Sasha Rubin <sasha.rubin@gmail.com>

AAAI

aniello murano <murano@na.infn.it> 25 October 2015 at 15:43
To: sasha <sasha.rubin@gmail.com>

Hi Sasha,

thank you for the review. Below I list all the reviews.

Nello

****

Summary of Received Reviews and Comments

Reviews superseded by other reviews are shown in the grey color in the table. All times are GMT.

date	PC member	subreviewer	Overall evaluation	Reviewer's confidence	Significance of the...	Technical Quality	Depth of Theoretical and/or...	Quality of Presentation	Breadth of Interest to the AI...
Oct 23	Alessio Lomuscio		**-2**	**2**	1	2	1	2	3
Oct 24	Simon Miles	Chris Haynes	**1**	**2**	2	2	2	2	2
Oct 25	Aniello Murano	Sasha Rubin	**-1**	**1**	2	3	2	3	3

Review 1

Paper:	657
Title:	Distant Responsibility in Multi-Agent Settings
PC member:	Alessio Lomuscio
Significance:	1: (minimal or no contribution)
Soundness:	2: (minor inconsistencies or small fixable errors)
Scholarship:	1: (important related work missing or mischaracterizes prior research)
Clarity:	2: (more or less readable)
Breadth of Interest:	3: (some interest beyond specialty area)
Summary Rating:	**-2**: (--)
Confidence:	**2**: (reasonably confident)

Summarize the Main Contribution of the Paper:

The paper puts forward an approach to formally represent "group responsibility" by means of ATL.

The paper contribution, as I understand it, is effectively the claim that ATL (Alur et al) in its original form can properly formalize the concept of "direct responsibility". Alur et al's readings of their modalities are considerably more modest and simply refer to processes being able or unable to enforce a temporal state of affairs.

The authors' starting point, as I understand it, is that if a group of agents can avoid something (a temporal statement) to take place through collective action, then they are "directly responsible" for it, if it takes place. The notion is further enriched by considering minimality and other requirements that appear to follow from this.

Comments for the Authors:

I have several concerns about this representation as it seems to me to be too simple, nearly naive, to capture any realistic notion of responsibility. This is at several levels - I am only listing some below.

* It is so weak that the concept of responsibility it aims to model seems to be uninteresting. Suppose, for simplicity someone kills a pedestrian with his car. It's not his fault as the pedestrian didn't see the car and jumped under it. Would we say the driver is responsible for the death? If I take a simplified model capturing what above, one can determine that the driver could have avoided the death only by choosing not to drive the car on that day. I do not recognize this concept of responsibility in ethical, legal or common-sense of the word.

I also find it hard to reason about responsibility without accounting for other states of affairs in the model. In the example of the political parties from the paper, it really seems too simple to say that since the party did not block the legislation they are responsible for it. That may seem reasonable if they could have vetoed it. But the model is not sensitive to what action is required. Instead this seems essential. What if to block the legislation the only viable action was to kill the speaker? Would they be responsible because they didn't? And if the only action that would block the legislation were to blow up the whole of the parliament? Would they be responsible for the legislation if they didn't? It's easy to see one can create rather natural examples showing our common understanding of "responsibility" goes well beyond not acting when an action would prevent a certain state of affairs.

* At the same time the plain ATL formalization also seems too strong.

The agents in a coalition may have a way of avoiding a certain sequence of states of affairs but they might not know how to do it. The formalism does not account for this.

Even the agents in the coalition do want to avoid a certain event and know how to do it, doing so may require coordination and cooperation so that the correct individual actions are selected. See the attacking prisoners' dilemma for example.

Note that neither of these is a problem in the normal readings of ATL, as these concern the plain statement of being able to enforce/avoid a certain states of affairs. But by using a very loaded term such as "responsible" (one that all ATL's literature avoids as far as I know), then one would expect to be able to account for the ability, intentionality, and the opportunity to carry out actions.

To summarize, I believe that concepts of interest here are considerably more subtle than how they are described here. They are closely related to the notions of agency, bringing it about, and power that are discussed widely in the philosophical logic literature. Of course the logic STIT (Belnap and colleagues) is an also an attempt to deal with these issues. The treatment of STIT is very different from the one here presented. The conclusions that would be drawn are different, radically the opposite in many cases. It may even be that the current proposal is a better approximation of these concepts, but a discussion and a comparison is required to establish exactly what the contribution is, however limited. In doing so, I would also link and compare the current approach to work by Porn on "bringing it about".

In addition to the formalization in ATL; the rest of the paper is rather light on technicalities as they mostly refer to minimality or considerations on subsets.

My suggestion to the authors is to clarify the notions above perhaps by restricting their applicability to scenarios where these and other issues are not problematic (closed-world assumptions, perhaps?). In any follow-up paper these points could be further discussed and some proofs could be omitted. Many of them are routine and do not seem to add much to the paper. An in-depth comparison and a more deep and detailed exploration of these issues is, in my opinion, required.

Confidential Comments to Program Committee:	
Nominate for a Best Paper Award:	
Time:	Oct 23, 16:17

Review 2

PC member:	Simon Miles
Reviewer:	Chris Haynes <christopher.haynes@kcl.ac.uk>
Significance:	2: (modest or incremental contribution)
Soundness:	2: (minor inconsistencies or small fixable errors)
Scholarship:	2: (relevant literature cited but could expand)

Clarity:	2: (more or less readable)
Breadth of Interest:	2: (interest limited to specialty area)
Summary Rating:	**1**: (+ (weak accept))
Confidence:	**2**: (reasonably confident)

Summarize the Main Contribution of the Paper:

The authors propose a concept of 'distant responsibility' that grades groups of agents with respect to the amount of responsibility they have for bringing about a certain state of affairs. This degree of responsibility is based on how many steps they have to take in order to prevent the state of affairs.
They distinguish between preventing a state of affairs once and maintaining this prevention.

Comments for the Authors:

The concept seems sound, but I don't feel the paper makes clear the motivation for it, nor how this would be used in a system.

It seems a little odd not to distinguish between action and inaction, when talking of responsibility. Perhaps not in the furnace example, but in the example of the undesired bill in the introduction, surely the group actively pushing for the bill is responsible to a higher degree than a group that allowed it to happen by inaction?

Basing degree on number of decision steps (or actions), rather than time or cost (i.e. effort) seems to limit practical application, especially if one is going to use the notion of responsibility by distance to invest resources (as suggested by the authors).
The initial definition refers to "collective decision steps". Does this mean the agents have to collectively commit to a certain decision of action?

The two sentences with the citations in the first paragraph seem to be saying the same thing; I don't think both are necessary. Or, at least you don't need the citations in the first of the sentences.

Typos/Minor Issues:

“These approaches, however, does not” should be “do not”.
“(and a1a2a3)” does not seem to make sense. Should this be “(and a1a2 and a2a3)”?
“two member groups have larger share” should be “have a larger share”.

The caption to Figure 1, mentions variable i. Is this the subscript to q? It's not clear.
In Proof of Proposition 1, “all sates” should be “all states”.
In Distant Group Responsibility first paragraph - “state of affairs form” should be “state of affairs from”

“due to minimally concerns” seems wrong. Should it be “due to minimality concerns”?
“We believe that in neither of these cases” - should that be “either”?
“intuition that when the a state” - remove the “the” or the “a”

In Lemma 1, why use k instead of |N|? I think the latter would be clearer for the reader, who might have forgotten that |N| = k. You remind us in the subsequent proof, but I don't see the benefit of using k here.

“For collaborative state” should be either “For a collaborative state” or “For the collaborative state” “gradation of responsibility on a the degree” - remove the “a”

Confidential Comments to Program Committee:	
Nominate for a Best Paper Award:	
Time:	Oct 24, 11:13

Review 3

PC member:	Aniello Murano
Reviewer:	Sasha Rubin <sasha.rubin@gmail.com>
Significance:	2: (modest or incremental contribution)
Soundness:	3: (correct)
Scholarship:	2: (relevant literature cited but could expand)
Clarity:	3: (crystal clear)
Breadth of Interest:	3: (some interest beyond specialty area)
Summary Rating:	**-1**: (- (weak reject))
Confidence:	**1**: (educated guess)

Summarize the Main Contribution of the Paper:

This paper is about assigning degrees of responsibilities to groups of agents in a multi-agent setting.

More precisely, the goal (of a given concurrent game structure with N agents) is to reach a given set of global states S in a given number of steps d.

The main contribution is a definition (Dfn 6) of the degree to which a group C of agents is responsible for achieving (or maintaining) the goal. In the special case that all N agents are required to achieve the goal, the degree of responsibility of a group C turns out to be |C|/|N|.

The paper supplies some easy consequences of the main definition.

The authors motivate their work in a few ways:
1. as extending existing work by Bulling and Dastani (2013) from single-agent responsibility to group responsibility.
2. as being able to model responsibility and blame in political discourse, or to model situations where managers must decided where to invest resources to achieve a particular goal.

Comments for the Authors:

The paper is very well written (modulo some minor typos).

The scope of the paper is certainly within AAAI.

The related work should be expanded, especially the relationship with Chockler and Halpern (2004) that also studies a quantitative notion of responsibility (there is only one sentence describing the relationship, it is on page 6, and this sentence is neither clear nor very informative).

My main concern is that the definition, which is the main contribution of the paper, is ad hoc. In more detail:

1. The only formal justification for the main definition (Dfn 6) is that more members can't decrease the responsibility of a group (Prop 7).

This is certainly not enough to justify the definition (any number of definitions would satisfy Prop 7). The authors would do well to state other natural properties that a notion of responsibility should satisfy, and prove their definition satisfies these properties (or find an alternative definition that does).

2. Even in the special case that one needs all N agents to achieve the goal, the authors only give two properties (Prop 9), i.e., that their definition of responsibility is additive and scalable, but do not argue why these are desirable/natural properties of responsibility.

Minor comments:
- "The emergence of autonomous agents and multi-agent systems
requires formal models to represent and reason about
the responsibility of agents and agent groups for the outcome
of their actions" (pg 1).

A citation here is important to justify the very strong claim that MAS "requires" reasoning about responsibility.

- The sentence "this would be in correspondence" (pg 3) is vague. What is the relationship with the cited work?

- "of affairs form" --> "of affairs from"

- "a the"

- in Lem 1 it might be clearer to replace k by |N|.

- Sometimes the authors write that the goal is to achieve and sometimes that it is to preclude. It would be clearer if one of these was chosen and used consistently.

Confidential Comments to Program Committee:	
Nominate for a Best Paper Award:	
Time:	Oct 25, 14:35

[Quoted text hidden]

[Quoted text hidden]