Submitted on August 20, 2023 to the Notre Dame Journal of Formal Logic
Volume ??, Number ??,

Responsibility in Infinite Games

Xiulin Cui and Pavel Naumov

Abstract There are two distinct forms of responsibility that can be found in literature: counterfactual responsibility and responsibility for "seeing to it that". It has been previously observed that, in the case of strategic games, the counterfactual form of responsibility can be defined through responsibility for "seeing to it that", but not the other way around.

The article considers these two forms of responsibility in the case of infinite extensive form games. The main technical result is that, in this new setting, neither of the two forms of responsibility can be defined through the other. Some preliminary results for finite extensive form games are also given.

1 Introduction

In this article, we study two forms of responsibility in infinite games. As an example, consider a situation in which Alice and Bob just got married and received a box of fancy chocolate candies as a wedding gift. They have eaten all but one candy and each of them feels uncomfortable eating the last candy.

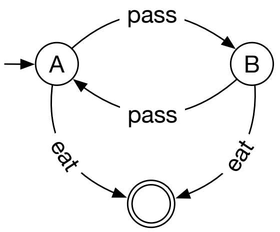

Figure 1 Infinite game played on a finite graph.

This situation can be modelled as an infinite game on a finite graph [12] depicted in Figure 1. This graph has three states. In two of these states, labelled by a and b, players Alice and Bob, respectively, make a decision either to "pass" the choice to the other player or to "eat" the last candy. The game can either run forever or terminate if one of the players eats the candy and, thus, transitions the game into the third (final) state. For the sake of this example, we assume that Alice starts the game. To show this, we marked state A as the initial state of the game.

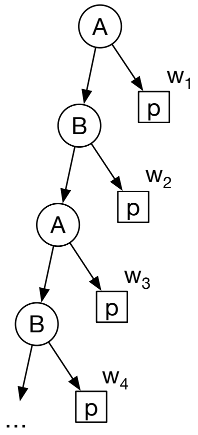

Figure 2 Infinite extensive form game. Propositional variable p represents the statement "The last candy is eaten".

To analyse the responsibility of agents in infinite games on finite graphs, it is convenient to represent them as *infinite extensive form* games [7, 5, 1]. Such a representation of our game is depicted in Figure 2. Descending paths in this tree starting at the root node correspond to different plays of the game. Our game has a unique infinite path that corresponds to a play under which the candy is never eaten. It also has infinitely many *finite* paths, terminating in the leaf nodes, that correspond to the plays under which the candy is eaten. In this article, we study statements about the outcomes (leaf nodes) of the games. An example of such a statement is "the last candy is eaten". We denote it by propositional variable p. Note that each leaf node in our diagram is labelled with variable p. This reflects the fact that the statement "the last candy is eaten" is true in each of the outcomes of the game.

Although each infinite game on a finite graph, like the one depicted in Figure 1, can be unfolded into an infinite extensive form game, like in Figure 2, the converse is not true. Infinite extensive form games capture a more general class of games than infinite games on finite graphs. In this article, we study responsibility in the more general class of infinite extensive form games.

Let us now suppose that the game ends in outcome w_1, see Figure 2. In other words, Alice eats the last candy the first moment she is given a chance to do this. Note that by doing this (going right on the tree) she *made* the statement "the last candy is eaten" to be *unavoidably true*. Indeed, if she would have chosen a different action (going left) the last candy might have never been eaten. Each time when an agent takes an action that makes a statement unavoidably true (and it was not unavoidably true before the action), we say that the agent is responsible for *seeing to* *it that* the statement is true. We write this as

$$w_1 \Vdash \mathsf{S}_{\text{Alice}}(\text{“the last candy is eaten”}). \tag{1}$$

We read statement (1) as “on the path of play leading to outcome w_1, Alice has seen to it that the last candy is eaten”. Modality S has been extensively studied in STIT logic [3, 9, 10, 8, 19]. The version of seeing-to-it that we consider in this article is known under the name “achievement stit” [4, 10].

Let us now return to our example and consider outcome w_2. Note that the last candy is also eaten on the path leading to this outcome, but it is Bob, not Alice, who made the statement “the last candy is eaten” unavoidably true. Thus,

$$w_2 \Vdash \mathsf{S}_{\text{Bob}}(\text{“the last candy is eaten”}), \tag{2}$$
$$w_2 \nVdash \mathsf{S}_{\text{Alice}}(\text{“the last candy is eaten”}).$$

In general, statement

$$w_i \Vdash \mathsf{S}_{\text{Alice}}(\text{“the last candy is eaten”})$$

is true iff number i is odd. By the *truth set* $[\![\varphi]\!]$ of a formula φ we mean the set of all outcomes in which φ is satisfied. In our case, $[\![\mathsf{S}_{\text{Alice}}(\text{“the last candy is eaten”})]\!]$ is the set $\{w_{2i+1} \mid i \geq 0\}$. We visualise this set on the left diagram in Figure 3.

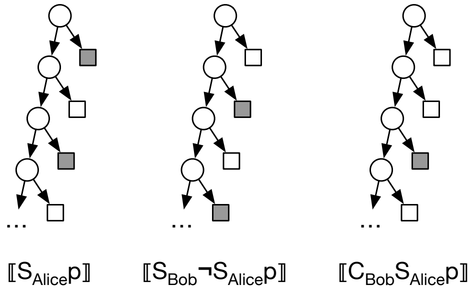
$[\![\mathsf{S}_{\text{Alice}}p]\!]$
$[\![\mathsf{S}_{\text{Bob}}\neg\mathsf{S}_{\text{Alice}}p]\!]$
$[\![\mathsf{C}_{\text{Bob}}\mathsf{S}_{\text{Alice}}p]\!]$
...

Figure 3 Propositional variable p represents the statement “the last candy is eaten”. Grey squares on each diagram show the elements of the corresponding truth set.

Recall from statement (2) that, in outcome w_2, Bob is responsible for seeing to it that the last candy is eaten. Although he can be blamed for this, he has an excuse. By making the move into leaf node w_2, Bob made it unavoidable that statement $\mathsf{S}_{\text{Alice}}p$ is false, see the left diagram in Figure 3. Thus,

$$w_2 \Vdash \mathsf{S}_{\text{Bob}}\neg\mathsf{S}_{\text{Alice}}(\text{“the last candy is eaten”}).$$

In other words, in outcome w_2, the kind newlywed husband Bob has seen to it that his wife is not responsible for seeing to it that the last candy is eaten.

In general, the statement

$$w_i \Vdash \mathsf{S}_{\text{Bob}}\neg\mathsf{S}_{\text{Alice}}(\text{“the last candy is eaten”})$$

is true iff number i is even. We show the truth set of the formula

$$[\![\mathsf{S}_{\text{Bob}}\neg\mathsf{S}_{\text{Alice}}(\text{“the last candy is eaten”})]\!]$$

on the middle diagram in Figure 3.

Let us go back to outcome w_1, where Alice eats the last candy the first moment she is given a chance to do this. We have previously noticed, see statement (1), that in this outcome she has seen to it that the last candy is eaten. Observe that by making the move into leaf node w_1 she made it unavoidable that statement $\mathsf{S}_{\text{Bob}}\neg\mathsf{S}_{\text{Alice}}p$ is false, see the middle diagram in Figure 3. Hence,

$$w_1 \Vdash \mathsf{S}_{\text{Alice}}\neg\mathsf{S}_{\text{Bob}}\neg\mathsf{S}_{\text{Alice}}(\text{“the last candy is gone”}).$$

In other words, she made sure that her darling husband has no excuse to eat the last candy!

Seeing-to-it modality S captures one possible form of responsibility. A very different definition of responsibility is proposed by Frankfurt: “a person is morally responsible for what he has done only if he could have done otherwise” [6]. Although Frankfurt himself discusses many limitations to this definition, it became one of the standard approaches to defining responsibility in philosophy [21]. We refer to this form of responsibility as *counterfactual responsibility*. This form of responsibility is also sometimes called “backward responsibility” [22]. We use modality C to capture counterfactual responsibility. Logical systems for reasoning about counterfactual responsibility have been proposed in [13, 16, 15, 14].

As an example, recall from (1) that, in outcome w_1, Alice is seeing to it that the last candy is eaten. Note that she can easily avoid seeing to this by never eating the last candy. Thus, in outcome w_1, Alice is counterfactually responsible for seeing to it that the last candy is eaten:

$$w_1 \Vdash \mathsf{C}_{\text{Alice}}\mathsf{S}_{\text{Alice}}(\text{“the last candy is gone”}).$$

At the same time, if the game ends in outcome w_1, then Bob has no chance to prevent Alice from seeing to it that the last candy is eaten:

$$w_1 \nVdash \mathsf{C}_{\text{Bob}}\mathsf{S}_{\text{Alice}}(\text{“the last candy is gone”}).$$

It is interesting to point out that the situation is different in outcome w_3, where Alice is also responsible for seeing to it that the last candy is eaten. Prior to reaching this outcome, Bob had an opportunity to eat the last candy himself (go to outcome w_2). By doing this, he would prevent Alice from being responsible for seeing to it that the last candy is eaten. Thus, in outcome w_3, Bob is counterfactually responsible for Alice seeing to it that the last candy is eaten:

$$w_3 \Vdash \mathsf{C}_{\text{Bob}}\mathsf{S}_{\text{Alice}}(\text{“the last candy is gone”}).$$

To put it in other words, on the path of play leading to outcome w_3, Bob had a chance to spare Alice from the temptation to eat the last candy. He did not do this and, as a result, is counterfactually responsible. In general, the statement $w_i \Vdash \mathsf{C}_{\text{Bob}}\mathsf{S}_{\text{Alice}}p$ is true for each odd integer $i \geq 3$, see right diagram in Figure 3.

Finally, observe from the right diagram in Figure 3 that along the path of play leading to outcome w_3, Alice could have easily prevented $\mathsf{C}_{\text{Bob}}\mathsf{S}_{\text{Alice}}p$ from being true if she would have eaten the candy herself the first moment she had a chance to do this:

$$w_3 \Vdash \mathsf{C}_{\text{Alice}}\mathsf{C}_{\text{Bob}}\mathsf{S}_{\text{Alice}}(\text{“the last candy is gone”}).$$

2 Contribution

In this article, we formally define modalities S and C and study the properties of the interplay between them. One possible way to study these properties is to develop an axiomatic system for the language containing both modalities. Another is to study the definability of one of these modalities through the other. In this work, we focus on the definability results.

The definability of modalities S and C through each other has been previously investigated by Naumov and Tao [17, 18]. The setting of their work is significantly different from the setting of the last candy game because they consider *strategic* games in which all agents act simultaneously and just once. In that setting, they establish two important results. First, they show that counterfactual modality C can be defined through seeing-to-it modality S as follows:

$$\mathsf{C}_a\varphi \equiv \varphi \wedge \mathsf{S}_a\neg\mathsf{S}_a\neg\varphi. \tag{3}$$

The expression $\mathsf{S}_a\neg\mathsf{S}_a\neg\varphi$ by itself has been studied in philosophy literature, where it is referred to as "forbearing" [20, p.45] and "refraining" [2]. Second, Naumov and Tao have shown that, in the case of strategic games, seeing-to-it modality S is *not* definable through counterfactual modality C.

In this article, we formally define modalities S and C for infinite extensive form games and show that, unlike the strategic games' case, neither of them is definable through the other. Indirectly, these results, together with the observations in [17, 18], show that infinite extensive form games provide a significantly richer than the strategic game setting for modelling multiagent interactions. We also discuss the possibility of extending our results to *finite* extensive form games and state some partial results there.

The rest of the article is structured as follows. First, we define infinite extensive form games and related notations. Then, we introduce the formal syntax and semantics of our modal language. In Section 5, we show that modality S is not definable via modality C. In Section 6, we show that modality C is not definable via modality S. In Appendix 7, we discuss some preliminary results for *finite* extensive form games. Section 8 concludes.

3 Infinite Games

Throughout the article, we assume a fixed set of propositional variables and a fixed set of agents. By an *infinite extensive form game* we mean a (possibly infinite) tree whose non-leaf nodes are labelled by agents and whose leaf nodes are labelled by sets of propositional variables. Different nodes might have the same label and not all labels have to be used. When we say that a tree is possibly infinite, we mean that any node of the tree might have infinitely many children and that the tree might have (at most ω) infinite depth. We assume that the path from the root of the tree to each node has a finite length. We refer to leaf nodes as *outcomes*.

An example of an infinite extensive form game is depicted in Figure 4. In this game, for example, non-leaf nodes n_1 and n_2 are labelled by agents a and b, respectively. Outcomes w_1 and w_2 are labelled with sets $\{p\}$ and $\varnothing$, respectively.

By $Ancestors(n)$ of a node n we mean the finite set of all nodes on the path from the root node of the game to node n. In our example, $Ancestors(n_2) = \{n_1, n_2\}$ and $Ancestors(w_1) = \{n_1, n_2, n_4, w_1\}$.

By $Subtree(n)$ we mean the (possibly infinite) set of nodes located at the subtree starting at node n. In our example, $Subtree(n_3)$ is the finite set $\{n_3, w_3, n_6, w_4, w_5\}$, but $Subtree(n_2)$ is the *infinite* set $\{n_2, n_4, n_5, w_1, w_2, \dots\}$.

By $Outcomes(n)$ we mean the (possibly infinite) set of all outcomes in the set $Subtree(n)$. In our example, $Outcomes(n_3)$ is the set $\{w_3, w_4, w_5\}$.

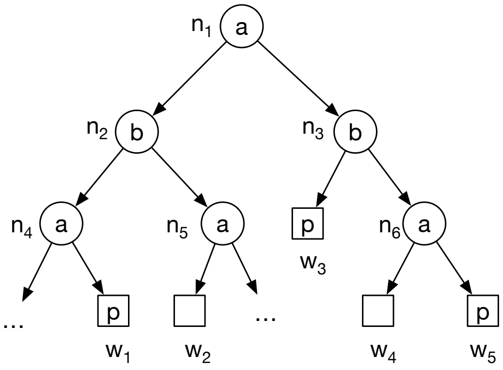

Figure 4 An infinite extensive form game.

By a strategy of an agent a in the subtree starting at node n, we mean a function that to each node labelled with agent a in the set $Subtree(n)$ assigns a specific outgoing edge of this node. The set of all such functions is denoted by $Strategies_a(n)$. In our example, "choose left" function $\ell(m)$, that assigns the left outgoing edge to each node m labelled with agent a in the entire tree, belongs to set $Strategies_a(n_1)$. As another example, consider "choose left on the left, choose right on the right" function $f(m)$ that assigns the *left* outgoing edge to each node m labelled with agent a in $Subtree(n_2)$ and assigns the *right* outgoing edge to each node m labelled with agent a in $Subtree(n_3)$. This function also belongs to the set $Strategies_a(n_1)$.

For any strategy $s \in Strategies_a(n)$, by $Outcomes_a(n,s)$ we mean the set of outcomes $w \in Outcomes(n)$ such that for each non-leaf node m on the path from node n to outcome w, the path contains the edge $s(m)$. For example, for our "choose left" function $\ell \in Strategies_a(n_1)$ set $Outcomes_a(n_1,\ell)$ contains outcome w_2 and does *not* contain outcomes w_1, w_3, w_4, and w_5. At the same time, for a similar "choose left" function $\ell' \in Strategies_a(n_3)$ for the subtree starting at node n_3, set $Outcomes_a(n_3,\ell')$ is the set $\{w_3,w_4\}$.

4 Syntax and Semantics

The language Φ that we consider in this article is defined by the grammar:

$$\varphi := p \mid \neg\varphi \mid \varphi\vee\varphi \mid \mathsf{S}_a\varphi \mid \mathsf{C}_a\varphi,$$

where p is a propositional variable and a is an agent. We read $\mathsf{S}_a\varphi$ as "agent a sees to φ" and $\mathsf{C}_a\varphi$ as "agent a is counterfactually responsible for φ". We assume the conjunction $\wedge$, constant false $\bot$, and constant true $\top$ are defined through negation $\neg$ and disjunction $\vee$ in the standard way.

Definition 1 *For any outcome w of an infinite extensive form game and any formula $\varphi \in \Phi$, the satisfaction relation $w \Vdash \varphi$ is defined as follows:*

1. *$w \Vdash p$, if outcome w is labelled with a set containing propositional variable p,*
2. *$w \Vdash \neg\varphi$, if $w \nVdash \varphi$,*
3. *$w \Vdash \varphi\vee\psi$, if $w \Vdash \varphi$ or $w \Vdash \psi$,*

4. *$w \Vdash \mathsf{S}_a\varphi$, if there exists a non-root node $n \in Ancestors(w)$ such that $n \rightsquigarrow \varphi$, node $parent(n)$ is labelled with agent a, and $parent(n) \not\rightsquigarrow \varphi$,*
5. *$w \Vdash \mathsf{C}_a\varphi$ if $w \Vdash \varphi$ and there is a node $n \in Ancestors(w)$ and a strategy $s \in Strategies_a(n)$ of agent a such that $u \nVdash \varphi$ for each outcome $u \in Outcomes_a(n,s)$.*

where for any node n and any formula φ relation $n \rightsquigarrow \varphi$ hold when $Subtree(n)$ is finite and $u \Vdash \varphi$ for each outcome $u \in Outcomes(n)$.

For any given game, let the truth set $[\![\varphi]\!]$ of a formula $\varphi \in \Phi$ be the set of all *outcomes* (leaf nodes) w of the game such that $w \Vdash \varphi$.

Definition 2 *Formulae $\varphi, \psi \in \Phi$ are semantically equivalent if $[\![\varphi]\!] = [\![\psi]\!]$ in each infinite extensive form game.*

5 Undefinability of S through C

In this section, we show that, in the infinite extensive form game setting, modality S is not definable through modality C. Without loss of generality, we assume that our language contains a single propositional variable p and just two agents, a and b. We can assume that the language has only two agents because we do not require all agents to be used as labels of non-leave nodes.

Traditionally, the undefinability of one modality through another is established using the bisimulation technique. In our case, this approach would consist in specifying two games that are indistinguishable in the language without modality S and are distinguishable using modality S.

In this article, we use a different method of proving undefinability. This new method, called "truth set algebra", has been recently proposed by Knight and Naumov [11]. The method is based on the analysis of "truth sets" of formulae for a given single game.

In our case, we use the infinite extensive form game depicted atop of Figure 5. In the game, first, agent a is given a chance to make proposition variable p false. If agent a does not do this, variable p will never become false. Instead, agents a and b will take turns to decide either to terminate the game with p being true or to continue the game.

We visualise truth set $[\![\varphi]\!]$ by shading grey all outcomes that belong to the set. The four middle diagrams in Figure 5 visualise sets $[\![p]\!]$, $[\![\neg p]\!]$, $[\![\bot]\!]$, and $[\![\top]\!]$.

Lemma 1 *$[\![\mathsf{C}_a\varphi]\!], [\![\mathsf{C}_b\varphi]\!] \in \{[\![p]\!], [\![\neg p]\!], [\![\bot]\!], [\![\top]\!]\}$ for any $\varphi \in \Phi$, such that $[\![\varphi]\!] \in \{[\![p]\!], [\![\neg p]\!], [\![\bot]\!], [\![\top]\!]\}$.*

Proof We consider the following three cases separately:
Case I: $[\![\varphi]\!] = [\![p]\!]$. Then, $[\![\varphi]\!] = \{w_2, w_3, w_4, \dots\}$, see Figure 5. Thus, formula φ is satisfied in outcomes w_2, w_3, w_4, ... Agent a can prevent φ in each of these outcomes by going to outcome w_1. Hence, agent a is counterfactually responsible for φ in each of the outcomes w_2, w_3, w_4, ... Thus, $[\![\mathsf{C}_a\varphi]\!] = \{w_2, w_3, w_4, \dots\} = [\![p]\!]$.

At the same time, agent b cannot prevent φ in each of the outcomes w_2, w_3, w_4, ... Thus, $[\![\mathsf{C}_b\varphi]\!] = \varnothing = [\![\bot]\!]$.

In Figure 5, we show these two observations by arrows labelled with C_a and C_b from the diagram representing set $[\![p]\!]$ to the diagrams representing sets $[\![p]\!]$ and $[\![\bot]\!]$, respectively.

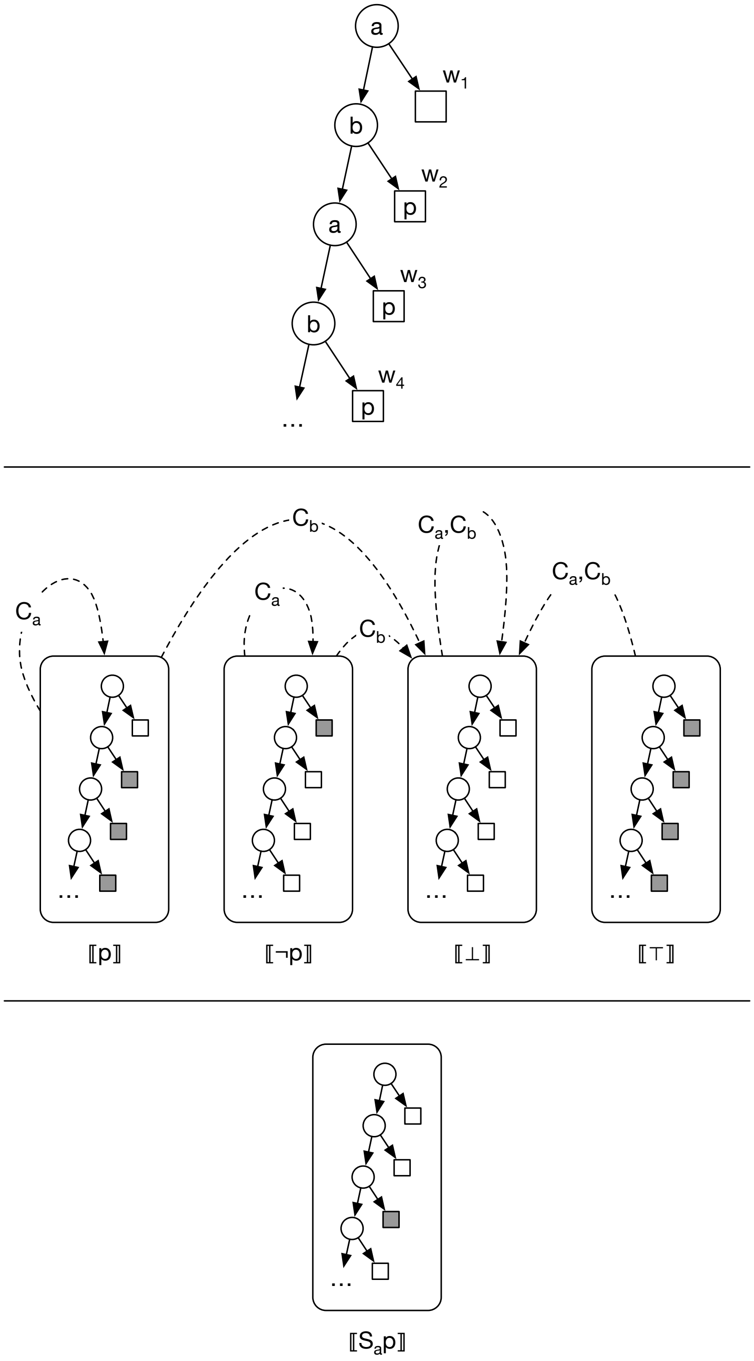

Figure 5 Towards undefinability of modality S via modality C. In the bottom diagram, the colours of leaf nodes alternate starting from the second leaf from the top.

Case II: $[\![\varphi]\!] = [\![\neg p]\!]$. Thus, $[\![\varphi]\!] = \{w_1\}$, see Figure 5. Note that agent *a* had a strategy to guarantee that the game does not end in an outcome in which *p* is true.

The strategy consists in going left on the first step. What agent a does after that is not important. Thus, $w_1 \Vdash \mathsf{C}_a p$. Statement $\mathsf{C}_a p$ is not satisfied in all other outcomes by item 5 of Definition 1 because p is not satisfied in all of the other outcomes. Therefore, $[\![\mathsf{C}_a\varphi]\!] = \{w_1\} = [\![\neg p]\!]$.

At the same time, along the path leading to outcome w_1, agent b had no strategy to guarantee that the game does not end in an outcome in which φ is true. Thus, $w_1 \nVdash \mathsf{C}_b\varphi$. Statement $\mathsf{C}_b\varphi$ is not satisfied in all other outcomes by item 5 of Definition 1 because φ is not satisfied in all of the other outcomes. Therefore, $[\![\mathsf{C}_b\varphi]\!] = \varnothing = [\![\bot]\!]$.

In Figure 5, we show these two observations by arrows labelled with C_a and C_b from the diagram representing set $[\![\neg p]\!]$ to the diagrams representing sets $[\![\neg p]\!]$ and $[\![\bot]\!]$, respectively.

Case III: $[\![\varphi]\!] = [\![\bot]\!]$. Thus, $[\![\varphi]\!] = \varnothing$, see Figure 5. Hence, formula φ is not satisfied in each outcome of the game. Thus, by item 5 of Definition 1, formulae $\mathsf{C}_a\varphi$ and $\mathsf{C}_b\varphi$ also are not satisfied in each outcome of the game. Therefore, $[\![\mathsf{C}_a\varphi]\!] = [\![\mathsf{C}_b\varphi]\!] = \varnothing = [\![\bot]\!]$.

Case IV: $[\![\varphi]\!] = [\![\top]\!]$. Hence, the truth set $[\varphi]\!]$ contains *all* outcomes of the game, see Figure 5. This means that the only way to guarantee that the game does not end in an outcome in which φ is true is *to guarantee that the game does not end at all*. Neither of the agents has such ability for the game in Figure 5. Therefore, $[\![\mathsf{C}_a\varphi]\!] = [\![\mathsf{C}_b\varphi]\!] = \varnothing = [\![\bot]\!]$. □

Lemma 2 $[\![\varphi]\!] \in \{[\![p]\!], [\![\neg p]\!], [\![\bot]\!], [\![\top]\!]\}$ *for any formula* $\varphi \in \Phi$ *that uses only modality* C.

Proof We prove the lemma by induction on the structural complexity of formula φ. If formula φ is a propositional variable p, then the statement of the lemma is true because set $\{[\![p]\!], [\![\neg p]\!], [\![\bot]\!], [\![\top]\!]\}$ contains $[\![p]\!]$.

Suppose formula φ has the form $\neg\psi$. Thus, for any outcome w,

$$w \in [\![\varphi]\!] \Leftrightarrow w \in [\![\neg\psi]\!] \Leftrightarrow w \Vdash \neg\psi \Leftrightarrow w \nVdash \psi \Leftrightarrow w \notin [\![\psi]\!],$$

by the definition of $[\![\cdot]\!]$, item 2 of Definition 1, and again the definition of $[\![\cdot]\!]$. In other words, the set of outcomes $[\![\varphi]\!]$ is the *complement* of the set of outcomes $[\![\psi]\!]$. Note that for each of the four truth sets depicted in the middle of Figure 5, the complement of this set is also among those four sets. For example, the complement of $[\![\neg p]\!]$ is $[\![p]\!]$. Thus, the induction hypothesis $[\![\psi]\!] \in \{[\![p]\!], [\![\neg p]\!], [\![\bot]\!], [\![\top]\!]\}$ implies that $[\![\varphi]\!] \in \{[\![p]\!], [\![\neg p]\!], [\![\bot]\!], [\![\top]\!]\}$.

Let formula φ have the form $\psi \vee \chi$. Then, for any outcome w,

$$\begin{aligned} &w \in [\![\varphi]\!] \Leftrightarrow w \in [\![\psi \vee \chi]\!] \Leftrightarrow w \Vdash \psi \vee \chi \Leftrightarrow w \Vdash \psi \text{ or } w \Vdash \chi \\ &\Leftrightarrow w \in [\![\psi]\!] \text{ or } w \in [\![\chi]\!] \Leftrightarrow w \in [\![\psi]\!] \cup [\![\chi]\!]. \end{aligned}$$

by the definition of $[\![\cdot]\!]$, item 3 of Definition 1, and again the definition of $[\![\cdot]\!]$. In other words, the set of outcomes $[\![\varphi]\!]$ is the *union* of the set of outcomes $[\![\psi]\!]$ and the set of outcomes $[\![\chi]\!]$. Note that for any pair of the four truth sets depicted in the middle of Figure 5, the union of these sets is also among those four sets. For example, the union of $[\![p]\!]$ and $[\![\neg p]\!]$ is $[\![\top]\!]$. Thus, the induction hypothesis $[\![\psi]\!], [\![\chi]\!] \in \{[\![p]\!], [\![\neg p]\!], [\![\bot]\!], [\![\top]\!]\}$ implies that $[\![\varphi]\!] \in \{[\![p]\!], [\![\neg p]\!], [\![\bot]\!], [\![\top]\!]\}$.

Finally, let formula φ have the form $\mathsf{C}_g\psi$, where $g\in\{a,b\}$. In this case, the induction assumption $[\![\psi]\!]\in\{[\![p]\!],[\![\neg p]\!],[\![\bot]\!],[\![\top]\!]\}$ implies that $[\![\varphi]\!]\in\{[\![p]\!],[\![\neg p]\!],[\![\bot]\!],[\![\top]\!]\}$ by Lemma 1. □

Lemma 3 $[\![\mathsf{S}_a p]\!]\notin\{[\![p]\!],[\![\neg p]\!],[\![\bot]\!],[\![\top]\!]\}$.

Proof It suffices to show that the truth set $[\![\mathsf{S}_a p]\!]$ is depicted at the bottom of Figure 5.

Consider any node w_i, where $i\geq 1$. By item 4 of Definition 1, for $w_i\Vdash\mathsf{S}_a p$ to be true, along the path leading to w_i, there must exist a non-root node n such that

A. node $\mathit{parent}(n)$ is labelled with agent a,
B. $\mathit{parent}(n)\not\rightsquigarrow\varphi$,
C. $n\rightsquigarrow\varphi$.

By Definition 1, the condition $n\rightsquigarrow\varphi$ implies that $\mathit{Subtree}(n)$ is finite. Thus, the conditions (A), (B), and (C) potentially can be satisfied *only* if $n=w_i$. Hence, for the game depicted atop of Figure 5, statement $w_i\Vdash\mathsf{S}_a p$ is true if

(A′) node $\mathit{parent}(w_i)$ is labelled with agent a,
(B′) $\mathit{parent}(w_i)\not\rightsquigarrow\varphi$,
(C′) $w_i\rightsquigarrow\varphi$.

Condition (A′) is satisfied iff index i is an odd number. Condition (B′) is true for each index i because $\mathit{Subtree}(\mathit{parent}(w_i))$ is *not* finite for each i, see definition of relation $\rightsquigarrow$ in Definition 1. Finally, condition (C′) is satisfied for each integer $i\geq 2$, see the game as depicted atop of Figure 5. Therefore, $w_i\Vdash\mathsf{S}_a p$ iff i is an odd number such that $i\geq 2$. □

The next result follows from the two lemmas above and Definition 2.

Theorem 1 (undefinability) *Formula $\mathsf{S}_a p$ is not semantically equivalent to any formula in language Φ that does not use modality S.*

6 Undefinability of C through S

In this section, we show that modality C is not definable through modality S. Without loss of generality, we again assume that our language contains a single propositional variable p and just two agents, a and b. To prove the desired result, it suffices to show that modality C_a is not definable through any combination of modalities S_a and S_b. We actually prove a slightly *stronger* result that C_a is not definable through any combination of S_a, S_b, and C_b.

Just like in the previous section, we apply the "truth set algebra" technique. This time, we use the game and the truth sets depicted in Figure 6.

Lemma 4 $[\![\mathsf{S}_a\varphi]\!],[\![\mathsf{S}_b\varphi]\!],[\![\mathsf{C}_b\varphi]\!]\in\{[\![p]\!],[\![\neg p]\!],[\![\bot]\!],[\![\top]\!]\}$, *for any $\varphi\in\Phi$ where $[\![\varphi]\!]\in\{[\![p]\!],[\![\neg p]\!],[\![\bot]\!],[\![\top]\!]\}$.*

Proof The proof of the lemma is similar to the proof of Lemma 1. The dashed lines between the four diagrams in the centre of Figure 6 show how the modalities map the truth sets into each other. For example, the dashed arrow labelled with S_a from the truth set $[\![p]\!]$ to the truth set $[\![\bot]\!]$ denotes the fact that if $[\![\varphi]\!]=[\![p]\!]$, then $[\![\mathsf{S}_a\varphi]\!]=[\![\bot]\!]$. □

The proof of the next lemma is identical to the proof of Lemma 2 except that it uses Lemma 4 instead of Lemma 1.

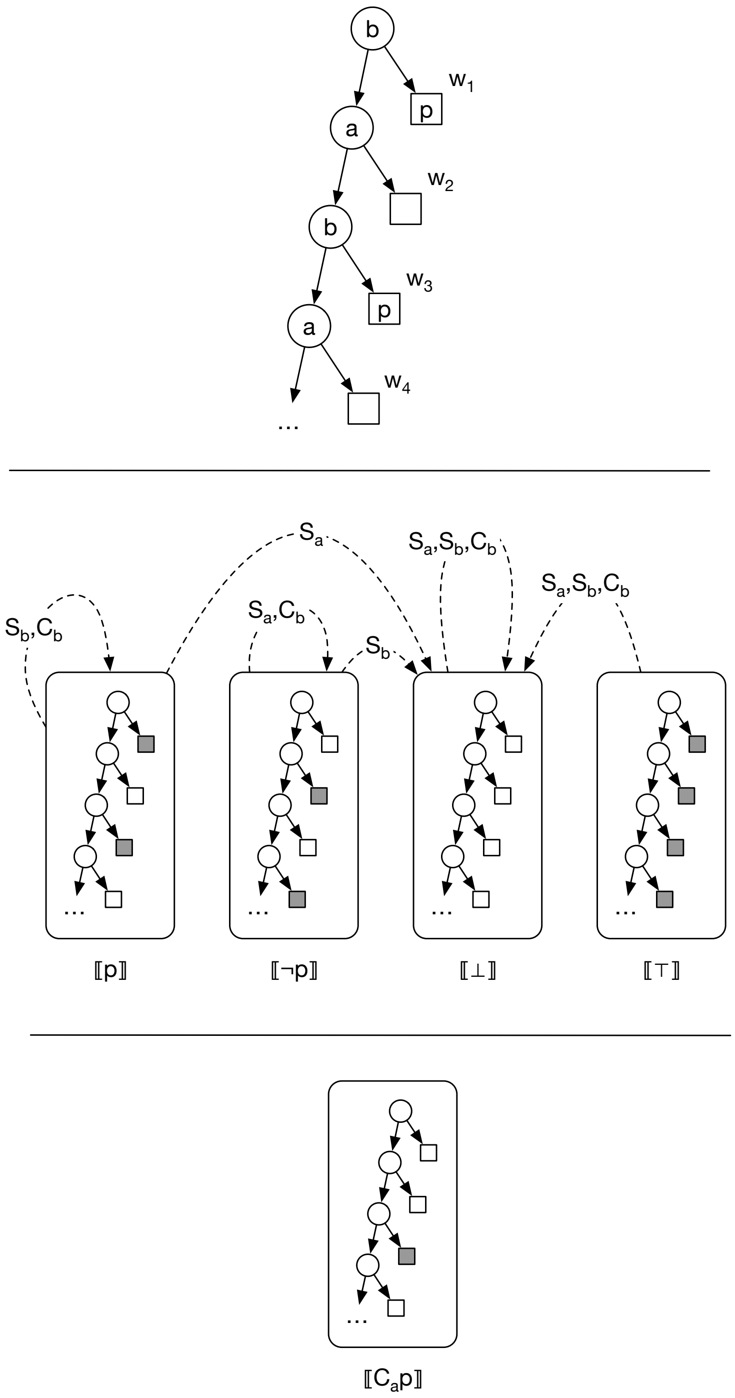

Figure 6 Towards undefinability of modality C via modality S. In the bottom diagram, the colours of leaf nodes alternate starting from the second leaf from the top.

Lemma 5 *$[\![\varphi]\!] \in \{[\![p]\!], [\![\neg p]\!], [\![\bot]\!], [\![\top]\!]\}$ for any formula $\varphi \in \Phi$ that uses only modalities S_a, S_b, and C_b.*

Lemma 6 $[\![\mathsf{C}_a p]\!] \notin \{[\![p]\!], [\![\neg p]\!], [\![\bot]\!], [\![\top]\!]\}$.

Proof It suffices to show that the truth set $[\![\mathsf{C}_a p]\!]$ is depicted at the bottom of Figure 6.

Consider any node w_i, where $i \geq 1$. By item 5 of Definition 1, for $w_i \Vdash \mathsf{C}_a p$ to be true, the following conditions should be satisfied:

A. $w_i \Vdash p$,
B. there must exist a node $n \in Ancestors(w_i)$ and a strategy $s \in Strategies_a(n)$ of agent a such that $u \nVdash p$ for each outcome $u \in Outcomes_a(n,s)$.

Note that condition (A) above is satisfied if and only if number i is odd, see the game as depicted atop of Figure 6.

Observe also that node $n = parent(w_2)$ and strategy "always go right" (into a leaf node) satisfy condition (B) for each integer $i \geq 2$. At the same time, condition (B) cannot be satisfied for $i = 1$, see the game as depicted atop of Figure 6.

Therefore, $w_i \Vdash \mathsf{C}_a p$ iff i is an odd number such that $i \geq 2$. □

The next result follows from the two lemmas above and Definition 2.

Theorem 2 (undefinability) *Formula $\mathsf{C}_a p$ is not semantically equivalent to any formula in language Φ that does not use modality C_a.*

Note that, as stated earlier, we proved slightly more than the undefinability of C through S. Namely, we have shown that C_a is not definable through any combination of modalities S_a, S_b, and C_b.

7 Future Work: Finite Games

By a finite extensive form game we mean any infinite extensive form game, as defined in Section 3, whose tree has finitely many nodes. Thus, we treat finite games as a *subclass* of infinite games.

In this article, we have shown that modalities S and C are not definable through each other for *infinite* extensive form games. The existing proofs of Theorem 1 and Theorem 2 do not apply to the subclass of *finite* extensive form games because the games depicted in Figure 5 and Figure 6 are not finite.

Natural and interesting questions for future research are if Theorem 1 and Theorem 2 hold for the class of finite extensive form games. Although we do not know an answer to either of these questions, we have partial results for each of them.

Definition 3 *Formulae $\varphi, \psi \in \Phi$ are semantically equivalent over finite games if $[\![\varphi]\!] = [\![\psi]\!]$ in each finite extensive form game.*

7.1 On definability of S via C

Theorem 1 shows that modality S is not definable through modality C. More specifically, the proof of this theorem gives an example of an infinite two-player game in which modality S_a is not definable through modalities C_a and C_b.

Although we do not know if the same can be done for finite games, we can show that, over the class of finite games, modality S_a is not definable through modality C_a only. To do this, we again use the "truth set algebra" technique.

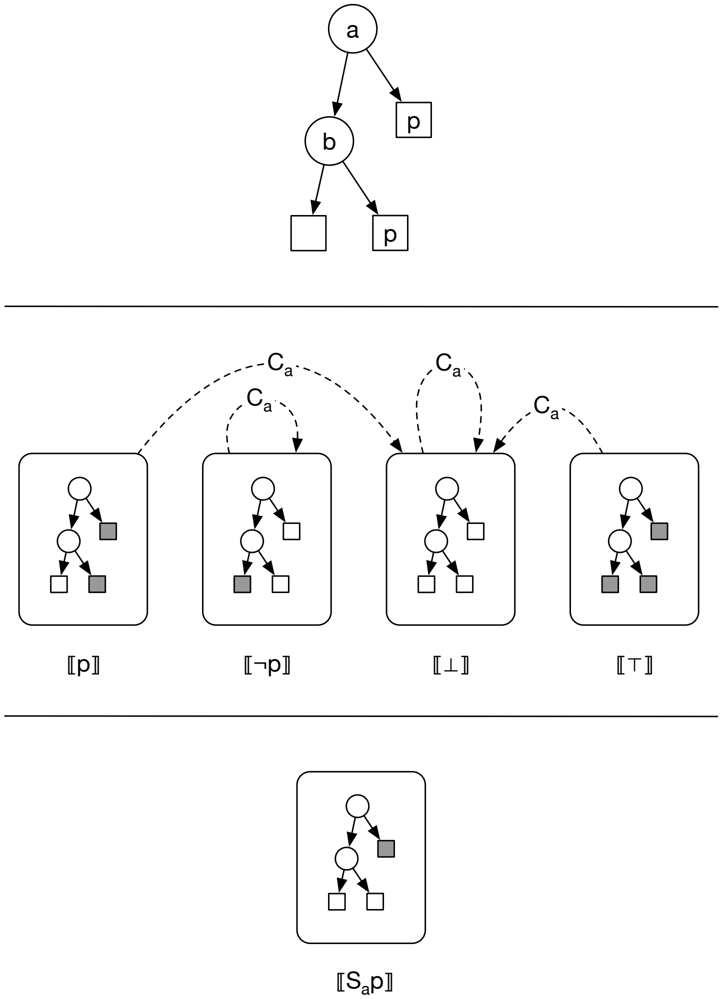

Figure 7 Towards undefinability of S_a through C_a over the class of finite games.

Consider the finite extensive form game depicted atop of Figure 7. The proofs of the following three lemmas are similar to the proof of Lemma 1, Lemma 2, and Lemma 3. The cases in the proof of Lemma 7 are shown using dashed arrows in Figure 7 similar to how it was done in Figure 5 for Lemma 1.

Lemma 7 *If* $[\![\varphi]\!] \in \{[\![p]\!], [\![\neg p]\!], [\![\bot]\!], [\![\top]\!]\}$, *then* $[\![\mathsf{C}_a\varphi]\!] \in \{[\![p]\!], [\![\neg p]\!], [\![\bot]\!], [\![\top]\!]\}$, *for any formula* $\varphi \in \Phi$.

Lemma 8 $[\![\varphi]\!] \in \{[\![p]\!], [\![\neg p]\!], [\![\bot]\!], [\![\top]\!]\}$ *for any formula* $\varphi \in \Phi$ *that uses only modality* C_a.

Lemma 9 $[\![\mathsf{S}_a p]\!] \notin \{[\![p]\!], [\![\neg p]\!], [\![\bot]\!], [\![\top]\!]\}$.

The next result follows from the two lemmas above and Definition 3.

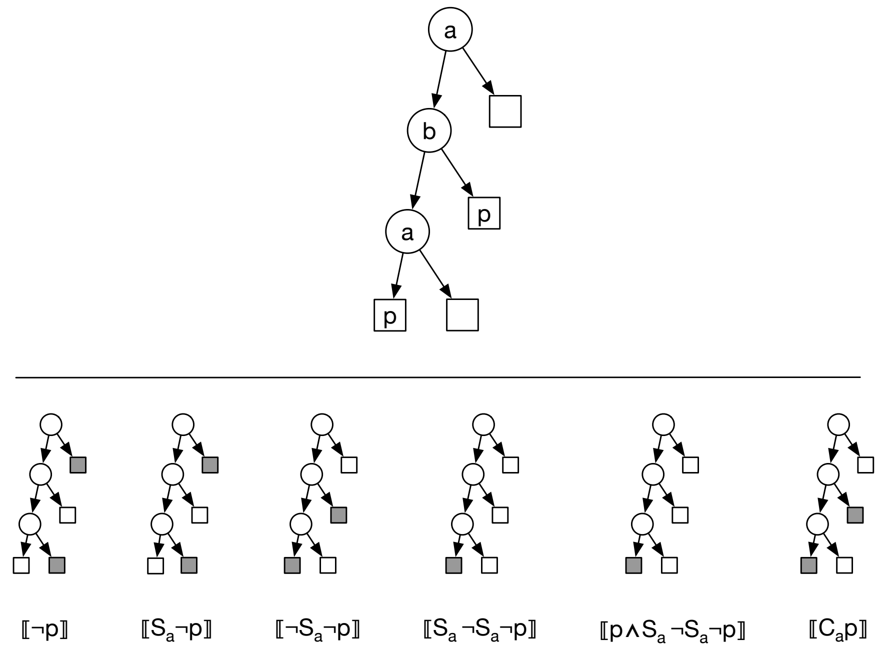

Figure 8 Counterexample for equivalence (3).

Theorem 3 (undefinability) *Formula $\mathsf{S}_a p$ is not semantically over finite games equivalent to any formula in language Φ that does not use modalities S_a, S_b, and C_b.*

7.2 On definability of C via S Although we do not know if modality C is definable through modality S over the class of finite extensive form games, we do know that equivalence (3) does *not* hold for the finite games. To observe this, it suffices to construct a single finite game and to show that the sets $[\![C_a p]\!]$ and $[\![p \wedge \mathsf{S}_a \neg \mathsf{S}_a \neg p]\!]$ are not equal for that specific game.

An example of such a game is depicted atop Figure 8. Below it, we show the computation of the truth set $[\![p \wedge \mathsf{S}_a \neg \mathsf{S}_a \neg p]\!]$ by constructing diagrams for truth sets of all subformulae of the formula $p \wedge \mathsf{S}_a \neg \mathsf{S}_a \neg p$. Finally, in the right-most position of the bottom row in Figure 8, we show the truth set $[\![C_a p]\!]$. As the diagrams show, sets $[\![C_a p]\!]$ and $[\![p \wedge \mathsf{S}_a \neg \mathsf{S}_a \neg p]\!]$ are not equal for this game.

8 Conclusion

In this article, we defined and studied counterfactual and seeing-to-it forms of responsibility in infinite extensive form games. We have shown that, unlike the case of strategic games, neither of these two forms of responsibility can be defined through the other. We have also discussed preliminary undefinability results for the class of finite extensive form games. Note that although we stated all our undefinability results in terms of arbitrary infinite games, a slightly stronger version of these results holds. Namely, it is easy to see that all games used to prove undefinability could be "folded" into *finite* graphs similar to the one depicted in Figure 1. Thus, the undefinability results hold for a more restricted class of "periodic" infinite games that can be obtained by unfolding finite graphs.

Interesting direction for future research is the axiomatisation of the interplay between modalities S and C.

References

[1] Aumann, R. J., "Mixed and behavior strategies in infinite extensive games," Technical report, Princeton University, 1961. 2

[2] Belnap, N., and M. Perloff, "Seeing to it that: a canonical form for agentives," *Theoria*, vol. 54 (1988), pp. 175–199. 5

[3] Belnap, N., and M. Perloff, "Seeing to it that: A canonical form for agentives," pp. 167–190 in *Knowledge representation and defeasible reasoning*, Springer, 1990. 3

[4] Belnap, N., and M. Perloff, "The way of the agent," *Studia Logica*, (1992), pp. 463–484. 3

[5] Capucci, M., N. Ghani, C. Kupke, J. Ledent, and F. N. Forsberg, "Infinite horizon extensive form games, coalgebraically,", in *Mathematics for Computation*, World Scientific Publishing Co. Pte Ltd, 2022. 2

[6] Frankfurt, H. G., "Alternate possibilities and moral responsibility," *The Journal of Philosophy*, vol. 66 (1969), pp. 829–839. 4

[7] Gale, D., and F. M. Stewart, "Infinite games with perfect information," *Contributions to the Theory of Games*, vol. 2 (1953), pp. 2–16. 2

[8] Horty, J., and E. Pacuit, "Action types in STIT semantics," *The Review of Symbolic Logic*, vol. 10 (2017), pp. 617–637. 3

[9] Horty, J. F., *Agency and deontic logic*, Oxford University Press, 2001. 3

[10] Horty, J. F., and N. Belnap, "The deliberative STIT: A study of action, omission, ability, and obligation," *Journal of Philosophical Logic*, vol. 24 (1995), pp. 583–644. 3

[11] Knight, S., P. Naumov, Q. Shi, and V. Suntharraj, "Truth set algebra: A new way to prove undefinability," *arXiv:2208.04422*, (2022). 7

[12] McNaughton, R., "Infinite games played on finite graphs," *Annals of Pure and Applied Logic*, vol. 65 (1993), pp. 149–184. 1

[13] Naumov, P., and J. Tao, "Blameworthiness in strategic games," in *Proceedings of Thirty-third AAAI Conference on Artificial Intelligence (AAAI-19)*, 2019. 4

[14] Naumov, P., and J. Tao, "Blameworthiness in security games," in *Proceedings of Thirty-Fourth AAAI Conference on Artificial Intelligence (AAAI-20)*, 2020. 4

[15] Naumov, P., and J. Tao, "Duty to warn in strategic games," in *Proceedings of the 19th International Conference on Autonomous Agents and Multiagent Systems, AAMAS '20, Auckland, New Zealand, May 9-13, 2020*, pp. 904–912. International Foundation for Autonomous Agents and Multiagent Systems, 2020. 4

[16] Naumov, P., and J. Tao, “An epistemic logic of blameworthiness,” *Artificial Intelligence*, vol. 283 (2020). 103269. 4

[17] Naumov, P., and J. Tao, “Two forms of responsibility in strategic games,” in *30th International Joint Conference on Artificial Intelligence (IJCAI-21)*, 2021. 5

[18] Naumov, P., and J. Tao, “Counterfactual and seeing-to-it responsibilities in strategic games,” *Annals of Pure and Applied Logic*, vol. 174 (2023), p. 103353. 5

[19] Olkhovikov, G. K., and H. Wansing, “Inference as doxastic agency. part i: The basics of justification STIT logic,” *Studia Logica*, vol. 107 (2019), pp. 167–194. 3

[20] von Wright, G. H., *Norm and Action: A Logical Enquiry*, Routledge & Kegan Paul, 1963. 5

[21] Widerker, D., *Moral responsibility and alternative possibilities: Essays on the importance of alternative possibilities*, Routledge, 2017. 4

[22] Yazdanpanah, V., M. Dastani, W. Jamroga, N. Alechina, and B. Logan, “Strategic responsibility under imperfect information,” in *Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems*, pp. 592–600. International Foundation for Autonomous Agents and Multiagent Systems, 2019. 4

Cui
School of Electronics and Computer Science
University of Southampton
United Kingdom
13932660492@163.com

Naumov
School of Electronics and Computer Science
University of Southampton
United Kingdom
pgn2@cornell.edu