Coalitional Responsibility in Strategic Settings

Nils Bulling[1] and Mehdi Dastani[2]

[1] Clausthal University, Germany
bulling@in.tu-clausthal.de
[2] Utrecht University, The Netherlands
M.M.Dastani@uu.nl

Abstract. This paper focuses on the concept of group responsibility and presents a formal analysis of it from a strategic point of view. A group of agents is considered to be responsible for an outcome if the group can avoid the outcome. Based on this interpretation of group responsibility, different notions of group responsibility are provided and their properties are studied. The formal analysis starts with the semantics of different notions of group responsibility followed by their logical characterizations. The presented work is compared and related to the existing work on responsibility.

1 Introduction

Responsibility is a central concept in philosophy and social sciences. Various types of responsibility such as moral, legal, social, and organizational responsibility have been identified [10]. Moreover, responsibility is classified along different dimensions such as individual or collective, normative or descriptive, forward-looking or backward-looking, and action-based or state-based [16]. An example of an individual forward-looking responsibility is the obligation of an academic researcher to see to it that the outcome of his research is truthful, not plagiarized, original, etc. This responsibility can be moral, legal, social or organizational. In general, responsibility that is based on the obligation to see to it that a state of affairs is the case is often seen as a forward-looking responsibility. An example of a collective backward-looking responsibility is the responsibility for the low-ranked teaching quality of a university department. In such a case, the teaching members of the department can collectively be held responsible for the low teaching quality.

The attribution of responsibility to agents or groups of agents is often characterized by means of specific (fairness) conditions [11,15]. For example, the conditions that characterize accountability and blameworthiness, which are considered as two instances of backward-looking responsibility, are formulated as follows. Agents can be held accountable for a state of affairs (or an action) if they have intentionally, deliberately and actively been involved in realizing the state of affairs (or the action). On the other hand, agents can be blamed for a state of affairs (or an action) if they can be held accountable for it, and moreover, the involvement is based on free choice (agents were not enforced or compelled), and they know that the state of affairs (or the action) has negative consequences.

J. Leite et al. (Eds.): CLIMA XIV, LNAI 8143, pp. 172–189, 2013.

Although the concept of responsibility has been studied for quite a long time, there is yet no consensus of what this concept exactly and formally means. Most formal work on responsibility is concerned with specific instantiation of this concept such as having moral (legal, social, or organizational) obligations and being accountable or blameworthy for something. To our knowledge there is not much work on formalizing and analyzing the very abstract concept of responsibility without considering its instantiations. The abstract concept of responsibility that we have in mind captures the power dimension of responsibility as illustrated by the quote "*With great power, comes great responsibility*". More specifically, this notion of responsibility can be used to hold a group of agents responsible for a state of affairs if they can ensure avoiding the state of affairs. In other words, agents can be held responsible for a state of affairs if they have the power to preclude the state of affairs.

This notion of responsibility is neither forward-looking nor backward-looking since it neither requires the agents to see to it that a state of affairs is the case nor implies that the agents are accountable because the state of affairs may not be realized. Moreover, most work on responsibility is concerned with individual agents, ignoring responsibility of coalitions of agents with a strategic flavor. A coalition of agents can be held responsible for some state of affairs due to strategic reasons. For example, two political parties that jointly have a majority in the parliament are responsible for the enactment of a law because they can form a coalition to block the enactment of the law. Note that the involved agents can also strategically reason and decide to be absent at the voting session in order to abdicate their responsibilities. Our proposed framework can be applied to analyze the responsibility of agent coalitions in multi-agent scenario's where different agents have different sets of actions/options available to them, e.g., elections and collective decision making, distributed problem solving and collaborative systems.

This paper aims at formalizing this abstract concept of state-based responsibility for coalitions of agents. We consider a coalition of agents as being responsible for some states of affairs if the coalition can preclude it. This abstract notion of responsibility is formalized in concurrent game structures where the strategic behavior of a set of agents can be represented and analyzed. The proposed framework allows defining various notions of this abstract concept of responsibility. It also allows reasoning about responsibilities of agents' coalitions and deciding which coalition of agents is responsible for specific states of the system.

The structure of this paper is as follows. Section 2 presents the formal framework in which the notion of responsibility is characterized. Section 3 provides a semantic analysis of various notions of group responsibility and study their properties. In Section 4 we show how a coalition logic with quantification can be used to characterize and to reason about group responsibility. The provided notion of group responsibility is put in the context of related work in Section 5. Finally, Section 6 concludes the paper and we point out some future work directions.

2 Preliminaries: Models and Power

In this paper, the behavior of a multi-agent system is modeled by concurrent game structures (CGS). A *concurrent game structure* [3] (CGS) is a tuple $\mathcal{M} = (N, Q, Act, d, o)$

which includes a nonempty finite set of all agents $N = \{1, \ldots, k\}$, a nonempty set of system states Q, and a nonempty finite set of (atomic) actions Act. The function $d : N \times Q \to \mathcal{P}(Act)$ defines sets of actions available to agents at each state, and o is a deterministic and partial transition function that assigns the outcome state $q' = o(q, \alpha_1, \ldots, \alpha_k)$ to state q and a tuple of actions $\alpha_i \in d(i, q)$ that can be executed by N in q. An action profile is a sequence $(\alpha_1, \ldots, \alpha_k)$ consisting of an action for each player. We require that if $o(q, \alpha_1, \ldots, \alpha_k)$ is undefined then $o(q, \alpha'_1, \ldots, \alpha'_k)$ is undefined for each action profile $(\alpha'_1, \ldots, \alpha'_k)$. We write $d_i(q)$ for $d(i, q)$ and $d_C(q) := \prod_{i \in C} d_i(q)$.

A *state of affairs* is defined as a set $S \subseteq Q$ of states. In the rest of this paper, we use $\bar{S}$ to denote the set $Q \backslash S$ of states. Let $\mathcal{M}$ be a CGS, q a state in it and S be a state of affairs. We say that:

– *C can q-enforce S in $\mathcal{M}$* iff there is a joint action $\alpha_C \in d_C(q)$ such that for all joint actions $\alpha_{N \backslash C} \in d_{N \backslash C}(q)$ we have that $o(q, (\alpha_C, \alpha_{N \backslash C})) \in S$. That is, coalition C must have an action profile that guarantees to end up in a state from S, independent of what the agents outside C do.

– *C q-controls S in $\mathcal{M}$* iff C can q-enforce S as well as $\bar{S}$ in $\mathcal{M}$.

– *C can q-avoid S in $\mathcal{M}$* iff for all $\alpha_{N \backslash C} \in d_{N \backslash C}(q)$ there is $\alpha_C \in d_C(q)$ such that $o(q, (\alpha_{N \backslash C}, \alpha_C)) \in \bar{S}$.

In the following we shall omit "in $\mathcal{M}$" whenever $\mathcal{M}$ is clear from context. We note that the notions of enforcement and avoidance correspond to the game-theoretic notions of α-effectivity and β-effectivity, respectively (e.g. [13]). More, precisely, we have that C can q-enforce S in $\mathcal{M}$ iff C is α-effective for S in q; and C can q-avoid S in $\mathcal{M}$ iff C is β-effective for $\bar{S}$ in q[1].

In general, a coalition that q-controls S is not unique; that is, there is a CGS $\mathcal{M}$, state q, state of affairs S, and different coalitions C and C' that q-control S. In this case we have that $C \cap C' \neq \emptyset$. Moreover, if C can q-enforce $\bar{S}$ then C can q-avoid S.

It is often the case that agents have incomplete information about the world. In CGSs this is modeled by equivalence relations $\sim_a$, one for each $a \in N$. A *uniform* strategy for a player a is a function $s_a : Q \to Act$ such that $s_a(q) = s_a(q')$ for all $q \sim_a q'$. A collective uniform strategy for C is a tuple of strategies consisting of a uniform strategy for each member of C. Moreover, we defined the mutual knowledge relation $\sim_C$ as $\bigcup_{a \in C} \sim_a$. Consequently, we say that a coalition *knows* that it can q-enforce S in $\mathcal{M}$ if there is a collective joint uniform strategy s_C of C such that for all states q' with $q \sim_C q'$ and all actions $\alpha_{N \backslash C} \in d_{N \backslash C}(q')$ we have that $o(q', (\alpha_C, \alpha_{N \backslash C})) \in S$. Analogously, we say that C knows that it q-controls and can q-avoid S.

3 Coalitional Strategic Responsibility

This section provides a semantical analysis of various notions of group responsibility. The intuitive idea of responsibility that we have in mind is that a group of agents can be said to be responsible for some state of affairs if they have the preclusive power to prevent the state of affairs, regardless of what the other agents can do. Under this interpretation, a group of agents is responsible for a state of affairs in the sense that the state of affairs can only be realized if they allow the state of affairs to become the case.

[1] In this context, we consider the normal form game naturally associated to the state q in $\mathcal{M}$.

3.1 Basic Definitions of Responsibility

In the following let $\mathcal{M}$ be a CGS, q a state of $\mathcal{M}$ and S a state of affairs in $\mathcal{M}$. We consider two definitions of responsibility. Both notions are *preclusive* in the sense of [12]. The first notion assigns a coalition responsible for a state of affairs if it is the smallest coalition (provided it exists) that can prevent that state of affairs. Our concept of responsibility is local in the sense that it is defined regarding some origin state. A coalition can be responsible for a state of affairs from some state and not responsible from others.

Definition 1 (Responsibility). *We say that a group $C \subseteq N$ is* q-responsible for S in $\mathcal{M}$ *iff C can q-enforce $\bar{S}$ and for all other coalitions C' that can q-enforce $\bar{S}$ we have that $C \subseteq C'$.*

Again, we omit "in $\mathcal{M}$" if clear from context and proceed in the same way in the rest of the paper. This definition ensures that a coalition is q-responsible for S if there is no other coalition that does not contain the coalition and which can prevent S. This notion of responsibility has the property that a responsible coalition is unique.

Proposition 1. *If C_1 and C_2 are q-responsible for S in $\mathcal{M}$ then $C_1 = C_2$.*

The proposition shows that responsibility is a *very strong concept*. Often there is no smallest group of agents which can preclude a state of affairs. This is for example the case when there are agents with identical preclusive powers. The next definition captures this intuition. A coalition is *weakly responsible* for a state of affairs if it has the power to preclude it and if the coalition is minimal. We do not require, however, that it is the smallest coalition having such preclusive power. It is important to note that if there are some weakly responsible coalition but no responsible one that does *not* mean that there is not responsible coalition in the colloquial sense. It simply means that there are several coalitions that are responsible–again, in the colloquial sense–but no unique one.

Definition 2 (Weak Responsibility). *We say that a group $C \subseteq N$ is* weakly q-responsible for S in $\mathcal{M}$ *iff C is a minimal coalition that can q-enforce $\bar{S}$.*

We note that both notions of responsibility are based on *preclusive power* in terms of enforcement and not in terms of avoidance. Clearly, we have the following result.

Proposition 2. *If C is q-responsible for S then it is also weakly q-responsible for S and there is no other weakly q-responsible coalition for S. Also if $\emptyset$ is weakly q-responsible for S; then, $\emptyset$ is q-responsible for S.*

Proof. The first part of the proof is obvious. Suppose C' is a weakly q-responsible coalition for S with $C' \neq C$. We cannot have $C' \subsetneq C$ as this would contradict the minimality of C. Analogously, we cannot have $C \subsetneq C'$. Thus, we must have $C \not\subseteq C'$ which contradicts that C is q-responsible for S. Clearly, if $\emptyset$ is weakly q-responsible for S; then, it is the smallest such coalition. □

Example 1. We consider the CGS shown in Figure 1[2]. We refer to player 1 as "Driver 1", to 2 as "Driver 2", and to 3 as "family member of Driver 2". The story is as follows.

[2] We thank an anonymous CLIMA reviewer for this example.

Two drivers can decide to drive or to wait. If both chose to drive their cars will crash, with one exception: a family member of Driver 2 can poison Driver 2, making him/her unable to drive and thus avoids a crash. In this example the weakly q_0-responsible coalitions for $\{q_2\}$ are exactly $\{1\}$, $\{2\}$, and $\{3\}$. However, no coalition is q_0-responsible for $\{q_2\}$! Again, it is important to note that this does not mean that no coalition is responsible in the colloquial sense but simply that there are three (weakly) responsible coalitions.

Also note that our notion of responsibility is *free* of any moral connotation. The family member who has not poisoned the driver is as responsible for a crash (i.e. state $\{q_2\}$) as Driver 1 and Driver 2; although, intuitively poisoning should not be a serious alternative.

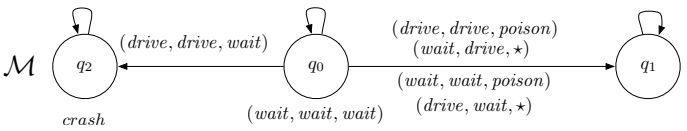

Fig. 1. The CGS $\mathcal{M}_1 = (\{1,2,3\}, \{q_0, q_1, q_2\}, \{drive, wait, poison\}, d, o)$ where $d_1(q_0) = d_2(q_0) = \{drive, wait\}, d_3(q_0) = \{poison, wait\}$ and $d_i(q) = \{wait\}$ for all $i \in \{1,2,3\}$ and $q \in \{q_1, q_2\}$. The outcome function o is shown in the figure, e.g. $o(q_0, (drive, drive, wait)) = q_2$. The star $\star$ represents any available action, i.e. $\star \in \{wait, poison\}$.

3.2 Degrees of Responsibility: Crucial and Necessary Coalitions

Responsible as well as weakly responsible coalitions have the preclusive power to prevent a specific state of affairs. A natural question is whether all members of a coalition are equally responsible or if it is possible to assign different degrees of responsibility to subcoalitions of agents.

Crucial Coalitions. Firstly, we consider subcoalitions of a responsible coalition which cannot be replaced by other coalitions without losing their status of being responsible. We call such responsible subcoalition the crucially responsible coalition, or simply, a crucial coalition.

Definition 3 (Crucial coalition). *Let C be (weakly) q-responsible for S in $\mathcal{M}$. We say that a (sub)coalition $\hat{C} \subseteq C$ is* q-crucial for S in C *and $\mathcal{M}$ iff for all coalitions $C' \subseteq N$, if $(C \backslash \hat{C}) \cup C'$ is weakly q-responsible for S then $\hat{C} \subseteq C'$.*

Example 2. Let $N = \{1,2,3,4\}$ and $\mathcal{M} = (N, \{q_0, q_1, q_2\}, \{1,2\}, d, o)$ with $d(q_0) = \{1,2\}$, $d_i(q_2) = d_i(q_1) = \{1\}$, and $d_i(q_1) = d_i(q_2) = \emptyset$ where $i \in N$. The transition function is defined as follows $o(q_0, (1,1,1,\star)) = o(q_0, (\star, 2, \star, 2)) = q_2$, and $o(q_0, \alpha) = q_1$ for $\alpha \in d_N(q_0) \backslash \{(1,1,1,\star), (\star,2,\star,2)\}$ where $\star \in \{1,2\}$. We have that $C_1 = \{1,2,3\}$ and $C_2 = \{2,4\}$ are the weakly q_0-responsible coalitions for $S = \{q_1\}$. We also have that all subsets of C_1 except $\{1,3\}$ and $\{1,2,3\}$ are q_0-crucial for S in C_1. For example, to see that $\hat{C} = \{2,3\}$ is q_0-crucial for S in C_1, we have to check if for all $C' \subseteq N$ it holds that if $(\{1,2,3\} \setminus \{2,3\}) \cup C'$ is weakly q_0-responsible for S in C_1 (i.e., if $\{1\} \cup C' \in \{C_1, C_2\}$), then $\hat{C} = \{2,3\} \subseteq C'$. Clearly, the antecedent is only true if C' equals $\{2,3\}$ or $\{1,2,3\}$. In both cases we have that $\hat{C} \subseteq C'$. As the

previous case shows, we can replace $\hat{C}$ by $C' = \{2, 4\}$ in C_1 and the resulting coalition can q-enforce $\bar{S}$; though, it is not minimal. Moreover, note that $\hat{C}$ is not q_0-crucial for S in C_2 because it can be replaced by $\{2\}$. Similarly, we have that $\{1\}$, $\{2\}$, and $\{1, 2\}$ are q_0-crucial for S in C_1. $\{1\}$ and $\{2\}$ are q_0-crucial for S in C_2. To some extend one may argue that $\{2\}$ is more responsible than $\{2, 3\}$ as it is crucial in C_1 as well as in C_2. We will further discuss the latter statement.

Note that a weakly responsible coalition can have several crucial subcoalitions, i.e., in general a crucial coalition is not unique. In the following proposition we analyze some properties of crucial coalitions.

(i) The first property states that subcoalitions that are crucial for a weakly responsible coalition are characteristic for the weakly responsible coalition, i.e., they cannot be replaced to form a different weakly responsible coalition. (ii) The second property states that cruciality is closed under subset relation in the sense that crucial coalitions of one (weak) responsible coalition cannot have non-crucial subcoalitions. (iii) The third property states that the intersection of all weakly responsible coalitions is always crucial for all these weakly responsible coalitions. Note that the empty coalition is always crucial. (iv) The fourth property states that cruciality is not closed under union, i.e., the union of two crucial coalition is not necessarily crucial. (v) The fifth property states that the proper subsets of non-overlapping weakly responsible coalitions are crucial while the weakly responsible coalitions themselves are not crucial when there is more than one. (vi) Finally, the sixth property states that the subtraction of weakly responsible coalitions is not a crucial coalition.

Proposition 3 (Properties). *Let C be weakly q-responsible for S in $\mathcal{M}$ and $\hat{C}$ be q-crucial for S in C.*

1. *For any $C' \subseteq N$ such that $(C \backslash \hat{C}) \cup C'$ is weakly q-responsible for S we have that $\hat{C} \subseteq C' \subseteq C$; hence, $(C \backslash \hat{C}) \cup C' = C$.*
2. *Any subcoalition $\hat{C}' \subseteq \hat{C}$ is q-crucial for S in C. In particular, this shows that cruciality is closed under intersection and subtraction: if $\hat{C}_1$ and $\hat{C}_2$ are q-crucial for S in C; then, so is $\hat{C}_1 \cap \hat{C}_2$, $\hat{C}_1 \backslash \hat{C}_2$, and $\hat{C}_2 \backslash \hat{C}_1$.*
3. *Let W be the set of all weakly q-responsible coalitions for S. Then, $\bigcap W$ is q-crucial for S for all coalitions in W.*
4. *Given another q-crucial coalition $\hat{C}'$ for S in C the union $\hat{C} \cup \hat{C}'$ is not necessarily q-crucial for S in C.*
5. *Let W be the set of all weakly q-responsible coalitions for S such that for all $C_i, C_j \in W$ with $i \neq j$ it holds $C_i \cap C_j = \emptyset$. Then, every strict subcoalition $\hat{C} \subset C \in W$ is q-crucial for S in C. Moreover, if $|W| > 1$, then coalition $C \in W$ is not q-crucial for S in C.*
6. *If C_1 and C_2 are weakly q-responsible coalitions for S in $\mathcal{M}$ and $C_1 \setminus C_2 \neq \emptyset$ and $C_1 \neq \emptyset \neq C_2$, then $C_1 \setminus C_2$ is not q-crucial for S in C_1.*

Proof. **1.** By definition $\hat{C} \subseteq C'$. Now suppose that $C' \not\subseteq C$. Then, $Y := C' \backslash C \neq \emptyset$. We have $C = (C \backslash \hat{C}) \cup (C' \backslash Y) \subsetneq (C \backslash \hat{C}) \cup C'$. This shows that $(C \backslash \hat{C}) \cup C'$ is not a minimal coalition that can q-enforce S; hence, it is not weakly q-responsible for S. Contradiction! **2.** Clearly, this is the case for $\hat{C}' = \hat{C}$. Now, suppose there is a coalition

$\hat{C}' \subsetneq \hat{C}$ which is not q-crucial for S in C. Then, there is $C' \subseteq N$ such that $\hat{C}' \not\subseteq C'$ and $(C \backslash \hat{C}') \cup C'$ is weakly q-responsible for S in C. Let $Y := C' \backslash \hat{C}'$. We have that $Y \not\subseteq C$; for, if $Y \subseteq C$ then $(C \backslash \hat{C}') \cup C' \subsetneq C$. This would contradict the minimality of C. We define $D := (\hat{C} \backslash \hat{C}') \cup C'$. Because $Y \not\subseteq C$ we have $\hat{C} \not\subseteq D$. Moreover, $(C \backslash \hat{C}) \cup D = (C \backslash \hat{C}') \cup C'$ is weakly q-responsible for S. But this implies that $\hat{C}$ cannot be q-crucial for S in C. Contradiction! **3.** Suppose $\bigcap W$ is not q-crucial for S in $C \in W$. Then, there is $C' \subseteq N$, such that $(C \backslash \bigcap W) \cup C'$ is weakly q-responsible for S and $\bigcap W \not\subseteq C'$. But this contradicts $\bigcap W \subseteq (C \backslash \bigcap W) \cup C' \in W$. **4.** To see that consider the q_0-crucial coalitions $\{1, 2\}$ and $\{2, 3\}$ for S in C_1 from Example 2. The union equals $C_1 = \{1, 2, 3\}$ which is not q_0-crucial coalitions for S in C. **5.** Suppose $\hat{C} \subset C \in W$ is not q-crucial for S in C, i.e., for some $C' \subseteq N$ it holds that $(C \backslash \hat{C}) \cup C'$ is weakly q-responsible for S in C and $\hat{C} \not\subseteq C'$. We make the following case distinction: 1) $(C \setminus \hat{C}) \cup C' = C$ and 2) $(C \setminus \hat{C}) \cup C' = C^* \neq C$ for some $C^* \in W$. In the first case we have $\hat{C} \subseteq C'$. Contradiction. In the second case, we must have $\hat{C} = C$ and $C' = C^*$ because C^* and C are disjoint by assumption. This also yields a contradiction because we have assumed that $\hat{C} \subset C$. Now, let $C, C' \in W$ with $C \neq C'$ (note, by Proposition 2 no set can be empty). If C is q-crucial for S in C; then $C \subseteq C'$ because $C' = (C \backslash C) \cup C'$ is weakly q-responsible for S. But this is a contraction as C and C' are disjoint. **6.** Suppose $C_1 \setminus C_2$ were q-crucial for S in C_1. This implies that for all $C' \subseteq N$ it holds that if $(C_1 \setminus (C_1 \setminus C_2)) \cup C'$ is weakly q-responsible for S, then $(C_1 \setminus C_2) \subseteq C'$. Now take $C' = C_2$. By assumption $(C_1 \setminus (C_1 \setminus C_2)) \cup C_2 = C_2$ is weakly q-responsible for S in C_1. But we have $(C_1 \setminus C_2) \not\subseteq C_2$. Contradiction! □

The next lemma gives a characterization of responsible coalitions. As expected from Proposition 1 a responsible coalition consists only of crucial subcoalitions.

Lemma 1. *Coalition C is q-responsible for S in $\mathcal{M}$ iff every (sub)coalition $\hat{C} \subseteq C$ is q-crucial for S in C and $\mathcal{M}$.*

Proof. "⇒": Suppose C is q-responsible for S and there is a subcoalition $\hat{C} \subseteq C$ which is not q-crucial for S in C. Then, there is an C' with $\hat{C} \not\subseteq C'$ such that $(C \backslash \hat{C}) \cup C'$ is weakly q-responsible for S. This means that $(C \backslash \hat{C}) \cup C'$ can q-enforce $\bar{S}$ and that ($\star$) $C \not\subseteq (C \backslash \hat{C}) \cup C'$ which contradicts the assumption that C is q-responsible for S. To see that ($\star$) holds, we consider the following cases. (i) If $C' \subsetneq \hat{C}$ then $(C \backslash \hat{C}) \cup C' \subsetneq C$; hence, ($\star$). (ii) Let $Y := C' \backslash \hat{C} \neq \emptyset$. If $Y \subseteq C$ then $(C \backslash \hat{C}) \cup C' \subsetneq C$; hence, ($\star$); else, if $Y \not\subseteq C$ then $(C \backslash \hat{C}) \cup C' \not\subseteq C$. Hence, if it would be the case that $C \subseteq (C \backslash \hat{C}) \cup C'$ then also $C \subsetneq (C \backslash \hat{C}) \cup C'$. But this contradicts the minimality of $(C \backslash \hat{C}) \cup C'$ that has to hold because $(C \backslash \hat{C}) \cup C'$ is weakly q-responsible for S.

"⇐": Suppose every (sub)coalition $\hat{C} \subseteq C$ is q-crucial for S in C and that C is not q-responsible for S. Then, there is another coalition C' which can q-enforce $\bar{S}$ and $C \not\subseteq C'$. Let $C'' \subseteq C'$ be the coalition that is q-weakly responsible for S (it has to exist!). However, this means that C is not q-crucial for S in C, because $(C \backslash C) \cup C'' = C''$. This contradicts the assumption that every subset of C is crucial! □

Thanks to the previous lemma and Proposition 3.2 we obtain the following result relating responsible coalitions with crucial ones.

Proposition 4 (Characterization of responsibility). *A coalition C is q-responsible for S iff C is q-crucial for S in C.*

Proof. If C is q-responsible for S then C is q-crucial for S in C by Lemma 1. On the other hand, if C is q-crucial for S in C then any subcoalition is q-crucial for S in C by Proposition 3.2. Then, by Lemma 1 we can deduce that C is q-responsible for S. □

Necessary Coalitions. We consider subcoalitions of responsible coalitions with stronger properties. The notion of *necessary coalition* is stronger than the one of crucial coalitions in the sense that they are an indispensable part of any replacing coalition which maintains the preclusive power. This is realized by relaxing the condition of weak responsibility underlying the concept of cruciality.

Definition 4 (Necessary coalition). *Let C be (weakly) q-responsible for S. We say that a (sub)coalition $\hat{C} \subseteq C$ is q-necessary for S in C iff for all coalitions $C' \subseteq N$ it holds that if $(C \backslash \hat{C}) \cup C'$ can q-enforce $\bar{S}$ we have that $\hat{C} \subseteq C'$.*

Example 3. We continue Example 2. The coalition $\{2\}$ is q_0-necessary for S in C_1 as well as in C_2. Now, let C be any weakly q_0-responsible coalition for $\{q_2\}$ of Example 1. Then, the only q_0-crucial and q_0-necessary coalition of C is $\emptyset$. Intuitively, this shows that all coalitions are "equally responsible" in a colloquial sense.

Proposition 5 (Properties). *Let C be weakly q-responsible for S and $\hat{C}$ be q-necessary for S in C.*

1. *For any other coalition C' which is weakly q-responsible for S we have that $\hat{C} \subseteq C \cap C'$.*
2. *Let C' be another weakly q-responsible coalition for S. Then, $\hat{C}$ is q-necessary for S in C'.*
3. *$\hat{C}$ is q-crucial for S in C.*
4. *Given another q-necessary coalition $\hat{C}'$ for S in C the union $\hat{C} \cup \hat{C}'$ is also q-necessary for S in C.*

Proof. **1.** Let C and C' be two different weakly q-responsible coalitions for S as stated in the proposition. Any supercoalition of C' can q-enforce $\bar{S}$, in particular also $(C \backslash \hat{C}) \cup C'$. Because $\hat{C}$ is q-necessary for S in C we have $\hat{C} \subseteq C'$ which proves that $\hat{C} \subseteq C \cap C'$. **2.** By Proposition 5.1, $\hat{C} \subseteq C'$. Now suppose that $\hat{C}$ were not q-necessary for S in C'. Then, there is a coalition $C'' \subseteq N$, such that $\hat{C} \not\subseteq C''$ and $(C' \backslash \hat{C}) \cup C''$ can q-enforce $\bar{S}$. We have that $\hat{C} \not\subseteq (C' \backslash \hat{C}) \cup C''$. Moreover, $(C \backslash \hat{C}) \cup ((C' \backslash \hat{C}) \cup C'')$ can q-enforce $\bar{S}$. But this contradicts that $\hat{C}$ is q-necessary for S in C. **3.** Suppose $\hat{C}$ is not q-crucial for S in C. Then, there is an $C' \subseteq N$ such that $(C \backslash \hat{C}) \cup C'$ is weakly q-responsible for S and $\hat{C} \not\subseteq C'$. However, in particular $(C \backslash \hat{C}) \cup C'$ can q-enforce $\bar{S}$. This contradicts the assumption that $\hat{C}$ is q-necessary for S in C. **4.** Suppose that $\hat{C} \cup \hat{C}'$ were not q-necessary for S in C. Then, there is an $C' \subseteq N$ such that $(C \backslash (\hat{C} \cup \hat{C}')) \cup C'$ can q-enforce $\bar{S}$ and $\hat{C} \cup \hat{C}' \not\subseteq C'$. Then, $\hat{C} \not\subseteq C'$ or $\hat{C}' \not\subseteq C'$. Without loss of generality, assume that $\hat{C} \not\subseteq C'$. Because $(C \backslash (\hat{C} \cup \hat{C}')) \cup C'$ can q-enforce $\bar{S}$ we also have that $(C \backslash \hat{C}) \cup C'$ can q-enforce $\bar{S}$. But this means that $\hat{C}$ cannot be q-necessary for S in C. Contradiction! □

In particular, note that every necessary coalition is crucial and that necessary coalitions are closed under union which is not the case for crucial coalitions.

3.3 The Most Responsible Coalition

In this section we study a special type of necessary coalition, the *most responsible coalition*. In principle there can be many coalitions that are crucial or necessary for a weakly responsible coalition. In Proposition 5.2 we have shown that a coalition necessary for *some* weakly responsible coalition is necessary for *all* weakly responsible coalitions. In the next theorem we show that each weakly responsible coalition has a largest necessary coalition and that this is actually the largest necessary coalition in all weakly responsible coalitions. Hence, members of this coalition may be seen as more responsible than other members as they are part of all possible coalitions that can prevent a specific state of affairs.

Theorem 1 (Uniqueness). *Let coalition C be weakly q-responsible for S. Then, there is a unique maximal q-necessary coalition C^u for S in C and this coalition is also the unique maximal q-necessary coalition for S in any other coalition which is weakly q-responsible for S. In particular, if C is q-responsible for S then $C^u = C$.*

Proof. Let W be the set of all weakly q-responsible coalitions for S. By Proposition 5.4, each $C \in W$ has a largest q-necessary coalition for S in C. By Proposition 5.2 a q-necessary coalition for S for some coalition in W is q-necessary coalition for S for all coalitions in W. The claim follows. □

Definition 5 (Most responsible coalition). *We call the coalition C^u from Theorem 1 the* most q-responsible coalition for S.

In Proposition 5.3 we have shown that every necessary coalition is also crucial. Note, that the reverse is not necessarily true. The following lemma is important for our Characterization Theorem and shows that a coalition that is crucial in *all* weakly responsible coalitions is also necessary.

Lemma 2. *Suppose $\hat{C}$ is q-crucial for S in all weakly q-responsible coalitions for S. Then, $\hat{C}$ is q-necessary for S in all weakly q-responsible coalitions for S.*

Proof. Suppose the claim is false; then there is a weakly q-responsible coalitions C for S such that $\hat{C}$ is not q-necessary for S in C. Hence, there is a coalition $C' \subseteq N$ such that $(C \backslash \hat{C}) \cup C'$ can q-enforce S and $\hat{C} \not\subseteq C'$. Then, there also is a weakly q-responsible coalition $C'' \subseteq (C \backslash \hat{C}) \cup C'$ with $\hat{C} \not\subseteq C''$. Contradiction, as $\hat{C}$ is q-crucial for S for all weakly q-responsible coalitions for S by assumption. □

Finally, we show that exactly the agents that are part of all weakly responsible coalitions form the most responsible coalition which nicely matches with the intuition that these agents can be seen more responsible than others.

Theorem 2 (Characterization: most responsible coalition). *Let W be the set of all (weakly) q-responsible coalitions for S. The most q-responsible coalition C^u for S equals $\bigcap W$.*

Proof. Let C^u denote the most q-responsible coalition for S. "$C^u \subseteq \bigcap W$": By definition C^u is a member of any weakly q-responsible coalition C for S which shows

that $C^u \subseteq \bigcap W$. "$\bigcap W \subseteq C^u$": $\bigcap W$ is q-crucial for S in any $C \in W$ by Proposition 3.3. Thanks to Lemma 2 we have that $\bigcap W$ is also q-necessary for S. Then, because C^u is the largest q-necessary coalition in each weakly q-responsible coalition we have $\bigcap W \subseteq C^u$ by Theorem 2. □

Example 4. We continue Example 3. The coalition $\{2\}$ is the most q-responsible coalition for S. In Example 1, the most q_0-responsible coalition for $\{s_2\}$ is $\emptyset$ (cf. Example 3).

3.4 Evidence Sets and Responsibility

If a coalition C is q-responsible for S and we collect some evidence A by either observing or being informed about some of the agents' actions in q, then we can ask whether C can be held responsible for S in q under the collected evidence A. Moreover, we can ask which particular agents in C can be held responsible for S in q under the collected evidence A. On the other hand, we can ask which (minimal set of) evidence needs to be collected to hold a coalition or particular agents responsible for S in q.

The intuition for *holding* a group of agents responsible under an evidence set is as follows. Suppose a group C of agents is (weakly) q-responsible for some states S in a model $\mathcal{M}$ because they have actions that prevent the state of affairs S, i.e., C can prevent S in state q of $\mathcal{M}$. In Example 2 the group $C_2 = \{2, 4\}$ is weakly q_0-responsible for $S = \{q_1\}$ as they can prevent S by performing action 2. Suppose we have evidence that some agents have performed some actions. For example, we have evidence that agent 4 has performed action 1. This evidence can be used to modify the model $\mathcal{M}$ by removing the transitions that are inconsistent with the evidence, i.e., those transitions that contradict the evidence are removed from $\mathcal{M}$. In our example, we can remove actions $(*, *, *, 2)$ from the model presented in Example 2.

Now, the idea is that if the group C of agents is not (weakly) q-responsible for the states S in the modified model any more, then some of the agents from which the evidence is collected have decided not to prevent S such that these agents can be held responsible for S. In our example, agent 4 has performed action 1 which does not ensure the prevention of q_2. It has to be emphasized that the statement "C can be held responsible for S in q under the collected evidence" should be interpreted as "the evidence suggests that C has acted irresponsibly or incautious" since C has not performed actions to prevent S. This interpretation is aligned with the following quote on responsibility: "*It is not only for what we do that we are held responsible, but also for what we do not do*". It should also be stressed that under this interpretation it is not necessarily the case that S is actually being realized such that C cannot be held accountable or blameworthy for S in q. In our example, the evidence that agent 4 has performed action 1 does not imply that state q_1 is realized. The collected evidence indicates that C_2 is not q-responsible for S any more which means that in q the agent group C_2 has no preclusive power for S any more.

Formally, we assume that we are given an *evidence set* $A \subseteq Q \times N \times Act$ of (occurred) events. A tuple (q, i, α_i) states that agent i has executed action α_i in state q. We assume that our information is correct; that is, for all states q and players i there is at most one tuple $(q, i, \alpha) \in A$ and if such a tuple exists that $\alpha \in d_i(q)$. Given an evidence set A, we use $\mathcal{M}|_A$ to denote the update of model $\mathcal{M}$ that is obtained by

removing all transitions not consistent with A from $\mathcal{M}$. Note that the update operation here is considered as a meta-model operation.

Definition 6. *Let coalition C be (weakly) q-responsible for S in $\mathcal{M}$ and A be an evidence set. The coalition C can be* held responsible for S in q under evidence A *if C is not (weakly) q-responsible for S in $\mathcal{M}|_A$. Moreover, a subset $A' \subseteq A$ is said to be a* relevant evidence set of A for coalition C *with respect to S and $\mathcal{M}$ (i.e., C can be held responsible for S in $\mathcal{M}$ based on A') iff A' is a minimal subset of A such that C is not (weakly) q-responsible for S in $\mathcal{M}|_{A'}$ but it is (weakly) q-responsible for S in $\mathcal{M}|_{A\setminus A'}$.*

Let $\mathcal{A}$ be the set of all possible evidence sets and $Ag : \mathcal{A} \to 2^N$ be a function that determines the agents from which evidence is collected. The following proposition states that the evidence under which an agent can be held responsible should be about the agent's actions. The result follows directly from the fact that the loss of preclusive power of an agent group can only be due to their own actions, and not the actions of other agents.

Proposition 6. *Let coalition C be (weakly) q-responsible for S in $\mathcal{M}$ and A be an evidence set. (1) If C is not (weakly) q-responsible for S in $\mathcal{M}|_A$, then $Ag(A) \cap C \neq \emptyset$. (2) If $A' \subseteq A$ is a relevant evidence set for coalition C in $\mathcal{M}$, then $Ag(A') \subseteq C$.*

An implication of the above is that in order to hold a coalition responsible, one needs to collect evidence against at least one of the agents involved in the coalition. It should be noted that our aim was not to characterize the concept of "responsibility under evidence" in the proposed framework as it involves a meta-model update operation. Of course, our framework can be extended to make this possible, but we leave an elaboration on this concept for future work.

3.5 Reasoning about Responsibility

In the previous section we have introduced definitions of responsibility and the notions of crucial and necessary coalitions. How can we make use of these notations? Suppose that we have observed S–nothing more, nothing less–and would like to determine which coalition(s) is (are) responsible for S. As we have no more knowledge, we consider all states leading to S. These are all states in $X_S = \{q \mid \exists \alpha_N \in d_N(q) : o(q, \alpha_N) \in S\}$. We follow a conservative strategy and only consider a coalition (weakly) responsible for S if it is (weakly) responsible for any state in X_S. That is, a coalition is (weakly) responsible for S iff it is *(weakly) X_S-responsible for S* iff it is (weakly) q-responsible for S for all $q \in X_S$. Now, if there is a coalition C which is X-responsible for S we can say that C is responsible for S because at all states in X coalition C is the unique coalition that can prevent S. Similarly, we can interpret the notions of crucial, necessary and most responsible to sets of states X.

However, it might not be fair to assign responsibility to a coalition if the agents are not aware that they can prevent S. To model this, we have introduced incomplete information models and the concept of knowledge. Thus, we say that the coalition C *knows* that it is X-responsible for S if the coalition is responsible from all states it considers possible. For this purpose we replace "C is q-responsible" by "C knows it is q-responsible" etc. This means that responsibility is not only verified from X_S but from all states $\{q' \mid q \in X_S \text{ and } q \sim_C q'\}$ with the limitation that only uniform strategies are considered. Knowledge and responsibility are strongly interweaved and we would like to study the connection in more depth in our future research.

4 Logical Characterization

In this section we propose a logical characterization of responsibility. Therefore, we use a slight variant of the logic *coalition logic with quantification* (CLQ) proposed in [2, in the Proof of Theorem 8][3]. The logic is an extension of *coalition logic* [13,14] that allows to quantify over coalitions with specific properties. It is worth mentioning that the quantified versions are, in the finite case, not more expressive than coalition logic but often allow for an exponentially more succinct specification [2].

4.1 Preliminaries: Coalition Logic with Quantification

Formulae of *coalition logic* [13] (over $\mathcal{P}(N)$) are given by the following grammar: $\varphi ::= p \mid \neg\varphi \mid \varphi \vee \varphi \mid [A]\varphi$ where $A \in \mathcal{P}(N)$, $\mathsf{p} \in \Pi$. We define $\langle A\rangle\varphi$ as $\neg[A]\neg\varphi$ and Boolean connectives as usual. The semantics is defined over a CGS and a state q:

$\mathcal{M}, q \models [C]\varphi$ iff there is a joint action $\alpha_C \in d_C(q)$ such that for all joint actions $\alpha_{N\setminus C} \in d_{N\setminus C}(q)$ we have that $\mathcal{M}, o(q, (\alpha_C, \alpha_{N\setminus C})) \models \varphi$

Let $\mathcal{M}$ be a CGS and q a state in it. It is easy to verify that we have that: (i) C *can q-enforce* φ iff $\mathcal{M}, q \models [C]\varphi$; (ii) C *q-controls* φ iff $\mathcal{M}, q \models [C]\varphi \wedge [C]\neg\varphi$; and (iii) C *can q-avoid* φ iff $\mathcal{M}, q \models \langle N\setminus C\rangle\neg\varphi$.

We use an extension of coalition logic, introduced in [2], that allows to quantify over coalitions. Firstly, we introduce *coalitional predicates over* $\mathcal{P}(N)$: $P ::= \mathsf{sub}(\mathsf{C}) \mid \mathsf{super}(\mathsf{C}) \mid \neg P \mid P \vee P$ where $C \in \mathcal{P}(N)$ is a set of agents. The semantics of these predicates is defined over $A \subseteq N$ in a straight forward way: $A \models \mathsf{sub}(\mathsf{C})$ iff $A \subseteq C$ and $A \models \mathsf{super}(\mathsf{C})$ iff $A \supseteq C$. Negation and disjunction are treated as usual. We define equality between coalitions as macro: $\mathsf{eq}(\mathsf{C}) \equiv \mathsf{sub}(\mathsf{C}) \wedge \mathsf{super}(\mathsf{C})$. Note, we assume that the coalitional symbols C have their canonical semantic meaning–we do not discern between semantic and syntactic constructs in this paper.

Now let $\mathcal{V}$ be a set of coalitional variables. We define the logic *coalition logic with quantification*[4] (CLQ) [2, in the Proof of Theorem 8] as follows:

$$\varphi ::= \psi \mid \neg\varphi \mid \varphi \vee \varphi \mid \exists X|_P\, \varphi \mid \forall X|_P\, \varphi$$

where $X \in \mathcal{V}$, P is a coalitional predicate over $\mathcal{P}(N) \cup \mathcal{V}$, and ψ a coalition logic formula over $\mathcal{V}$. Moreover, we assume that all coalitional variables are bound. As for coalition logic the semantics is given over a CGS, a state in it, and a coalition valuation $\xi : \mathcal{V} \to \mathcal{P}(N)$. We define $\xi[X := C]$ as the valuation that equals ξ for all $Y \neq X$, i.e. $\xi[X := C](Y) = \xi(Y)$, and $\xi[X := C](X) = C$. We also define a special valuation ξ_0 with $\xi_0(X) = \emptyset$ for all $X \in \mathcal{V}$. We just give the semantics for the cooperation modality and the quantifiers, all other cases are standard:

$\mathcal{M}, q, \xi \models [X]\psi$ iff there is a joint action $\alpha_{\xi(X)} \in Act_{\xi(X)}$ such that for all joint actions $\alpha_{N\setminus\xi(X)} \in Act_{N\setminus\xi(X)}$ we have that $\mathcal{M}, o(q, (\alpha_{\xi(X)}, \alpha_{N\setminus\xi(X)})), \xi \models \varphi$

[3] Note, that CLQ is different from the better known logic Quantified Coalition Logic (QCL) also presented in [2].

[4] We would like to note that in comparison to [2] our definition of CLQ is somewhat more general as we allow coalitional variables within coalitional predicates.

$\mathcal{M}, q, \xi \models \exists X|_P \psi$ iff there is $C \subseteq N$ such that $C, \xi[X := C] \models P$ and $\mathcal{M}, q, \xi[X := C] \models \psi$

$\mathcal{M}, q, \xi \models \forall X|_P \psi$ iff for all $C \subseteq N$ with $C, \xi[X := C] \models P$ we have that $\mathcal{M}, q, \xi[X := C] \models \psi$]

where $C, \xi \models P$ is defined as $C \models P[\xi]$ and $P[\xi]$ is obtained from P where each coalitional variable Y occurring in P is replaced by $\xi(Y)$. We simply write $\mathcal{M}, q \models \varphi$ if φ is a closed formula. For a set of states $Q' \subseteq Q$ we write $\mathcal{M}, Q', \xi \models \varphi$ iff for all $q \in Q'$ we have $\mathcal{M}, q, \xi \models \varphi$. For a closed formula φ we define $[\![\varphi]\!]_{\mathcal{M}} = \{q \in Q \mid \mathcal{M}, q \models \varphi\}$. In [2] it was shown that model checking Quantified Coalition logic (QCL) is **PSPACE**-complete over compact models and can be done in polynomial time over an explicit representation based on effectivity functions. These results do not straightforwardly transfer to our setting as we use (i) a different representation of models, and (ii) a slightly generalized version of CLQ–which is somewhat different from QCL. A detailed study in our setting is out of the scope of this paper and we leave it for future work.

4.2 Logical Characterization of Responsibility

Given a closed formula φ, $[\![\varphi]\!]_{\mathcal{M}}$ associates to it the set of states in which φ holds. Moreover, instead of writing "C is q-responsible for $[\![\varphi]\!]_{\mathcal{M}}$ in $\mathcal{M}$", we simply say "C is q-responsible for φ (in $\mathcal{M}$)" etc. In the following we show that our notions of responsibility can be formalized within *coalition logic with quantification*. We assume that $\mathcal{M}$ is a CGS, q a state in it, C a coalition and φ a closed formula. Again, we omit mentioning $\mathcal{M}$ whenever clear from context. Firstly, we define the following two formulae:

$$\mathsf{R}^{\mathsf{s}}_{\mathsf{C}}\varphi \equiv \exists X|_{\mathsf{eq(C)}}([X]\neg\varphi \wedge \forall Y|_{\neg\mathsf{super(X)}}\neg[Y]\neg\varphi)$$
$$\mathsf{R}^{\mathsf{w}}_{\mathsf{C}}\varphi \equiv \exists X|_{\mathsf{eq(C)}}([X]\neg\varphi \wedge \forall Y|_{\neg\mathsf{eq(X)}\wedge\mathsf{sub(X)}}\neg[Y]\neg\varphi)$$

Proposition 7. *C is q-responsible (resp. weakly q-responsible) for φ in $\mathcal{M}$ iff $\mathcal{M}, q \models \mathsf{R}^{\mathsf{s}}_{\mathsf{C}}\varphi$ (resp. φ iff $\mathcal{M}, q \models \mathsf{R}^{\mathsf{w}}_{\mathsf{C}}\varphi$).*

Proof (Sketch). We only show the case for weak responsibility. We have that $\mathcal{M}, q, \xi \models \mathsf{R}^{\mathsf{w}}_{\mathsf{C}}\varphi$ iff $\mathcal{M}, q, \xi \models \exists X|_{\mathsf{eq(C)}}([X]\neg\varphi \wedge \forall Y|_{\neg\mathsf{eq(X)}\wedge\mathsf{sub(X)}}\neg[Y]\neg\varphi)$ iff $\mathcal{M}, q, \xi[X := C] \models [X]\neg\varphi$ and for all C' with $C' \neq C$ and $C' \subseteq C$ we have $\mathcal{M}, q, \xi[X := C, Y := C'] \not\models [Y]\neg\varphi$ iff C can q-enforce $\neg\varphi$ and all $C' \subset C$ cannot q-enforce $\neg\varphi$ iff C can q-enforce $\neg\varphi$ and there is no subcoalition of C that can q-enforce $\neg\varphi$ iff C is a minimal coalition that can q-enforce $\neg\varphi$ iff C is weakly q-responsible for φ. □

Now, we can simply express properties as given in Proposition 2 by $\models \mathsf{R}^{\mathsf{s}}_{\mathsf{C}}\varphi \rightarrow \mathsf{R}^{\mathsf{w}}_{\mathsf{C}}\varphi$. We can also easily define crucial coalitions, necessary coalitions, and the most responsible coalition:

$$\mathsf{Crucial}_{\hat{\mathsf{C}},\mathsf{C}}\varphi \equiv (\mathsf{R}^{\mathsf{s}}_{\mathsf{C}}\varphi \vee \mathsf{R}^{\mathsf{w}}_{\mathsf{C}}) \wedge \exists X_C|_{\mathsf{eq(C)}}\exists X_{\hat{C}}|_{\mathsf{eq}(\hat{\mathsf{C}})}\forall X|_{\neg\mathsf{super}(\mathsf{X}_{\hat{\mathsf{C}}})}\neg\mathsf{R}^{\mathsf{w}}_{(\mathsf{X}_{\mathsf{C}}\setminus\mathsf{X}_{\hat{\mathsf{C}}})\cup\mathsf{X}}\varphi$$
$$\mathsf{Nec}_{\hat{\mathsf{C}},\mathsf{C}}\varphi \equiv (\mathsf{R}^{\mathsf{s}}_{\mathsf{C}}\varphi \vee \mathsf{R}^{\mathsf{w}}_{\mathsf{C}}) \wedge \exists X_C|_{\mathsf{eq(C)}}\exists X_{\hat{C}}|_{\mathsf{eq}(\hat{\mathsf{C}})}\forall X|_{\neg\mathsf{super}(\mathsf{X}_{\hat{\mathsf{C}}})}\exists Y|_{\mathsf{eq}((\mathsf{X}_{\mathsf{C}}\setminus\mathsf{X}_{\hat{\mathsf{C}}})\cup\mathsf{X})}\neg[Y]\neg\varphi$$
$$\mathsf{Most}_{\mathsf{C}^u}\varphi \equiv \exists X|_{\mathsf{sub}(\mathsf{N})}(\mathsf{Nec}_{\mathsf{C}^u,\mathsf{X}}\varphi \wedge \forall Y|_{\neg\mathsf{sub}(\mathsf{C}^u)}\neg\mathsf{Nec}_{\mathsf{Y},\mathsf{X}}\varphi)$$

Proposition 8. *We have that $\hat{C}$ is q-crucial (resp. q-necessary and most q-responsible) for φ in the (weakly) q-responsible coalition C for φ iff $\mathcal{M}, q \models \mathsf{Crucial}_{\hat{\mathsf{C}},\mathsf{C}}\varphi$ (resp. $\mathcal{M}, q \models \mathsf{Nec}_{\hat{\mathsf{C}},\mathsf{C}}\varphi$ and $\mathcal{M}, q \models \mathsf{Most}_{\mathsf{C}}\varphi$).*

Proof (Sketch). **Cruciality:** We have that $\mathcal{M}, q, \xi \models \mathsf{Crucial}_{\hat{\mathsf{C}},\mathsf{C}}\varphi$ iff C is q-responsible for φ or weakly q-responsible for φ (by Proposition 7) and for all $C' \not\supseteq \hat{C}$ we have that $(C \backslash \hat{C}) \cup C'$ is not weakly q-responsible for φ iff C is q-responsible for φ or weakly q-responsible for φ and for all $C' \subseteq N$ such that if $(C \backslash \hat{C}) \cup C'$ is weakly q-responsible for φ then $\hat{C} \subseteq C'$ iff $\hat{C}$ is q-crucial for φ in C.

Necessary: We have that $\mathcal{M}, q, \xi \models \mathsf{Nec}_{\hat{\mathsf{C}},\mathsf{C}}\varphi$ iff C is q-responsible for φ or weakly q-responsible for φ (by Proposition 7) and for all $C' \not\supseteq \hat{C}$ we have that $(C \backslash \hat{C}) \cup C'$ cannot q enforce $\neg\varphi$ iff C is q-responsible for φ or weakly q-responsible for φ and for all $C' \subseteq N$ such that if $(C \backslash \hat{C}) \cup C'$ can q-enforce $\neg\varphi$ then $\hat{C} \subseteq C'$ iff $\hat{C}$ is q-necessary for φ in C.

Most responsible: Now, let us consider the most responsible coalition. $\mathcal{M}, q, \xi \models \mathsf{Most}_{\mathsf{C}^u}\varphi$ iff there is $C \subseteq N$ such that C^u is q-necessary for φ in C and for all $C' \not\subseteq C^u$ we have that C' is not q-necessary for φ in C iff there is $C \subseteq N$ such that C^u is the maximal coalition that is q-necessary for φ in C iff C^u is the most q-responsible coalition for φ by Theorem 1 and Definition 5. □

The logical formulation shows that our notions of responsibility are fully based on strategic ability; there are no other hidden concepts. Also, it provides a first step to reasoning about group responsibility. We leave a detailed study for future work, including a deeper analysis of epistemic concepts and meta logical properties.

5 Related Work

Existing work on formalizing responsibility can be categorized in two approaches. The first type of work considers backward-looking responsibility formalized in dynamic logic while the second type of work considers forward-looking responsibility formalized in deontic and STIT logics. In the following, we provide an example of each approach.

Grossi et al. [8] investigate the concept of responsibility in an organizational setting where role playing agents operate within organizational structures defined by power, coordination and control relations. They distinguish causal and task-based responsibility, and investigate when agents in an organization can be held accountable for or be blamed for some (undesirable) state of affairs. For example, an agent A who delegates a task to an agent B using its organizational power can be held responsible for the failure of performing the task even though agent B has actually failed in performing the task. In order to formalize these notions of responsibility, they propose a dynamic deontic logic framework in which agents' activities as well as the organizational setting are specified. In this framework, an agent is defined to be causally responsible for ϕ by performing action α if and only if ϕ is the necessary effect of α and moreover ϕ would not have been the case if α was not performed. An agent is then said to be causally blameworthy (backward-looking responsible) if the agent is casually responsible for a violation and

the agent knows that his/her action would cause the violation which he could prevent by not performing the action. The task-based responsibility (forward-looking responsibility) is defined in terms of organizational tasks/plans that the agent is obliged to perform. Finally, an agent is said to be accountable for a violation caused by performing an action α if the agent is blameworthy for performing α causing the violation and moreover if the agent is task-responsible to perform α.

The characterizing feature of this approach is the formalization of different notions of responsibility in the context of organizational structures. The formalization of causal and task-based responsibility are defined with respect to individual agents and based on violation of an agent's obligations.

Another formal framework for analyzing the concept of (forward-looking) responsibility is proposed by Mastop [11]. The main focus of this work is the normative dimension of the "many hands problem", which is formulated as the problem of attributing the responsibility for the violation of a global norm to individual agents. This is a challenging problem because agents may not be responsible for the violation of a global norm even when they clearly violate their individual norms and thereby cause the violation of the global norm. The problem with attributing responsibility in such cases is argued to be the fairness issue. The fairness considered in this work is explained in terms of conditions such as "agents should be able to obey their individual norms", "agents should be aware of their individual norms", or "the violation of individual norms should be intentional and caused by some accidents". The framework is based on an extension of the XSTIT logic with intentions [5]. This logic is extended with, among other things, a set of designated constants denoting the responsibility of agents. The semantics of the extended XSTIT framework explicitly attributes to each agent a set of possibilities (history-state pairs) in which the agent fulfills all of its responsibilities (i.e., possibilities in which the agent's designated responsibility constant is true). An agent is defined to be responsible for ϕ if and only if the set of possibilities attributed to the agent satisfy ϕ. Based on this definition of forward-looking responsibility, the fairness conditions are formulated as axiom schemes. The author claims that the introduction of these axioms ensures that the responsibility of any violation of global norms can be attributed to some individuals that violate their individual norms.

In another work [7], Ciuni and Mastop, the XSTIT is used and extended to analyze the concept of distributed responsibility, i.e., the responsibility that is attributed to a group of agents. The basic problem considered in this work is to distinguish the responsibilities of individuals within a group to which a group responsibility is attributed. For example, if a group of two agents is responsible for $\phi \wedge \psi$, the presented framework can distinguish whether one of the agents or both are responsible for this composite fact. The characterizing feature of this approach is the explicit introduction of responsibility constants as well as their corresponding semantics counterparts, i.e., the sets of possibilities in which agents' responsibilities are fulfilled. In fact, the proposed framework is assumed to be informed about agents' responsibilities such that the main contribution of this paper is not to define the concept of responsibility itself, but the formalization of the fairness conditions in order to solve the "many hands problem".

We would like to mention three other papers in which responsibility is related to other notions such as emotions, causality, and morality. In the first paper [9], the authors use

a STIT logic for counterfactual reasoning about emotions. Their characterizations are based on the group's potential power to *could have prevented* some state-of-affairs. In contrast to their setting, however, we do not assume that the state-of-affairs has actually been materialized as we do not model backward-looking responsibility. Moreover, our focus is on the inherent structure of the coalitions at hand rather than on their pure power to prevent some states of affairs. In [6] the authors argue that causality has mostly been studied as an all-or-nothing concept and propose an extension to capture the degree of responsibility and blame in causal relations. The proposed extension allows one to express that a phenomenon A is responsible (or blameworthy) for causing a phenomenon B to a certain degree depending on A's contribution in causing B. The contribution of A in causing B (the degree of A's responsibility) is determined based on some counterfactual reasoning and other factors relevant to B being caused. An obvious difference with our work is the central role of causality and the fact that responsibility and blame are directly defined in terms of causality, i.e., A is responsible for B iff A has caused B. In our work, however, a group of agents is responsible for some states, not because they have caused the states (as the states are not assumed to be materialized), but because they have the power to preclude the states. In our framework, it can even be the case that a group of agents has the power to ensure a certain outcome while a different group of agents is responsible for the outcome. In the last paper [4], the authors focus on moral responsibility and provide a set of conditions that are claimed to be necessary and sufficient for assigning moral responsibility for a certain outcome to individuals. These conditions require that an individual can be held responsible for an outcome if the individual is autonomous, has causal contribution to the outcome, and has the opportunity to avoid the outcome. Again, in contrast to our framework, this paper assumes that an outcome is materialized and that the responsible individual has causally contributed to the materialization of the outcome. Moreover, although the last condition seems to be related to our notion of precluding power for avoiding the outcome, it requires that the individual who has causally contributed to the outcome should have the power to avoid the outcome. This is different from our approach where responsible coalitions for an outcome may be unrelated to the coalition who can ensure the outcome. Finally, this paper considers only the responsibility of individuals as it aims to tackle 'the problem of many hands' while we investigate the responsibility of coalitions.

6 Conclusion and Future Work

In this paper, we provided an abstract notion of group responsibility that does not imply accountability or blameworthiness. The proposed notion of responsibility is based on the preclusive power of groups of agents and is defined as the responsibility to prevent some state of affairs. We have formalized this notion of responsibility in concurrent game structures, which model multi-agent system behaviors. Different notions of responsibility such as (weak) responsibility, crucial and necessary responsibility are formally defined and their properties are analyzed. We then presented the notion of "responsibility under evidence" according to which a group of agents can be held responsible for a state of affairs if there is evidence that they did not act to prevent the state of affairs. In this sense, it can be said that the agents have acted irresponsibly or

incautiously (as they did not act to prevent the state of affairs) even if their performed actions do not cause the realization of the state of affairs that have to be prevented. The main results of this paper are formulated in the characterization theorems. Finally, we show how our notions of responsibility can be characterized as formulas of coalition logic with quantification [2].

We plan to extend this framework with different levels of agents' knowledge and intention in order to distinguish different levels of responsibilities. Such extension would also allow us to instantiate the presented abstract notion of responsibility to capture different types of responsibilities, for example accountability and blameworthiness. In such extensions, one would be able to determine if a group of agents is accountable or blameworthy for some state of affairs. We also aim at generalizing the notion of responsibility to a strategic setting as the current notion of responsibility is relativized to a specific state q such that it can only be expressed that a group of agents is q-responsible. Based on such an extension and given a realized state of affair, one would be able to reason about which agents at which states were responsible for the realization of the state of affairs. We aim at extending the framework such that group responsibility can be distributed to individual agents and elaborating on the logical characterization. Finally, it would be interesting to relate the concepts of crucial and necessary coalitions to the concept of power as discussed in [1].

References

1. Ågotnes, T., van der Hoek, W., Tennenholtz, M., Wooldridge, M.: Power in normative systems. In: AAMAS 2009: Proceedings of The 8th International Conference on Autonomous Agents and Multiagent Systems, Richland, SC, pp. 145–152. International Foundation for Autonomous Agents and Multiagent Systems (2009)
2. Ågotnes, T., van der Hoek, W., Wooldridge, M.: Quantified coalition logic. Synthese 165(2), 269–294 (2008)
3. Alur, R., Henzinger, T.A., Kupferman, O.: Alternating-time Temporal Logic. Journal of the ACM 49, 672–713 (2002)
4. Braham, M., van Hees, M.: An anatomy of moral responsibility. Mind 121(483), 601–634 (2012)
5. Broersen, J.: Deontic epistemic stit logic distinguishing modes of mens rea. Journal of Applied Logic 9(2), 137–152 (2011)
6. Chockler, H., Halpern, J.Y.: Responsibility and blame: A structural-model approach. Journal of Artificial Intelligence Research 22, 93–115 (2004)
7. Ciuni, R., Mastop, R.: Attributing distributed responsibility in stit logic. In: He, X., Horty, J., Pacuit, E. (eds.) LORI 2009. LNCS, vol. 5834, pp. 66–75. Springer, Heidelberg (2009)
8. Grossi, D., Royakkers, L.M.M., Dignum, F.: Organizational structure and responsibility. Artificial Intelligence and Law 15(3), 223–249 (2007)
9. Lorini, E., Schwarzentruber, F.: A logic for reasoning about counterfactual emotions. In: IJCAI, pp. 867–872 (2009)
10. Lucas, J.: Responsibility. Oxford University Press (1993)
11. Mastop, R.: Characterising responsibility in organisational structures: The problem of many hands. In: Governatori, G., Sartor, G. (eds.) Deontic Logic in Computer Science, 10th International Conference (DEON 2010), pp. 274–287 (2010)

12. Miller, N.R.: Power in game forms. In: Holler, M.J. (ed.) Power, Voting, and Voting Power, pp. 33–51. Physica-Verlag, Wuerzberg-Vienna (1982); Reprinted in Homo Oeconomicus 15(2), 219–243 (1999)
13. Pauly, M.: Logic for Social Software. PhD thesis, University of Amsterdam (2001)
14. Pauly, M.: A modal logic for coalitional power in games. Journal of Logic and Computation 12(1), 149–166 (2002)
15. van der Poel, I.: The relation between forward-looking and backward-looking responsibility. In: Vincent, N., van de Poel, I., van den Hoven, J. (eds.) Moral Responsibility: Beyond Free Will and Determinism, pp. 37–52. Springer (2011)
16. Vincent, N., van de Poel, I., van den Hoven, J. (eds.): Moral Responsibility: beyond free will and determinism. Library of Ethics and Applied Philosophy. Springer (2011)