b formulae might have different $\mathsf{K}_{a,b}$ formulae. Such two states would be indistinguishable to agent a and agent b, however, the distributed knowledge of agents a and b in these states will be different. This situation is inconsistent with Definition 3. To solve this problem we define outcomes not as maximal consistent sets of formulae, but as nodes of a tree. This approach has been previously used to prove the completeness of several logics for know-how modality [19,21–24].

We start the proof of the completeness by defining the canonical game $G(X_0) = (I, \{\sim_a\}_{a\in\mathcal{A}}, \Delta, \Omega, P, \pi)$ for each maximal consistent set of formulae X_0. In this definition, Φ refers to the set of all formulae in our language, see Definition 1.

Definition 4. The set of outcomes Ω consists of all finite sequences $X_0, C_1, X_1, C_2, \dots, C_n, X_n$, such that

1. $n \geq 0$,
2. X_i is a maximal consistent subset of Φ for each $i \geq 1$,
3. C_i is a coalition for each $i \geq 1$,
4. $\{\varphi \mid \mathsf{K}_{C_i}\varphi \in X_{i-1}\} \subseteq X_i$ for each $i \geq 1$.

For any sequence $s = x_1, \dots, x_n$ and any element y, by $s :: y$ we mean the sequence $x_1, \dots, x_n, y$. By $hd(s)$ we mean element x_n. We define a tree structure on the set of outcomes Ω by saying that outcome (node) $\omega = X_0, C_1, X_1, C_2, \dots, C_n, X_n$ and outcome (node) $\omega :: C_{n+1} :: X_{n+1}$ are connected by an undirected edge labeled with all agents in coalition C_{n+1}, see Fig. 4.

Definition 5. For any outcomes $\omega, \omega' \in \Omega$ and any agent $a \in \mathcal{A}$, let $\omega \sim_a \omega'$ if all edges along the unique path between ω and ω' are labeled with agent a.

Lemma 27. *Relation $\sim_a$ is an equivalence relation on set Ω.* ⊠

Lemma 29 below shows that the tree construction overcomes the distributed knowledge challenge discussed in the preamble for this section. Lemma 28 lays ground for the induction step in the proof of Lemma 29.

Lemma 28. $\mathsf{K}_D\varphi \in X_n$ *iff* $\mathsf{K}_D\varphi \in X_{n+1}$ *for any formula* $\varphi \in \Phi$, *any* $n \geq 0$, *and any outcome* $X_0, C_1, X_1, C_2, \dots, X_n, C_{n+1}, X_{n+1} \in \Omega$, *and any coalition* $D \subseteq C_{n+1}$.

Proof. If $\mathsf{K}_D\varphi \in X_n$, then $X_n \vdash \mathsf{K}_D\mathsf{K}_D\varphi$ by Lemma 2. Hence, $X_n \vdash \mathsf{K}_{C_{n+1}}\mathsf{K}_D\varphi$ by the Monotonicity axiom, the assumption $D \subseteq C_{n+1}$, and the Modus Ponens inference rule. Thus, $\mathsf{K}_{C_{n+1}}\mathsf{K}_D\varphi \in X_n$ by the maximality of set X_n. Therefore, $\mathsf{K}_D\varphi \in X_{n+1}$ by Definition 4.

Suppose that $\mathsf{K}_D\varphi \notin X_n$. Hence, $\neg\mathsf{K}_D\varphi \in X_n$ by the maximality of set X_n. Thus, $X_n \vdash \mathsf{K}_D\neg\mathsf{K}_D\varphi$ by the Negative Introspection axiom and the Modus Ponens inference rule. Hence, $X_n \vdash \mathsf{K}_{C_{n+1}}\neg\mathsf{K}_D\varphi$ by the Monotonicity axiom, the assumption $D \subseteq C_{n+1}$, and the Modus Ponens inference rule. Then, $\mathsf{K}_{C_{n+1}}\neg\mathsf{K}_D\varphi \in X_n$ by the maximality of set X_n. Thus, $\neg\mathsf{K}_D\varphi \in X_{n+1}$ by Definition 4. Therefore, $\mathsf{K}_D\varphi \notin X_{n+1}$ because set X_{n+1} is consistent. ⊠

Lemma 29. *If $\omega \sim_C \omega'$, then $\mathsf{K}_C\varphi \in hd(\omega)$ iff $\mathsf{K}_C\varphi \in hd(\omega')$.*

Proof. If $\omega \sim_C \omega'$, then each edge along the unique path between nodes ω and ω' is labeled with all agents in coalition C.

We prove the lemma by induction on the length of the unique path between nodes ω and ω'. In the base case, $\omega = \omega'$. Thus, $\mathsf{K}_C\varphi \in hd(\omega)$ iff $\mathsf{K}_C\varphi \in hd(\omega')$. The induction step follows from Lemma 28. ⊠

Lemma 30. *If $\omega \sim_C \omega'$ and $\mathsf{K}_C\varphi \in hd(\omega)$, then $\varphi \in hd(\omega')$.*

Proof. By Lemma 29, assumptions $\omega \sim_C \omega'$ and $\mathsf{K}_C\varphi \in hd(\omega)$ imply that $\mathsf{K}_C\varphi \in hd(\omega')$. Thus, $hd(\omega') \vdash \varphi$ by the Truth axiom and the Modus Ponens inference rule. Therefore, $\varphi \in hd(\omega')$ because set $hd(\omega')$ is maximal. ⊠

The set of the initial states I of the canonical game is the set of all equivalence classes of Ω with respect to relation $\sim_{\mathcal{A}}$.

Definition 6. $I = \Omega/\sim_{\mathcal{A}}$.

Lemma 31. *Relation $\sim_C$ is well-defined on set I.*

Proof. Suppose that $\omega_1 \sim_C \omega_2$. Consider any outcomes ω_1' and ω_2' such that $\omega_1 \sim_{\mathcal{A}} \omega_1'$ and $\omega_2 \sim_{\mathcal{A}} \omega_2'$. It suffices to prove that $\omega_1' \sim_C \omega_2'$.

By Definition 5 and Lemma 27, assumption $\omega_1 \sim_{\mathcal{A}} \omega_1'$ implies that each edge along the unique path between nodes ω_1' and ω_1 is labeled with all agents in set $\mathcal{A}$. Also, assumption $\omega_1 \sim_C \omega_2$ implies that each edge along the unique path between nodes ω_1 and ω_2 is labeled with all agents in coalition C. Finally, assumption $\omega_2 \sim_{\mathcal{A}} \omega_2'$ implies that each edge along the unique path between nodes ω_2 and ω_2' is labeled with all agents in set $\mathcal{A}$. Hence, each edge along the unique path between nodes ω_1' and ω_2' is labeled with all agents in coalition C. Therefore, $\omega_1' \sim_C \omega_2'$ by Definition 5. ⊠

Lemma 32. *$\alpha \sim_C \alpha'$ iff $\omega \sim_C \omega'$, for any initial states $\alpha, \alpha' \in I$, any outcomes $\omega \in \alpha$ and $\omega' \in \alpha'$, and any coalition $C \subseteq \mathcal{A}$.* ⊠

Informally, in the canonical game each agent "votes" for a formula. In order for (α, δ, ω) to be a valid play, it must be true that if $\Box_C\varphi \in hd(\omega)$ and all members of coalition C vote for φ, then $\varphi \in hd(\omega)$.

Definition 7. The domain of actions Δ is set Φ.

Definition 8. The set $P \subseteq I \times \Delta^{\mathcal{A}} \times \Omega$ consists of all triples (α, δ, ω) such that $\omega \in \alpha$ and for any formula $\Box_C\varphi \in hd(\omega)$, if $\delta(a) = \varphi$ for each agent $a \in C$, then $\varphi \in hd(\omega)$.

Definition 9. $\pi(p) = \{(\alpha, \delta, \omega) \in P \mid p \in hd(\omega)\}$.

This concludes the definition of the canonical game $G(X_0)$. In Lemma 36 we will show the condition from item 5 of Definition 2 is satisfied. Namely, for each initial state $\alpha \in I$ and each complete action profile $\delta \in \Delta^{\mathcal{A}}$ there is at least one outcome $\omega \in \Omega$ such that $(\alpha, \delta, \omega) \in P$.

We state and prove the completeness later in this section as Theorem 1. As usual in the proofs of completeness, the key step of the proof is an "induction" or "truth" lemma. In our case it is Lemma 38. We start with auxiliary results that will be used in the proof of the truth lemma. The truth lemma is proven by induction on the structural complexity and, thus, the argument there is carried out in terms of the primitive modalities K and B. The auxiliary lemmas below, however, are stated for modalities K and $\Box$.

Lemma 33. *For any play $(\alpha, \delta, \omega) \in P$ of game $G(X_0)$ and any formula $\Box_C\varphi \in hd(\omega)$, there is an action profile $s \in \Delta^C$ such that for each play $(\alpha', \delta', \omega') \in P$, if $\alpha \sim_C \alpha'$ and $s =_C \delta'$, then $\varphi \in hd(\omega')$.*

Proof. Let $s \in \Delta^C$ be an action profile of coalition C such that $s(a) = \varphi$ for each agent $a \in C$. Consider any play $(\alpha', \delta', \omega') \in P$ such that $\alpha \sim_C \alpha'$ and $s =_C \delta'$. We will show that $\varphi \in hd(\omega')$.

Indeed, assumption $\Box_C\varphi \in hd(\omega)$ of the lemma implies that $hd(\omega) \vdash \mathsf{K}_C\Box_C\varphi$ by Lemma 25. Then, $\mathsf{K}_C\Box_C\varphi \in hd(\omega)$ because $hd(\omega)$ is a maximal consistent set. At the same time, assumption $\alpha \sim_C \alpha'$ implies that $\omega \sim_C \omega'$ by Lemma 32. Hence, $\Box_C\varphi \in hd(\omega')$ by Lemma 30. Therefore, $\varphi \in hd(\omega')$ by Definition 8 and because $\delta(a) = s(a) = \varphi$ for each $a \in C$. ⊠

Lemma 34. *For any play $(\alpha, \delta, \omega) \in P$ of game $G(X_0)$, any action profile $s \in \Delta^C$, and any formula $\neg\Box_C\varphi \in hd(\omega)$, there is a play $(\alpha', \delta', \omega') \in P$ such that $\alpha \sim_C \alpha'$, $s =_C \delta'$, and $\varphi \notin hd(\omega')$.*

Proof. Consider set of formulae

$$X = \{\neg\varphi\} \cup \{\psi \mid \mathsf{K}_C\psi \in hd(\omega)\} \cup \{\chi \mid \Box_D\chi \in hd(\omega), D \subseteq C, \forall a \in D(s(a) = \chi)\}.$$

Claim 1. *Set X is consistent.*

Proof of Claim. Suppose the opposite. Thus, there are formulae

$$\mathsf{K}_C\psi_1, \dots, \mathsf{K}_C\psi_m \in hd(\omega) \tag{7}$$

and formulae

$$\Box_{D_1}\chi_1, \dots, \Box_{D_n}\chi_n \in hd(\omega) \tag{8}$$

such that

$$D_1, \dots, D_n \subseteq C, \tag{9}$$

$$s(a) = \chi_i, \text{ for each } i \le n \text{ and each } a \in D_i, \tag{10}$$

and

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

Without loss of generality, we can assume that formulae $\chi_1, \dots, \chi_n$ are distinct. Then, sets $D_1, \dots, D_n$ are disjoint due to assumption (10). Thus, by Lemma 20 using statement (9),

$$\mathsf{K}_C\psi_1, \dots, \mathsf{K}_C\psi_m, \Box_{D_1}\chi_1, \dots, \Box_{D_n}\chi_n \vdash \Box_C\varphi.$$

Hence, $hd(\omega) \vdash \Box_C\varphi$ because of statements (7) and (8). Therefore, $\neg\Box_C\varphi \notin hd(\omega)$ because set $hd(\omega)$ is consistent, which is a contradiction. ⊠

Let X' be any maximal consistent extension of set X and ω' be sequence $\omega :: C :: X'$. Note that $\omega' \in \Omega$ by Definition 4 and the choice of set X, set X', and sequence ω'. Define α' to be the equivalence class of ω' with respect to equivalence relation $\sim_{\mathcal{A}}$. Then, $\alpha' \in I$ by Definition 6. Finally, let the complete action profile δ' be defined as

$$\delta'(a) = \begin{cases} s(a), & \text{if } a \in C, \\ \top, & \text{otherwise.} \end{cases} \tag{11}$$

Claim 2. *$\omega \sim_C \omega'$ and $\alpha \sim_C \alpha'$.*

Proof of Claim. Since $\omega' = \omega :: C :: X$, the edge between nodes ω and ω' is labeled with each agent $a \in C$. Thus, $\omega \sim_a \omega'$ for each agent $a \in C$ by Definition 5. Then, $\omega \sim_C \omega'$. Therefore, $\alpha \sim_C \alpha'$ by Lemma 32. ⊠

Claim 3. *$(\alpha', \delta', \omega') \in P$.*

Proof of Claim. Note that $\omega' \in \alpha'$ because α' is an equivalence class of ω'. Consider any formula $\Box_D\psi \in hd(\omega')$ such that

$$\delta'(a) = \psi, \text{ for each agent } a \in D. \tag{12}$$

By Definition 8, it suffices to show that $\psi \in hd(\omega')$. We consider the following two cases separately:

Case I: $D \subseteq C$. Assumption $\Box_D\psi \in hd(\omega')$ implies that $hd(\omega') \vdash \mathsf{K}_D\Box_D\psi$ by Lemma 25. Hence, $hd(\omega') \vdash \mathsf{K}_C\Box_D\psi$ by the Monotonicity axiom, the Modus Ponens inference rule, and because $D \subseteq C$. Thus, $\mathsf{K}_C\Box_D\psi \in hd(\omega')$ because set $hd(\omega')$ is maximal. Hence, $\Box_D\psi \in hd(\omega)$ by Lemma 30 and Claim 2. At the same time, $s(a) = \delta'(a) = \psi$ for each agent $a \in D \subseteq C$ by equation (11) and equation (12). Therefore, $\psi \in X \subseteq X' = hd(\omega')$ by the choice of set X, set X', sequence ω', and assumption (12).

Case II: there is an agent $a \in D \setminus C$. Thus, $\delta'(a) = \top$ by equation (11). Hence, $\psi = \delta'(a) = \top$ by equation (12). Therefore, $\psi \in hd(\omega')$ because set $hd(\omega')$ is maximal. ⊠

Note that $\neg\varphi \in X \subseteq X' = hd(\omega')$ by the choice of set X, set X', sequence ω'. Therefore, $\varphi \notin hd(\omega')$ because set $hd(\omega')$ is consistent. This concludes the proof of the lemma. ⊠

Lemma 35. *For any outcome $\omega \in \Omega$, there is an initial state $\alpha \in I$ and a complete action profile $\delta \in \Delta^{\mathcal{A}}$ such that $(\alpha, \delta, \omega) \in P$.*

Proof. Let α be the equivalence class of ω with respect to relation $\sim_{\mathcal{A}}$. Thus, $\omega \in \alpha$. Let $\delta(a) = \top$ for each agent $a \in \mathcal{A}$.

Consider any formula $\Box_C\varphi \in hd(\omega)$ such that

$$\delta(a) = \varphi \quad \text{for each } a \in C. \tag{13}$$

By Definition 8, it suffices to show that $\varphi \in hd(\omega)$. We consider the following two cases separately:

Case I: $C \neq \varnothing$. Thus, there is $a_0 \in C$. Hence, $\varphi = \delta(a_0) = \top$ by equation (13) and the choice of the complete action profile δ. Therefore, $\varphi \in hd(\omega)$ because set $hd(\omega)$ is maximal.

Case II: $C = \varnothing$. Then, assumption $\Box_C\varphi \in hd(\omega)$ implies $hd(\omega) \vdash \varphi$ by Lemma 26 and the Modus Ponens inference rule. Therefore, $\varphi \in hd(\omega)$ because set $hd(\omega)$ is maximal. ⊠

Next we show that the canonical model satisfies the condition from item 5 of Definition 2.

Lemma 36. *For each initial state $\alpha \in I$ and each complete action profile $\delta \in \Delta^{\mathcal{A}}$, there is an outcome $\omega \in \Omega$ such that $(\alpha, \delta, \omega) \in P$.*

Proof. By Definition 6, initial state α is an equivalence class. Since each equivalence class is not empty, there must exist an outcome $\omega_0 \in \Omega$ such that $\omega_0 \in \alpha$. By Lemma 35, there is an initial state $\alpha_0 \in I$ and a complete action profile $\delta_0 \in \Delta^{\mathcal{A}}$ such that $(\alpha_0, \delta_0, \omega_0) \in P$. Then, $\omega_0 \in \alpha_0$ by Definition 8. Hence, ω_0 belongs to equivalence classes α and α_0. Thus, $\alpha = \alpha_0$. Therefore, $(\alpha, \delta_0, \omega_0) \in P$. ⊠

Lemma 37. *For any $(\alpha, \delta, \omega) \in P$ and any $\neg\mathsf{K}_C\varphi \in hd(\omega)$, there is a play $(\alpha', \delta', \omega') \in P$ such that $\alpha \sim_C \alpha'$ and $\neg\varphi \in hd(\omega')$.*

Proof. Consider the set $X = \{\neg\varphi\} \cup \{\psi \mid \mathsf{K}_C\psi \in hd(\omega)\}$. First, we show that set X is consistent. Suppose the opposite. Then, there are formulae $\mathsf{K}_C\psi_1, \dots, \mathsf{K}_C\psi_n \in hd(\omega)$ such that $\psi_1, \dots, \psi_n \vdash \varphi$. Hence, $\mathsf{K}_C\psi_1, \dots, \mathsf{K}_C\psi_n \vdash \mathsf{K}_C\varphi$ by Lemma 4. Thus, $hd(\omega) \vdash \mathsf{K}_C\varphi$ because $\mathsf{K}_C\psi_1, \dots, \mathsf{K}_C\psi_n \in hd(\omega)$. Hence, $\neg\mathsf{K}_C\varphi \notin hd(\omega)$ because set $hd(\omega)$ is consistent, which contradicts the assumption of the lemma. Therefore, set X is consistent.

By Lemma 5, there is a maximal consistent extension X' of set X. Let ω' be the sequence $\omega :: C :: X'$. Note that $\omega' \in \Omega$ by Definition 4 and the choice of sets X and X'. Also, $\neg\varphi \in X \subseteq X' = hd(\omega')$ by the choice of sets X and X'.

By Lemma 35, there is an initial state $\alpha' \in I$ and a complete action profile δ' such that $(\alpha', \delta', \omega') \in P$. Note that $\omega \sim_C \omega'$ by Definition 5 and the choice of sequence ω'. Thus, $\alpha \sim_C \alpha'$ by Lemma 32. ⊠

The next lemma is the "induction" lemma, also known as the "truth" lemma, that connects the syntax of our logical system with the semantics of the canonical model.

Lemma 38. *$(\alpha, \delta, \omega) \Vdash \varphi$ iff $\varphi \in hd(\omega)$ for each play $(\alpha, \delta, \omega) \in P$ and each formula $\varphi \in \Phi$.*

Proof. We prove the lemma by induction on the complexity of formula φ. If φ is a propositional variable, then the lemma follows from Definition 3 and Definition 9. If formula φ is an implication or a negation, then the required follows from the maximality and the consistency of set $hd(\omega)$ by Definition 3 in the standard way.

Assume that formula φ has the form $\mathsf{K}_C\psi$.

($\Rightarrow$): Let $\mathsf{K}_C\psi \notin hd(\omega)$. Thus, $\neg\mathsf{K}_C\psi \in hd(\omega)$ by the maximality of set $hd(\omega)$. Hence, by Lemma 37, there is a play $(\alpha', \delta', \omega') \in P$ such that $\alpha \sim_C \alpha'$ and $\neg\psi \in hd(\omega')$. Then, $\psi \notin hd(\omega')$ by the consistency of set $hd(\omega')$. Thus, $(\alpha', \delta', \omega') \nVdash \psi$ by the induction hypothesis. Therefore, $(\alpha, \delta, \omega) \nVdash \mathsf{K}_C\psi$ by Definition 3.

($\Leftarrow$): Let $\mathsf{K}_C\psi \in hd(\omega)$. Thus, $\psi \in hd(\omega')$ for any $\omega' \in \Omega$ such that $\omega \sim_C \omega'$, by Lemma 30. Hence, by the induction hypothesis, $(\alpha', \delta', \omega') \Vdash \psi$ for each play $(\alpha', \delta', \omega') \in P$ such that $\omega \sim_C \omega'$. Thus, $(\alpha', \delta', \omega') \Vdash \psi$ for each $(\alpha', \delta', \omega') \in P$ such that $\alpha \sim_C \alpha'$, by Lemma 32. Therefore, $(\alpha, \delta, \omega) \Vdash \mathsf{K}_C\psi$ by Definition 3.

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

($\Rightarrow$): Suppose $\mathsf{B}_C\psi \notin hd(\omega)$. Thus, $\neg\mathsf{B}_C\psi \in hd(\omega)$ because set $hd(\omega)$ is maximal. Hence, by Lemma 21,

$$hd(\omega) \vdash \neg\psi \vee \neg\Box_C\neg\psi \vee \bigvee_{D \subsetneq C} \Box_D\neg\psi.$$

Then, because set $hd(\omega)$ is maximal, one of the following cases takes place:

Case I: $\neg\psi \in hd(\omega)$. Thus, $\psi \notin hd(\omega)$ because set $hd(\omega)$ is consistent. Hence, $(\alpha,\delta,\omega) \nVdash \psi$ by the induction hypothesis. Therefore, $(\alpha,\delta,\omega) \nVdash \mathsf{B}_C\psi$ by item 5(a) of Definition 3.

Case II: $\neg\Box_C\neg\psi \in hd(\omega)$. Hence, by Lemma 34, for any action profile $s \in \Delta^C$ there is a play $(\alpha',\delta',\omega') \in P$ such that $\alpha \sim_C \alpha'$, $s =_C \delta'$, and $\neg\psi \notin hd(\omega')$. Thus, because set $hd(\omega')$ is maximal, for any action profile $s \in \Delta^C$ there is a play $(\alpha',\delta',\omega') \in P$ such that $\alpha \sim_C \alpha'$, $s =_C \delta'$, and $\psi \in hd(\omega')$. Then, by the induction hypothesis, for any action profile $s \in \Delta^C$ there is a play $(\alpha',\delta',\omega') \in P$ such that $\alpha \sim_C \alpha'$, $s =_C \delta'$, and $(\alpha',\delta',\omega') \Vdash \psi$. Therefore, $(\alpha,\delta,\omega) \nVdash \mathsf{B}_C\psi$ by item 5(b) of Definition 3.

Case III: there is a proper subset $D \subsetneq C$ such that $\Box_D\neg\psi \in hd(\omega)$. Thus, by Lemma 33, there is an action profile $s \in \Delta^D$ such that for each play $(\alpha',\delta',\omega') \in P$, if $\alpha \sim_D \alpha'$ and $s =_D \delta'$, then $\neg\psi \in hd(\omega')$. Hence, because set $hd(\omega')$ is consistent, for each play $(\alpha',\delta',\omega') \in P$, if $\alpha \sim_D \alpha'$ and $s =_D \delta'$, then $\psi \notin hd(\omega')$. Then, by the induction hypothesis, for each play $(\alpha',\delta',\omega') \in P$, if $\alpha \sim_D \alpha'$ and $s =_D \delta'$, then $(\alpha',\delta',\omega') \nVdash \psi$. Therefore, $(\alpha,\delta,\omega) \nVdash \mathsf{B}_C\psi$ by item 5(c) of Definition 3.

($\Leftarrow$): Let $\mathsf{B}_C\psi \in hd(\omega)$. We prove that $(\alpha,\delta,\omega) \Vdash \mathsf{B}_C\varphi$ by verifying conditions 5(a), 5(b), and 5(c) of Definition 3 separately.

(a). Assumption $\mathsf{B}_C\psi \in hd(\omega)$ implies $hd(\omega) \vdash \varphi$ by the Truth axiom and the Modus Ponens inference rule. Hence, $\varphi \in hd(\omega)$ because set $hd(\omega)$ is maximal. Therefore, $(\alpha,\delta,\omega) \Vdash \varphi$ by the induction hypothesis.

(b). Assumption $\mathsf{B}_C\psi \in hd(\omega)$ implies $hd(\omega) \vdash \Box_C\neg\psi$ by Lemma 22 and the Modus Ponens inference rule. Hence, $\Box_C\neg\psi \in hd(\omega)$ because set $hd(\omega)$ is maximal. Thus, by Lemma 33, there is an action profile $s \in \Delta^C$ such that for each play $(\alpha',\delta',\omega') \in P$ if $\alpha \sim_C \alpha'$ and $s =_C \delta'$, then $\neg\psi \in hd(\omega')$. Then, because set $hd(\omega')$ is consistent, for each play $(\alpha',\delta',\omega') \in P$ if $\alpha \sim_C \alpha'$ and $s =_C \delta'$, then $\psi \notin hd(\omega')$. Therefore, by the induction hypothesis, for each play $(\alpha',\delta',\omega') \in P$ if $\alpha \sim_C \alpha'$ and $s =_C \delta'$, then $(\alpha',\delta',\omega') \nVdash \psi$.

(c). Consider any proper subset $D \subsetneq C$ and any action profile $s \in \Delta^D$. Assumption $\mathsf{B}_C\psi \in hd(\omega)$ implies $hd(\omega) \vdash \neg\Box_D\neg\psi$ by Lemma 23. Then, because set $hd(\omega)$ is maximal, $\neg\Box_D\neg\psi \in hd(\omega)$. Thus, by Lemma 34, there is a play $(\alpha',\delta',\omega') \in P$ such that $\alpha \sim_C \alpha'$, $s =_C \delta'$, and $\neg\psi \notin hd(\omega')$. Hence, because set $hd(\omega')$ is maximal, $\psi \in hd(\omega')$. Therefore, $(\alpha',\delta',\omega') \Vdash \psi$ by the induction hypothesis.

This concludes the proof of the lemma. ⊠

Finally, we are ready to state and prove the strong completeness of our logical system.

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

Proof. Assume that $X \nvdash \varphi$. Hence, set $X \cup \{\neg\varphi\}$ is consistent. By Lemma 5, there is a maximal consistent extension X_0 of set $X \cup \{\neg\varphi\}$. Let game $(I, \{\sim_a\}_{a\in\mathcal{A}}, \Delta, \Omega, P, \pi)$ be the canonical game $G(X_0)$. Also, let ω_0 be the single-element sequence X_0. Note that $\omega_0 \in \Omega$ by Definition 4. By Lemma 35, there is an initial state $\alpha \in I$ and a complete action profile $\delta \in \Delta^{\mathcal{A}}$ such that $(\alpha,\delta,\omega_0) \in P$. Hence, $(\alpha,\delta,\omega_0) \Vdash \chi$ for each $\chi \in X$ and $(\alpha,\delta,\omega_0) \Vdash \neg\varphi$ by Lemma 38 and the choice of set X_0. Therefore, $(\alpha,\delta,\omega_0) \nVdash \varphi$ by Definition 3. ⊠

8. Conclusion

In this article we proposed a definition of blameworthiness in strategic games with imperfect information. A coalition C is blamable for the statement φ if φ is true and C is a *minimal* coalition that had a *know-how* strategy to prevent φ. This work significantly extends our definition of blameworthiness for games with perfect information [14] by adding the minimality requirement on the coalition and the know-how requirement on the strategy. The main technical result is a sound and complete logical system that describes the interplay between the distributed knowledge and the blameworthiness modalities in the imperfect information setting. Because of the addition of the minimality and the know-how requirements, the proof of the completeness is significantly longer and substantially different from the one in [14]. Namely, the proof defines a counterfactual modality through the distributed knowledge and the blameworthiness modalities and uses the counterfactual modality to construct a canonical game.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] P. Singer, M. Eddon, Moral responsibility, problem of, Encyclopædia Britannica, https://www.britannica.com/topic/problem-of-moral-responsibility, 2013.
[2] L. Fields, Moral beliefs and blameworthiness: introduction, Philosophy 69 (270) (1994) 397–415.
[3] J.M. Fischer, M. Ravizza, Responsibility and Control: A Theory of Moral Responsibility, Cambridge University Press, 2000.
[4] S. Nichols, J. Knobe, Moral responsibility and determinism: the cognitive science of folk intuitions, Nous 41 (4) (2007) 663–685.
[5] E. Mason, Moral ignorance and blameworthiness, Philos. Stud. 172 (11) (2015) 3037–3057.

[6] D. Widerker, Moral Responsibility and Alternative Possibilities: Essays on the Importance of Alternative Possibilities, Routledge, 2017.
[7] H.G. Frankfurt, Alternate possibilities and moral responsibility, J. Philos. 66 (23) (1969) 829–839.
[8] F. Cushman, Deconstructing intent to reconstruct morality, Curr. Opin. Psychol. 6 (2015) 97–103.
[9] D. Lewis, Counterfactuals, John Wiley & Sons, 2013.
[10] J.Y. Halpern, Actual Causality, MIT Press, 2016.
[11] V. Batusov, M. Soutchanski, Situation calculus semantics for actual causality, in: Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, AAAI-18, 2018.
[12] J.Y. Halpern, M. Kleiman-Weiner, Towards formal definitions of blameworthiness, intention, and moral responsibility, in: Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, AAAI-18, 2018.
[13] N. Alechina, J.Y. Halpern, B. Logan, Causality, responsibility and blame in team plans, in: Proceedings of the 16th Conference on Autonomous Agents and MultiAgent Systems, International Foundation for Autonomous Agents and Multiagent Systems, 2017, pp. 1091–1099.
[14] P. Naumov, J. Tao, Blameworthiness in strategic games, in: Proceedings of Thirty-Third AAAI Conference on Artificial Intelligence, AAAI-19, 2019.
[15] American Law Institute, 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, 1985 Print.
[16] M. Pauly, Logic for social software, Ph.D. thesis, Institute for Logic, Language, and Computation, 2001.
[17] M. Pauly, A modal logic for coalitional power in games, J. Log. Comput. 12 (1) (2002) 149–166, https://doi.org/10.1093/logcom/12.1.149.
[18] T. Ågotnes, N. Alechina, Coalition logic with individual, distributed and common knowledge, J. Log. Comput. (01 2016), https://doi.org/10.1093/logcom/exv085.
[19] P. Naumov, J. Tao, Coalition power in epistemic transition systems, in: Proceedings of the 2017 International Conference on Autonomous Agents and Multiagent Systems, AAMAS, 2017, pp. 723–731.
[20] R. Fervari, A. Herzig, Y. Li, Y. Wang, Strategically knowing how, in: Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI-17, 2017, pp. 1031–1038.
[21] P. Naumov, J. Tao, Together we know how to achieve: an epistemic logic of know-how, in: 16th Conference on Theoretical Aspects of Rationality and Knowledge, TARK, July 24–26, 2017, in: EPTCS, vol. 251, 2017, pp. 441–453.
[22] P. Naumov, J. Tao, Together we know how to achieve: an epistemic logic of know-how, Artif. Intell. 262 (2018) 279–300, https://doi.org/10.1016/j.artint.2018.06.007.
[23] P. Naumov, J. Tao, Strategic coalitions with perfect recall, in: Proceedings of Thirty-Second AAAI Conference on Artificial Intelligence, 2018.
[24] P. Naumov, J. Tao, Second-order know-how strategies, in: Proceedings of the 2018 International Conference on Autonomous Agents and Multiagent Systems, AAMAS, 2018, pp. 390–398.
[25] P. Naumov, J. Tao, Knowing-how under uncertainty, Artif. Intell. 276 (2019) 41–56.
[26] Y. Wang, A logic of knowing how, in: Logic, Rationality, and Interaction, Springer, 2015, pp. 392–405.
[27] Y. Wang, A logic of goal-directed knowing how, Synthese (2016) 1–21.
[28] E. Lorini, F. Schwarzentruber, A logic for reasoning about counterfactual emotions, Artif. Intell. 175 (3) (2011) 814.
[29] M. Gant, Millennials being blamed for decline of American cheese, Fox News, www.foxnews.com/food-drink/millennials-kraft-american-cheese-sales-decline.amp, October 11, 2018.
[30] P. Naumov, J. Tao, Knowledge and blameworthiness, arXiv:1811.02446, 2018.
[31] V. Yazdanpanah, M. Dastani, W. Jamroga, N. Alechina, B. Logan, Strategic responsibility under imperfect information, in: Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems, International Foundation for Autonomous Agents and Multiagent Systems, 2019, pp. 592–600.
[32] M. Xu, Axioms for deliberative stit, J. Philos. Log. 27 (5) (1998) 505–552.
[33] J. Broersen, A. Herzig, N. Troquard, What groups do, can do, and know they can do: an analysis in normal modal logics, J. Appl. Non-Class. Log. 19 (3) (2009) 261–289, https://doi.org/10.3166/jancl.19.261-289.
[34] P. Naumov, J. Tao, Blameworthiness in security games, in: Proceedings of Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI-20, 2020.
[35] E. Mendelson, Introduction to Mathematical Logic, CRC Press, 2009.
[36] R. Fagin, J.Y. Halpern, Y. Moses, M.Y. Vardi, Reasoning About Knowledge, MIT Press, Cambridge, MA, 1995.