

Modeling the problem of many hands in organisations

Tiago de Lima¹ and Lambèr Royakkers² and Frank Dignum³

Abstract. In this paper we provide a formalism to reason about the problem of many hands in organisations. This is a problem that arises whenever the organisation is responsible for some undesirable outcome but none of its members can be held responsible for the outcome. The formalism proposed here is a logic that extends the Coalition Epistemic Dynamic Logic by adding a notion of group knowledge and also organisational structures to its semantics. An organisational structure is a set of agents and some relations between them. It defines the power and coordination links between agents, which defines how agents delegate tasks and communicate. We give formal definitions for individual and collective responsibility, as well as for the problem of the many hands.

Keywords: multi-agent systems, organised agency, responsibility, problem of many hands, coalition epistemic dynamic logic.

1 INTRODUCTION

The ‘problem of many hands’ (PMH) is a term coined by Thompson in [17], and studied in the context of organisations (e.g., by Bovens [4]). It is a problem that sometimes arises when we address collective responsibility: in some situations, a group of individuals can reasonably be held responsible for an undesirable outcome, while no member of the group can reasonably be held responsible for the outcome. Bovens argues that such situations are problematic because it “frustrates the need for compensation and retribution on the part of the victims” and, maybe more important, because “the fact that no one can be meaningfully called to account after the event also means ... that no one need feel responsible beforehand” [4].

Our aim in this article is to provide a formal tool that can help on detecting the possible occurrence of the PMH. For that, we build up a logic for reasoning about collective and individual responsibility. This logic extends the Coalition Epistemic Dynamic Logic (CEDL) proposed in [8]. We add a notion of group knowledge, generalise the definitions of individual responsibility to groups of agents, and give a formal definition for the PMH.

The PMH has primarily been studied in the context of organised agency. That is, the individuals involved are supposed to know their roles and be able to cooperate. Therefore, to address the PMH, we here focus the attention on situations where the set of individuals involved forms an ‘organisation’. To this end, we extend CEDL again by adding organisational structures to its semantics. These structures define the power and coordination links between agents, which define how agents delegate tasks and communicate in the context of the organisation.

The remainder of this article is organised as follows. Section 2 recalls CEDL and extends it with operators for group knowledge.

It also defines ability, obligations, forward-looking and backward-looking responsibility, generalising all of them to groups of agents. Section 3 presents the PMH in organisations, defines organisational structures and organisational actions, and presents an example of the application of the logic. Section 4 compares with related work and Section 5 concludes.

2 COALITION EPISTEMIC DYNAMIC LOGIC

CEDL has been proposed in [8]. It is a logic for reasoning about agency, such as Coalition Logic [14] and ATL [1]. But differently from those, it also enables reasoning about actions and knowledge. These features are useful to reason about responsibility and thus, to reason about the PMH.

2.1 Models

Assume a countable set Φ of atomic formulae, a finite set N of labels denoting agents, and a finite set A of atomic actions of the language. We use Δ to denote the set of all joint actions available for the agents, i.e., the set of all total functions δ with signature $N \rightarrow A$. And we use Γ to denote the set of all partial joint actions available for the agents, i.e., the set of all partial functions $\delta_G = \{(i, a) : (i, a) \in \delta \text{ and } i \in G\}$, where $\delta \in \Delta$ and $G \subseteq N$. CEDL models are quadruples $\langle W, \mathcal{K}, \mathcal{T}, \pi \rangle$, where:

- W is a non-empty set of states (or possible worlds);
- $\mathcal{K}$ is a function with signature $N \rightarrow (W \times W)$, defining, for each agent $i \in N$, an equivalence relation between possible worlds, representing i ’s knowledge;
- $\mathcal{T}$ is a function with signature $\Gamma \rightarrow (W \times W)$, defining, for each partial joint action δ_G , a relation between possible worlds, representing the transitions associated to the partial joint action δ_G ; and
- π is a function with signature $\Phi \rightarrow 2^W$, defining, for each $p \in \Phi$, the interpretation of the atomic formula p in the model.

To simplify notation we use $\mathcal{K}(i)(w)$ to denote the set of worlds that i considers possible at w . That is, $\mathcal{K}(i)(w)$ abbreviates the set $\{w' \mid \langle w, w' \rangle \in \mathcal{K}(i)\}$. Analogously, we use $\mathcal{T}(\delta_G)(w)$ to denote the set of possible outcomes of the occurrence of δ_G in w . That is, $\mathcal{T}(\delta_G)(w)$ abbreviates the set $\{w' \mid \langle w, w' \rangle \in \mathcal{T}(\delta_G)\}$.

CEDL models also satisfy the following five semantic constraints:

- (SC1) $\bigcup_{\delta \in \Delta} \mathcal{T}(\delta)(w) \neq \emptyset$.
- (SC2) $\mathcal{T}(\delta_{G_1})(w) \subseteq \bigcup_{\delta' \in \Delta} \mathcal{T}(\delta'_{G_1 \cup G_2})(w)$, where $\delta_{G_1} = \delta'_{G_1}$.
- (SC3) $\mathcal{T}(\delta_{G_1 \cup G_2})(w) \subseteq \mathcal{T}(\delta_{G_1})(w) \cap \mathcal{T}(\delta_{G_2})(w)$.
- (SC4) If $\mathcal{T}(\delta_N)(w) \neq \emptyset$ then $\mathcal{T}(\delta_N)(w)$ is a singleton.
- (SC5) $(\mathcal{T}(\delta_N) \circ \mathcal{K}(i))(w) \subseteq (\mathcal{K}(i) \circ \mathcal{T}(\delta_N))(w)$.

¹ CRIL, University of Artois, France

² IE&IS, Eindhoven University of Technology, The Netherlands

³ ICS, Utrecht University, The Netherlands

Constraint SC1 stipulates that there is at least one executable action for each agent. Constraints SC2 and SC3 stipulate how the outcome of actions performed simultaneously is calculated. SC2 stipulates that if a group of agents G_1 necessarily achieves a certain outcome by acting as determined by the joint action δ , then any supergroup of G_1 will possibly achieve the same outcome by acting as determined by another joint action δ' , provided that the agents in G_1 still act as determined by δ . SC3 is called ‘superadditivity’. It stipulates that if a group of agents $G_1 \cup G_2$ can possibly achieve a certain outcome by acting as determined by the joint action δ , then each of its subgroups can possibly achieve the same outcome, provided that they act as determined by the same joint action δ . In particular, this implies that the bigger the group the more deterministic the outcome is. SC4 is called ‘joint determinism’. It stipulates that, at any moment, the subsequent state is completely determined by the joint action that the entire set of agents decides to execute. (Recall that $\delta_N = \delta$.) And finally, SC5, sometimes called ‘no forgetting’ (or ‘perfect recall’ in [10]), defines the interaction between the knowledge relation and state transitions. The knowledge of an agent either increases or remains the same after the execution of an action. This means that agents never lose information, i.e., once an agent knows something, this agent will never forget it. Together with the fact that each $\mathcal{K}(i)$ is an equivalence relation, SC5 implies that action occurrences are perceived by all agents, which implies that the agents perceive the passage of time.

2.2 Language

The language $\mathcal{L}$ of CEDL is the set of formulae φ defined by the following grammar:

$$\varphi ::= \top \mid p \mid \neg\varphi \mid \varphi \vee \varphi \mid K_i\varphi \mid [\delta_G]\varphi$$

where p ranges over Φ , i ranges over N , and δ_G ranges over Γ .

Formula $K_i\varphi$ means ‘agent i knows that φ ’. Joint action $\{(i_1, a_1), \dots, (i_n, a_n)\}$ means ‘the agents in $\{i_1, \dots, i_n\}$ execute their corresponding actions in $\{a_1, \dots, a_n\}$ simultaneously (and we do not consider what the other agents are doing at the same time)’. And formula $[\delta_G]\varphi$ means ‘after every possible occurrence of δ_G , it is the case that φ ’.

In what follows, the common abbreviations for $\wedge$, $\rightarrow$, $\leftrightarrow$ and $\perp$ are also used. As Δ , N and 2^N are finite sets, existential and universal quantification over these sets just abbreviate a big disjunction and big conjunction, respectively.

Let a model $M = \langle W, \mathcal{K}, \mathcal{T}, \pi \rangle$ and a world $w \in W$ be given. The satisfaction relation $\models$ between pairs $\langle M, w \rangle$ (called pointed models) and formulae in $\mathcal{L}$ is as usual for Boolean operators, plus:

$$\begin{aligned} M, w \models K_i(\varphi) & \quad \text{iff} \quad M, w' \models \varphi \text{ for all } w' \in \mathcal{K}(i)(w) \\ M, w \models [\delta_G]\varphi & \quad \text{iff} \quad M, w' \models \varphi \text{ for all } w' \in \mathcal{T}(\delta_G)(w) \end{aligned}$$

As usual, a formula $\varphi \in \mathcal{L}$ is valid (notation: $\models \varphi$) if and only if every pointed model $\langle M, w \rangle$ satisfies φ .

The principles below constitute the axiomatic system of CEDL:

- (A1) All tautologies from classical propositional logic
- (A2) All S5 principles for each operator K_i
- (A3) All K principles for each operator $[\delta_G]$
- (A4) $\exists \delta \neg[\delta_i] \neg \top$
- (A5) $(\forall \delta' [\delta'_{G_1 \cup G_2}]\varphi) \rightarrow [\delta_{G_1}]\varphi$ (if $\delta'_{G_1} = \delta_{G_1}$)
- (A6) $([\delta_{G_1}]\varphi_1 \wedge [\delta_{G_2}]\varphi_2) \rightarrow [\delta_{G_1 \cup G_2}](\varphi_1 \wedge \varphi_2)$ (G_1, G_2 disjoint)
- (A7) $\neg[\delta_N]\varphi \rightarrow [\delta_N]\neg\varphi$

$$(A8) \quad K_i[\delta_N]\varphi \rightarrow [\delta_N]K_i\varphi$$

Principles A4–A8 correspond to constraints SC1–SC5, respectively.

Theorem 1 ([8]) *Principles A1–A8 are sound and complete with respect to CEDL models.*

2.3 Group knowledge

Since Aristotle [3], knowledge is considered to be an important component of the definition of responsibility. Because we aim at formalising collective responsibility, we first have to provide a formalisation of ‘group knowledge’. Unfortunately, there is no consensus in the literature of what group knowledge means (see, e.g., the discussion in [12]). Here, we choose to use the notion of ‘distributed knowledge’, found, e.g., in [10]. In the language, we replace operators K_i by the more general K_G , which semantics is formally defined as follows. Let $G \neq \emptyset$:

$$M, w \models K_G\varphi \quad \text{iff} \quad M, w' \models \varphi \text{ for all } w' \in \bigcap_{i \in G} \mathcal{K}(i)(w)$$

Distributed knowledge approximately describes the knowledge of someone who has complete knowledge of what each member of the community knows.

Proposition 2 *The following schemata are valid in CEDL with distributed knowledge:*

- (A2') All S5 principles for each operator K_G
- (A8') $K_G[\delta_N]\varphi \rightarrow [\delta_N]K_G\varphi$
- (A9) $(K_{G_1}\varphi_1 \wedge K_{G_2}\varphi_2) \rightarrow K_{G_1 \cup G_2}(\varphi_1 \wedge \varphi_2)$

Proof sketch. A2' is easy, and left for the reader. A8' is true because SC5 is preserved for groups G by the definition of K_G . And A9 is true because $\bigcap_{i \in G_1 \cup G_2} \mathcal{K}(i) \subseteq \bigcap_{i \in G_1} \mathcal{K}(i)$. $\square$

Note that with A9 and Axiom K, one can derive both: $K_{G_1}\varphi \rightarrow K_{G_1 \cup G_2}\varphi$ and $(K_{G_1}\varphi_1 \wedge K_{G_2}(\varphi_1 \rightarrow \varphi_2)) \rightarrow K_{G_1 \cup G_2}\varphi_2$. However, whether principles A1–A9 constitute a complete axiomatic system for CEDL with distributed knowledge, is left as an open question.

2.4 Ability

Another important component of the definition of responsibility is ability. A group is able to bring about a certain outcome whenever the group has the possibility to ensure the outcome. To define ability, we first define ‘ensuring’. Therefore, we introduce operator E in our logic. The formula $E_{\delta_G}\varphi$ means ‘by executing δ , group G ensures that φ ’. It is formally defined as follows:

$$E_{\delta_G}(\varphi) \stackrel{\text{def}}{=} \neg[\delta_G]\neg\top \wedge [\delta_G]\varphi$$

That is, ‘ensuring’ means that the action performed by G is executable and its execution invariably leads to a state where φ is true.

Several studies (e.g., [18], [5] and [11]) suggested that the meaning of the sentence ‘ G has the ability to bring about φ ’ should take agents’ knowledge into account. More precisely, ability to bring about φ should imply that agents ‘know how to act’ in order to bring it about. This suggests that we should define ability as follows:

$$A_G\varphi \stackrel{\text{def}}{=} \exists \delta (K_GE_{\delta_G}\varphi)$$

Formula $A_G\varphi$ means ‘group G is able to bring about φ in one step’.

Proposition 3 *Such definition of ability satisfies the properties:*

1. $\neg A_G \perp$
2. $(A_{G_1} \phi_1 \wedge A_{G_2} \phi_2) \rightarrow A_{G_1 \cup G_2} (\phi_1 \wedge \phi_2)$
3. $\models \phi_1 \rightarrow \phi_2 \text{ implies } \models A_G \phi_1 \rightarrow A_G \phi_2$

Proof sketch. The proof of items 1 and 3 are easy. The proof of item 2 first uses A5 to show that if there exists δ^1 and δ^2 such that $(\neg[\delta_{G_1}^1] \perp \wedge \neg[\delta_{G_2}^2] \perp)$ then there exists δ such that $\neg[\delta_{G_1 \cup G_2}] \perp$. Then, it uses A6 and A9 to derive the conclusion. $\square$

The operators that formalise agents' ability in Coalition Logic [14] and ATL [1] satisfy principles that are similar to the ones above. Maybe, the most important of them is superadditivity (Proposition 3.2). We remark however, that our operator A takes, not only agents' knowledge, but also time into account. Operator A_G can be seen as a fusion of ATL's $\langle\langle G \rangle\rangle$ and X (where the latter is the temporal operator 'next'). That is, formula $A_G \phi$ corresponds to ATL's $\langle\langle G \rangle\rangle X \phi$ and not to $\langle\langle G \rangle\rangle \phi$. Moreover, our choice of taking agents' knowledge into account makes our definition of ability differ in some other aspects. For instance, $\langle\langle G \rangle\rangle X \top$ is valid in ATL, while $A_G \top$ is not valid in CEDL. This is so because we do not assume that agents know which actions are executable in every state of the world.

Another interesting difference concerns the dual of A_N . In ATL, $\neg\langle\langle N \rangle\rangle X \neg \phi$ is equivalent to $\langle\langle \emptyset \rangle\rangle X \phi$, which means ' ϕ is unavoidable', while in CEDL, $\neg A_N \neg \phi$ means that 'it is not the case that the whole group of agents is able to bring about $\neg \phi$ ', which has the precise meaning 'for every action δ the whole group of agents considers it possible that executing δ does not ensure $\neg \phi$ '. Note that the latter does not imply that ϕ is unavoidable. Or, in other words, the impossibility to know for sure if we will achieve $\neg \phi$ does not imply the impossibility to achieve $\neg \phi$. We think that this is correct when considering agents' knowledge in the definition of ability.

Finally, note that ' ϕ is unavoidable' can be expressed in CEDL by a simple abbreviation. For it actually means that ϕ will be the case no matter what agents do. Therefore, we define it as:

$$X\phi \stackrel{\text{def}}{=} \forall \delta [\delta_N] \phi$$

We prefer to read $X\phi$ as ' ϕ is the case in the next state', or ' ϕ is the case after one step'. Note that $X\phi \rightarrow \neg A_G \neg \phi$ is valid, while $\neg A_G \neg \phi \rightarrow X\phi$ is not, and that $\neg A_N \neg \phi \rightarrow X\phi$ is not valid either, as explained above.

2.5 Obligations

We will also use obligations in the definition of responsibility. Since CEDL is a kind of dynamic logic, obligations can be defined adapting D'Altan et al.'s method [7], which is an implementation of Anderson's reduction of deontic logic to alethic modal logic [2].

First, we add to the set Φ of atomic formulae of the language the set $\{V_G : G \subseteq N\}$. The idea is that the atomic formula V_G flags the states of the model that contains a violation for G . In other words, V_G simply means ' G is in violation'. Obligations are then defined as follows:⁴

$$O_G \phi \stackrel{\text{def}}{=} (\neg \phi \rightarrow V_G)$$

A formula of the form $O_G \phi$ is read as 'it is obligatory for G that ϕ '. Defined as such, obligations satisfy some well-known properties:

⁴ Our definition has been inspired by D'Altan et al.'s work [7], but it is not exactly the same. D'Altan et al.'s obligation in our framework would rather be given by $XO_G \phi$, which we read as ' ϕ is obligatory for G in the next step'.

Proposition 4

1. $\models O_G \top$
2. $\models O_G (\phi_1 \wedge \phi_2) \leftrightarrow (O_G \phi_1 \wedge O_G \phi_2)$
3. $\models \phi \text{ implies } \models O_G \phi$

Note that we do not impose axiom D, i.e., $O_G \perp$ is satisfiable in CEDL with obligations. In our view it just means that dilemmas are possible. As in the classical Sartre's example, one can have the obligation to stay at home to look after an elderly mother and, at the same time, the obligation to join the resistance movement to fight the Nazis.

2.6 Responsibility

With all the ingredients in hand, we are now able to generalise the definitions of forward-looking and backward-looking responsibility given in [8] to groups of agents. The first one is forward-looking responsibility.

Definition 5 (We only define the first three here, R_G^3 , R_G^4 , etc., are analogous.)

$$\begin{aligned} R_G^0(\phi) &\stackrel{\text{def}}{=} O_G(\phi) \\ R_G^1(\phi) &\stackrel{\text{def}}{=} O_G(A_G \phi) \wedge X(R_G^0(\phi)) \\ R_G^2(\phi) &\stackrel{\text{def}}{=} O_G(A_G A_G \phi) \wedge X(R_G^1(\phi)) \end{aligned}$$

The formula $R_G^n(\phi)$ means ' G is forward-looking responsible for ϕ after n steps'. It approximates the idea that, different from obligations, ascribing responsibilities implies more than simply stating what are the undesirable situations. It also ascribes what are the undesirable behaviours. When ascribing forward-looking responsibility for ϕ to G , we set as violation all the states of the model from where G is not able to bring about ϕ any more.

The definition of backward-looking responsibility is given in two parts, which reflects the two questions we normally ask when trying to determine if someone is backward-looking responsible for some outcome. Given that an outcome ϕ has been reached, the first question we ask is: 'Did group G do something that contributed to the outcome?'. This is the case if G executed an action that possibly would not avoid the outcome, and that G knew this possibility. Formally:

$$\begin{aligned} I_G(\delta) &\stackrel{\text{def}}{=} \neg K_G E_{\delta_G} \neg V_G \wedge A_G \neg V_G \\ I_G(\delta^1; \dots; \delta^n) &\stackrel{\text{def}}{=} I_G(\delta^1) \vee \neg K_G [\delta_G^1] \neg I_G(\delta^2; \dots; \delta^n) \end{aligned}$$

Formula $I_G(\delta)$ means 'the performance of δ is irresponsible for group G '. The second question we normally ask is: 'Was group G supposed to reach $\neg \phi$?'. This is so if G had a forward-looking responsibility that $\neg \phi$. Putting both together, backward-looking responsibility is defined as:

$$\begin{aligned} B_{\delta_G}(\phi) &\stackrel{\text{def}}{=} R_G^1(\neg \phi) \wedge I_G(\delta) \\ B_{\delta_G^1; \dots; \delta_G^n} \phi &\stackrel{\text{def}}{=} R_G^n(\neg \phi) \wedge I_G(\delta^1; \dots; \delta^n) \end{aligned}$$

The latter question is asked because, if its answer is no, then there is no special reason that G should care about bringing about ϕ or not. Then we cannot hold G responsible for ϕ , because G was not supposed to pay attention about this outcome.

3 THE PROBLEM OF MANY HANDS

3.1 How to avoid the PMH

As said before, the problem of many hands arises in an organisation when the organisation is backward-looking responsible for some undesirable outcome but no organisation member can be held backward-looking responsible for this outcome. Let G be the set of members of a given organisation, it is formally given by:

$$\text{PMH}_{\delta_G} \varphi \stackrel{\text{def}}{=} B_{\delta_G}(\varphi) \wedge \forall i \neg B_{\delta_i}(\varphi)$$

A formula of the form $\text{PMH}_{\delta_G} \varphi$ means that ‘by performing δ_G , the problem of many hands arises in group G with respect to the outcome φ ’. If we follow the definitions of operator B we find out that we can paraphrase it as ‘by performing δ_G , group G is backward-looking responsible for φ , but no agent i in G is backward-looking responsible for φ by performing δ_i ’. Analysing the definitions of operator B a bit further we realise that the PMH may arise from three different sources: either (1) no forward-looking responsibility is ascribed to the individuals, or (2) the individuals that are forward-looking responsible for φ do not have the ability to bring it about (without the other individuals of the group), or (3) the individuals that are forward-looking responsible for φ do not have the necessary knowledge to bring about φ (without considering the knowledge of the other individuals of the group).

To address source 1, once a group is forward-looking responsible for outcome φ , we need to ascribe responsibility to at least one individual in that group. We call this individual the leader of the group. Formally we have:

$$(A10) \quad R_G^n \varphi \rightarrow R_i^n \varphi \quad (\text{if } i \text{ is the leader of } G)$$

and where the leader i is a distinguished agent in the model (this is defined formally in the next section).

Now, to address source 2, we have to be sure that the forward-looking responsibility can be “delegated” (directly or indirectly) to the individuals that have the ability to fulfill it. This means that the leader should have the power to ascribe forward-looking responsibilities to other individuals of the group.

Similarly, to address source 3, we have to be sure that individuals can share information. In other words, that knowledge can be transmitted among individuals of the group.

Therefore, a group that wants to avoid the PMH should organise itself in such a way that delegation actions and information actions are possible. In management theory, some ideas of how organisations can be built are given. Here, we use the ideas of [13], and define organisations as a kind of structure.

3.2 Organisational structures

An organisational structure is a quadruple $\langle G, I, P, i \rangle$ where:

- G is a group of agents in N , representing the members of the organisation;
- P is a subset of $G \times G$, representing a power relation among the members of the organisation;
- I is a subset of $G \times G$, representing an information flux relation among the members of the organisation;
- i is an agent in G , representing the leader of the group,

and that satisfies the following semantic constraints:

$$(SC6) \quad j \in P^+(i) \quad (\text{for all } j \in G)$$

$$(SC7) \quad P \subseteq I$$

where P^+ is the transitive closure of P .

SC6 implies that every individual in the organisation can eventually be ascribed forward-looking responsibility for some outcome. SC7 guarantees that if an individual may be ascribed forward-looking responsibility for some outcome, this individual can also be informed about it.

Let O be an organisational structure, the models of our logic are now tuples of the form $\langle O, W, \mathcal{T}, \mathcal{K}, \pi \rangle$. And its language is extended by the set of atomic formulae $\{\text{Power}(i, j) \mid i, j \in G\} \cup \{\text{Coord}(i, j) \mid i, j \in G\}$. The satisfaction relation is the same as before plus:

$$\begin{aligned} M, w \models \text{Power}(i, j) & \quad \text{iff} \quad (i, j) \in P \\ M, w \models \text{Coord}(i, j) & \quad \text{iff} \quad (i, j) \in I \end{aligned}$$

Note that from SC7 we immediately obtain:

$$(A11) \quad \text{Power}(i, j) \rightarrow \text{Coord}(i, j)$$

3.3 Organisational actions

Once organisational structures are in place, we can define the organisational actions mentioned above. The first kind of such actions consists of information actions. That is, we add the set of actions $\{\text{info}(G, \varphi) : G \subseteq N \text{ and } \varphi \in \mathcal{L}\}$ to the set A of atomic actions of the language. For example, action $\{(i, \text{info}(G, \varphi))\}$ means ‘agent i informs all members of G that φ ’. We require that such actions satisfy the following two properties:

$$(A12) \quad \neg[(i, \text{info}(G, \varphi))] \neg \top \rightarrow K_i \varphi$$

$$(A13) \quad \text{Coord}(i, j) \rightarrow [(i, \text{info}(G, \varphi))] K_j \varphi \quad (\text{for each } j \in G)$$

A12 restricts the circumstances in which action $\text{info}(G, \varphi)$ is executable. We impose that agents can inform only what they know will be true after the communication act. It follows that agents cannot lie. A13 stipulates that informing actions are successful communication actions when the agents involved are appropriately related by a coordination link. This also means that the agents trust information that comes through the coordination link.

The second kind of organisational action consists of delegation actions. That is, we add the set of actions $\{\text{deleg}((G_1, n_1, \varphi_1), \dots, (G_m, n_m, \varphi_m)) : G_1, \dots, G_m \subseteq N, n_1, \dots, n_m \in \mathbb{N} \text{ and } \varphi_1, \dots, \varphi_m \in \mathcal{L}\}$ to the set A of atomic actions of the language. In fact, this is a multiple-delegation action. Action $\{(i, \text{deleg}((G_1, n_1, \varphi_1), \dots, (G_m, n_m, \varphi_m)))\}$ means ‘agent i ascribes forward-looking responsibility for φ_k after n_k steps to G_k , for each $1 < k < m$ ’. We require that such actions satisfy:

$$(A14)$$

$$\text{Power}(i, j) \rightarrow [(i, \text{deleg}((G_1, n_1, \varphi_1), \dots, (G_m, n_m, \varphi_m))) K_j R_k^n \varphi_k \quad (\text{for each } j \in G \text{ and each } 1 < k < m)$$

A14 stipulates that, when agent i delegates φ to G , it creates a forward-looking responsibility that φ for every agent in G , but only if agent i has the power to do so. Note that, in this definition of delegation, the actor of the delegation remains responsible. In addition, the agents in G also know their new responsibility. This is why power links must also be coordination links. (Recall SC7.)

3.4 Indirect responsibility

Note that Definition 5 does not take indirect agency into account. It defines forward-looking responsibility as the case where the group

must guarantee the outcome by itself. This is not well adapted for an organisation setting, where delegation actions and information actions are available. Therefore, we redefine forward-looking responsibility in Definition 6 below. For the sake of readability it is given in two parts. In the first part we can see that the only difference from Definition 5 is the presence of operator IR. The latter operator stands for ‘indirect responsibility’.

Definition 6 (Again, we show only the first three, the others are analogous.)

$$\begin{aligned} R_G^0(\varphi) &\stackrel{\text{def}}{=} O_G(\varphi \vee IR_G^0(\varphi)) \\ R_G^1(\varphi) &\stackrel{\text{def}}{=} O_G(A_G\varphi \vee IR_G^1(\varphi)) \wedge XR_G^0(\varphi) \\ R_G^2(\varphi) &\stackrel{\text{def}}{=} O_G(A_G(A_G\varphi \vee IR_G^1(\varphi)) \vee IR_G^2(\varphi)) \wedge XR_G^1(\varphi) \end{aligned}$$

where IR_G^n is defined as:

$$\begin{aligned} IR_G^0(\varphi) &\stackrel{\text{def}}{=} \exists_{G'_1} K_G(\text{Power}(G, G'_1) \wedge R_{G'_1}^0(\varphi_1) \wedge \dots \wedge \\ &\quad \exists_{G'_m} K_G(\text{Power}(G, G'_m) \wedge R_{G'_m}^0(\varphi_m) \wedge \\ &\quad K_G((\varphi_1 \wedge \dots \wedge \varphi_m) \rightarrow \varphi)) \\ IR_G^1(\varphi) &\stackrel{\text{def}}{=} \exists_{G'_1} K_G(\text{Power}(G, G'_1) \wedge A_{G'_1}\varphi \wedge R_{G'_1}^1(\varphi_1) \wedge \dots \wedge \\ &\quad \exists_{G'_m} K_G(\text{Power}(G, G'_m) \wedge A_{G'_m}\varphi \wedge R_{G'_m}^1(\varphi_m) \wedge \\ &\quad K_G((\varphi_1 \wedge \dots \wedge \varphi_m) \rightarrow \varphi)) \\ IR_G^2(\varphi) &\stackrel{\text{def}}{=} \exists_{G'_1} K_G(\text{Power}(G, G'_1) \wedge A_{G'_1}(A_{G'_1}\varphi_1 \vee IR_{G'_1}^1(\varphi_m)) \wedge \\ &\quad R_{G'_1}^2(\varphi_1)) \wedge \dots \wedge \\ &\quad \exists_{G'_m} K_G(\text{Power}(G, G'_m) \wedge A_{G'_m}(A_{G'_m}\varphi_m \vee IR_{G'_m}^1(\varphi_m)) \wedge \\ &\quad R_{G'_m}^2(\varphi_m)) \wedge \\ &\quad K_G((\varphi_1 \wedge \dots \wedge \varphi_m) \rightarrow \varphi) \end{aligned}$$

where $\text{Power}(G, G') \stackrel{\text{def}}{=} \exists_{i \in G} \forall_{j \in G'} \text{Power}(i, j)$.

Informally, group G is forward-looking responsible for φ after n steps if and only if it is obligatory for G that it is able to bring about φ after n steps or that it is indirect forward-looking responsible for φ after n steps. The definition of indirect responsibility is the most complex part. Yet, it is intuitive. We consider that a group G is indirect responsible for φ if and only if there are groups $G'_1, \dots, G'_m$ under G 's power that are able to bring about $\varphi_1, \dots, \varphi_m$, or they are indirect responsible for $\varphi_1, \dots, \varphi_m$, and such that $\varphi_1, \dots, \varphi_m \rightarrow \varphi$.

3.5 Example

Consider an organisation with two bank accounts, 1 and 2. Alice (agent a) is the member of the organisation who manages the accounts. Betty (agent b) is the member who normally pays the bills for the organisation. And Carol (agent c) is the director of their department, but she has no access to the accounts.

Let $\langle M, w_0 \rangle$ be a pointed CEDL model, where M is depicted in Figure 1. It describes a situation where Betty will pay a bill after three steps using account 1. Alice knows that, but she does not know which account Betty will use. Alice also knows that none of the accounts has enough money to pay the bill, but she does know that both accounts together do. Group $G = \{a, b, c\}$ is ascribed the responsibility to have the bill paid and the balances non-negative after three steps, i.e., $\langle M, w_0 \rangle \models R_G^3(\text{paid} \wedge \text{bal}(n) \geq 0)$.

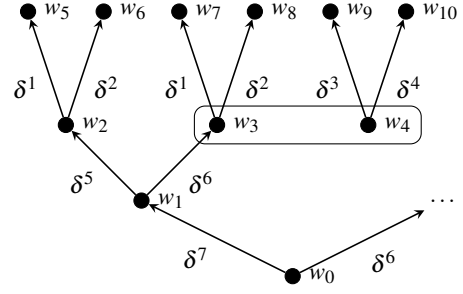


Figure 1. The CEDL model $M = \langle O, W, \mathcal{T}, \mathcal{X}, \pi \rangle$ for the example in Section 3.5. (The dots represent possible worlds in W , the arrows represent transitions in $\mathcal{T}$, and the rectangle represents $\mathcal{X}(a)$. The joint actions and the valuation function π are detailed in Table 1. The organisation $O = \langle G, P, I, c \rangle$, where $G = \{a, b, c\}$, $P = \{(c, a), (c, b)\}$, and $I = P \cup \{(b, a)\}$.)

Table 1. On the left, the joint actions for the model M in Figure 1. On the right, the valuation function π for the model M in Figure 1. (Action info abbreviates $\text{info}(a, [b : \text{pay}(2)] \neg \top)$ and action deleg abbreviates $\text{deleg}((a, 2, \text{bal}(n) \geq 0), (b, 1, A_a \text{bal}(n) \geq 0), (b, 2, \text{paid}))$. The bullets show where the atoms are true.)

	a	b	c		V_G	V_a	V_b	V_c	$\text{bal}(n) \geq 0$
δ^1	tr(2, 1)	pay(1)	wait	w_0					•
δ^2	tr(1, 2)	pay(1)	wait	w_1					•
δ^3	tr(2, 1)	pay(2)	wait	w_2					•
δ^4	tr(1, 2)	pay(2)	wait	w_3			•		•
δ^5	wait	info	wait	w_4					•
δ^6	wait	wait	wait	w_5	•	•	•	•	•
δ^7	wait	wait	deleg	w_6					•
				w_7	•	•	•	•	•
				w_8	•	•	•	•	•
				w_9	•	•	•	•	•
				w_{10}					•

Note that $\langle M, w_2 \rangle \models I_G(\delta^2)$. That is, at w_2 it is irresponsible for G that Alice places the money on account 2, Betty pays the bill from account 1 and Carol just waits. This is true because $\langle M, w_2 \rangle \models \neg K_G E_{\delta^2} \neg V_G$, i.e., G knows that δ^2 does not avoid the violation, and also $\langle M, w_2 \rangle \models A_G \neg V_G$, i.e., G knows that there is an action that ensures a non-violation state, namely δ^1 . Therefore, $\langle M, w_2 \rangle \models I_G(\delta^2)$, which implies $\langle M, w_0 \rangle \models I_G(\delta^7; \delta^5; \delta^2)$ and thus $\langle M, w_0 \rangle \models B_{\delta^7; \delta^5; \delta^2}(\neg \text{bal}(n) \geq 0)$.

Now, suppose for a moment that $P = \emptyset$, so that action deleg does not produce its normal effect of ascribing forward-looking responsibilities to Alice and Betty. That is, under this assumption $\pi(V_a) = \pi(V_b) = \emptyset$. Then, we have $\langle M, w_2 \rangle \models \neg I_a(\delta^2)$, because there is no violation for Alice in the model. Therefore, $\langle M, w_0 \rangle \models \neg I_a(\delta^7; \delta^5; \delta^2)$. Similarly, there is no violation for Betty either, so $\langle M, w_0 \rangle \models \neg I_b(\delta^7; \delta^5; \delta^2)$. Because of Axiom A10, the violations for Carol are present in the model. However, we have $\langle M, w_2 \rangle \models \neg I_c(\delta^2)$. The reason for this is that at w_2 Carol cannot ensure a non-violation state no matter which action she decides to execute. Therefore, $\langle M, w_0 \rangle \models \neg I_c(\delta^7; \delta^5; \delta^2)$. This means that neither Alice, Betty or Carol can be held backward-looking responsible for a violation state. Thus $\langle M, w_0 \rangle \models \text{PMH}_{\delta^7; \delta^5; \delta^2}(\neg \text{bal}(n) \geq 0)$, i.e., by performing such actions, the PMH arises in group G with respect to outcome $\neg \text{bal}(n) \geq 0$.

Now, let us return to the original configuration of P . In this case, after the execution of δ^7 , Alice is forward-looking responsible for $\text{bal}(n) \geq 0$. Then, if δ^2 is executed at w_2 , Alice is held backward-

looking responsible, i.e., $\langle M, w_2 \rangle \models B_{\delta^2}(\neg \text{bal}(n) \geq 0)$. That is, the PMH does not arise in the presence of the organisational structure proposed and the organisational actions info and deleg.

Now, let us see the importance of using the new Definition 6 (instead of the old Definition 5) in the formalisation of this example. If we were using the old definition, the PMH could not be avoided, because it does not take delegation and indirect agency into account. The leader of the group, Carol, has no action in her repertoire to ensure a non-violation state herself. It does not matter what she does, she will never be held backward-looking responsible. But with the new definition $\langle M, w_0 \rangle \models I_c(\delta^6)$, i.e., she can be held responsible, because δ^6 does not ensure a non-violation state for Carol, while there is an action that does, namely, δ^7 , because $\langle M, w_0 \rangle \models K_c E_{\delta^7} (\text{IR}_{\{a,b\}}^2 (\text{paid} \wedge \text{bal}(n) \geq 0))$.

Finally, it is also interesting to notice that the information action executed by Betty on w_1 is important to make Alice aware of which account she will use to pay the bill. We invite the reader to do the exercise of considering $I = \emptyset$. In this case, action info will not produce its effect and therefore, the PMH arises.⁵

4 RELATED WORK

We are not aware of other attempts to formalise the PMH in organisations. On the other hand, there are some attempts to formalise responsibility. Here, we mention four approaches that we find relevant.

The first one is Santos and Carmo's approach [16]. This is perhaps the first one to propose that forward-looking responsibility should be understood as 'obligation to ensure', i.e., $O_i E_i \varphi$. But different from ours, in Santos and Carmo's logic the operator E is the classical ensuring operator proposed by Pörn in [15].

The second one is Cholvy et al.'s approach [6]. They use ideas that are similar to Santos and Carmo's to formalise three different meanings of the term responsibility. The first two are similar to our forward-looking and backward-looking responsibility (where they call the latter, answerability). The third one is called accountability. It is described as "a position which enables you to make decisions in a given organization but implies that you must be prepared to justify your actions".

There are several differences between these two approaches and ours. But maybe the most important is the difference in perspective. By using Pörn's ensuring operator E, they abstract away from agents' actions. This means that those formalisms enable only to reason about the responsibilities that emerge from the behaviour of the agents. On the other hand, our formalism also enables us to reason about the behaviour that emerges from the responsibilities ascribed to agents. This is so because, if we suppose that our agents will try to fulfill their responsibilities, we can use our formalism to try to predict which actions they will decide to execute. That is, we may say that our formalism gives us the opportunity to reason about responsible actions, or, about what responsible agents would do.

The other two approaches we want to mention are Dignum and Dignum's approach [9] and Grossi et al.'s approach [13]. They share the same perspective as ours on agents' actions. For instance, Grossi et al.'s approach is also based on a kind of epistemic dynamic logic. However, it is simply not possible to formalise the PMH on these frameworks, for they account for individual responsibility only.

5 CONCLUSION

The main contribution of this work is the extension of CEDL in two ways. The first is the notion of distributed knowledge, and the second is the notion of organisational structure. We show that with the first extension it is possible to generalise the definitions of knowledge, ability, and responsibility to groups of agents. With the second, we show that we can formalise organised groups of agents. In particular, we show that we can formalise information and delegation actions. Now we are in position to model the problem of many hands in organisations, and to show how organised groups of agents are more likely to avoid such problem.

Possible future work includes a couple of questions and improvements. For instance, we did not address decidability, complexity, and expressiveness of the logic. For the improvements part, we expect to be able to enrich CEDL language with complex actions. This would enable to reason about conditional plans and maybe also repetition. Another possible improvement would be the incorporation of other temporal operators, such as 'always' and 'until', already present, e.g., in ATL.

REFERENCES

- [1] R. Alur, T. A. Henzinger, and O. Kupferman, 'Alternating-time temporal logic', *Journal of the ACM*, **5**(49), 672–713, (2002).
- [2] A. Anderson, 'A reduction of deontic logic to alethic modal logic', *Mind*, **67**, 100–103, (1958).
- [3] Aristotle, *Nicomachean Ethics*, 350 BC. Translated by W.D. Ross. <http://classics.mit.edu/Aristotle/nicomachaen.html>.
- [4] M. Bovens, *The Quest for Responsibility. Accountability and Citizenship in Complex Organisations*, Cambridge University Press, 1998.
- [5] J. Broersen, A. Herzig, and N. Troquard, 'A normal simulation of coalition logic and an epistemic extension', in *Proc. of TARK-2007*, ed., D. Samet, pp. 92–101, (2007).
- [6] L. Cholvy, F. Cuppens, and C. Saurel, 'Towards a logical formalization of responsibility', in *Proc. of the Sixth Int. Conf. on AI and Law*, pp. 233–242. ACM Press, (1997).
- [7] P. d'Altan, J.-J. Ch. Meyer, and R. Wieringa, 'An integrated framework for ought-to-be and ought-to-do constraints', *Artificial Intelligence and Law*, **4**, 77–111, (1996).
- [8] T. de Lima, L. Royakkers, and F. Dignum, 'A logic for reasoning about responsibility', *Logic Journal of IGPL*, **18**(1), 99–117, (2010). DOI: 10.1093/jigpal/jzp073.
- [9] V. Dignum and F. Dignum, 'A logic for agent organizations', in *Proc. of FAMAS-2007*, (2007).
- [10] R. Fagin, J. Halpern, Y. Moses, and M. Vardi, *Reasoning about Knowledge*, The MIT Press, 1995.
- [11] H. Ghaderi, H. Levesque, and Y. Lesprance, 'A logical theory of coordination and joint ability', in *Proc. of AAAI-2007*, pp. 421–426, (2007).
- [12] A. Goldman, 'Social epistemology', in *The Stanford Encyclopedia of Philosophy*, ed., E.N. Zalta, fall 2009 edn., (2009).
- [13] D. Grossi, L. Royakkers, and F. Dignum, 'Organizational structure and responsibility: An analysis in a dynamic logic of organized collective agency', *Artificial Intelligence and Law*, **15**, 223–249, (2007).
- [14] M. Pauly, 'A modal logic for coalitional power in games', *Journal of Logic and Computation*, **12**(1), 149–166, (2002).
- [15] I. Pörn, *Action Theory and Social Science*, volume 120 of *Synthese Library*, D. Reidel, 1977.
- [16] F. Santos and J. Carmo, 'Indirect action, influence and responsibility', in *Deontic Logic, Agency and Normative Systems*, eds., M. A. Brown and J. Carmo, C. J. van Rijsbergen, editor, Workshops in Computing series, Springer, (1996).
- [17] D.F. Thompson, 'Moral responsibility and public officials', *American Political Science Review*, **74**, 905–916, (1980).
- [18] W. van der Hoek and M. Wooldridge, 'Cooperation, knowledge, and time: Alternating-time temporal epistemic logic and its applications', *Studia Logica*, **75**, 125–157, (2003).

⁵ Note however, that under such assumption the model M must be changed. For instance, if action info does not produce any effect, worlds w_2 and w_3 should be bisimilar.