COLLECTIVE OBLIGATIONS, GROUP PLANS AND INDIVIDUAL ACTIONS

ALLARD TAMMINGA*, HEIN DUIJF†

Abstract: If group members aim to fulfil a collective obligation, they must act in such a way that the composition of their individual actions amounts to a group action that fulfils the collective obligation. We study a strong sense of joint action in which the members of a group design and then publicly adopt a group plan that coordinates the individual actions of the group members. We characterize the conditions under which a group plan successfully coordinates the group members' individual actions, and study how the public adoption of a plan changes the context in which individual agents make a decision about what to do.

Keywords: collective obligation, team reasoning, collective intentionality, cooperation, joint action

Of the major disciplines concerned with social behaviour, game theory alone, it seems, has eschewed, and got along to date without, group notions.

Michael Bacharach (2006: 72)

1. INTRODUCTION

If group members aim to fulfil a collective obligation, they must act in such a way that the composition of their individual actions amounts to a group action that fulfils the collective obligation. We use a game-theoretical formalism to study a strong sense of joint action in which the members of a group, using team reasoning, design and then publicly adopt a group plan. By highlighting particular group actions, a group

* University of Groningen, Oude Boteringestraat 52, 9712 GL Groningen, the Netherlands and Utrecht University, Janskerkhof 13, 3512 BL Utrecht, the Netherlands. Email: a.m.tamminga@rug.nl. URL: http://www.rug.nl/staff/a.m.tamminga/.
† Utrecht University, Janskerkhof 13, 3512 BL Utrecht, the Netherlands. Email: h.w.a.duijf@uu.nl. URL: http://www.uu.nl/staff/hwaduijf/.

187

plan specifies the individual actions that are the components of these highlighted group actions and thus specifies for every group member what she ought to do to contribute to the group's fulfilling its collective obligation. The public adoption of a group plan changes the context in which group members and other agents make a decision about what to do.[1] We give necessary and sufficient conditions under which a group plan that is designed to fulfil a collective obligation is a *good plan*, that is, it successfully coordinates the individual actions of the group members. Our central theorem concerns what happens if such a good plan is publicly adopted. We show that if the group members publicly adopt a good group plan, then the decision context is changed in such a way that for every group member it holds that she acts according to the group plan if and only if she performs an action that is one of the best things she can do in the changed decision context, regardless of the actions taken by all the others.

We illustrate our game-theoretical formalism with concepts and ideas from the philosophical and economics literature on deontic logic, team reasoning, collective intentionality and joint action. This serves two purposes. On the one hand, the concepts and ideas from philosophy and economics help to clarify what our formalism is meant to model and thereby provide a partial conceptual justification for our formal analysis of the relation between collective obligations, group plans and individual actions. On the other hand, we submit that our formalization captures important aspects of concepts and ideas that have been developed in the philosophical and economics literature and that therefore our formal results are of central relevance to the debate on team reasoning, collective intentionality and joint action. To illustrate this, we briefly discuss the relation between collective and individual rationality and the relation between collective and individual blameworthiness from the perspective of our formalism. We do not, however, engage in a prolonged philosophical discussion with the main protagonists in the debate, since our current purpose is to present our game-theoretical formalism and our results on the relation between collective obligations, group plans and individual actions.

The paper is set out as follows. In Section 2, we define individual and collective obligations in terms of deontically optimal actions in a (single-shot) coordination game involving a single deontic ideality function. An individual agent is said to fulfil her *individual obligation* if and only if she

---

[1] Van Hees and Roy (2008) and Roy (2009a, 2009b) present game-theoretical studies of how individual intentions might change the decision context for individual agents. Although their decision contexts include the intentions and actions of other agents, their approach differs from ours in that team reasoning, collective intentions and group actions play no role in their analyses.

performs one of her deontically optimal individual actions. Likewise, a group fulfils its *collective obligation* if and only if it performs one of its deontically optimal group actions. (For readability, we drop the adverb 'deontically' in 'deontically optimal'.) We show that the fulfilment of a collective obligation is neither necessary nor sufficient for the fulfilment of individual obligations. In Section 3, we define a *group plan* as a set of group actions that are available to the group as a whole. Given such a group plan, a group member fulfils her *member obligation* specified by the plan if and only if she performs an action that is her component action of one of the group actions in the plan. We model the public adoption of a group plan as an update of the deontic ideality of the action profiles in a coordination game. In Section 4, we give necessary and sufficient conditions under which a group plan is a *good plan*, that is, conditions under which a group plan guarantees that if every group member fulfils her member obligation specified by the plan, then the group itself fulfils its collective obligation. Our central theorem is proved in Section 4.3: if a coordination game is updated with a good plan, then for every individual group member it holds that she fulfils her member obligation specified by the plan if and only if she fulfils her individual obligation in the coordination game that results from updating the original coordination game with the plan. Consequently, if a group publicly adopts a good plan, then a collective failure to fulfil a collective obligation is always due to an individual failure to fulfil a member obligation specified by the plan, or equivalently, an individual failure to fulfil an individual obligation in the coordination game that results from updating the original coordination game with that plan. A brief discussion of the relation between collective and individual rationality follows. In Section 5, we show that almost all our findings on coordination games transfer to what we call 'zero-preserving cooperation games' involving group-relative utility functions. On the basis of our discussion of cooperation games we then compare our approach to cooperation with the team-reasoning account of cooperation. In the concluding section, we use our formal analysis of collective obligations, group plans and individual actions to study some logical aspects of backward-looking collective moral responsibility.

2. INDIVIDUAL AND COLLECTIVE OBLIGATIONS IN COORDINATION GAMES

We use *coordination games* to study relations between collective obligations, group plans and obligations of individual agents.[2] (In Section 5, we investigate these relations using 'cooperation games'.)

---

[2] Coordination games are similar to Schelling's (1960: 84) *pure-collaboration games* and Bacharach's (2006: 122) *coordination contexts*.

A coordination game involves a finite set $N = \{i_1, \ldots, i_n\}$ of individual agents. Each individual agent i in N has a non-empty and finite set A_i of available individual actions. We use a_i and a_i' as variables for individual actions in the set A_i. The Cartesian product $\times_{i \in N} A_i$ of all the individual agents' sets of actions gives the full set A of action profiles. We use a and a' as variables for action profiles in the set A.[3] A deontic ideality function d assigns to each action profile a in A a value $d(a)$ that is either 1 (if a is deontically ideal) or 0 (if a is not deontically ideal):[4]

Definition 1 (Coordination Game). *A coordination game S is a triple $\langle N, (A_i), d \rangle$, where N is a finite set of individual agents, for each agent i in N it holds that A_i is a non-empty and finite set of actions available to agent i, and d is a deontic ideality function that assigns to each action profile a in $A (= \times_{i \in N} A_i)$ a value $d(a) \in \{0, 1\}$.*

To prove some of our claims about collective obligations, group plans and obligations of individual agents, we must rule out coordination games in which no action profile is deontically ideal. Such coordination games are *flat*:

Definition 2. *Let $S = \langle N, (A_i), d \rangle$ be a coordination game. Then S is* flat *if and only if for all action profiles $a \in A$ it holds that $d(a) = 0$.*

For each group $\mathcal{G} \subseteq N$ we define the set $A_{\mathcal{G}}$ of group actions that are available to group $\mathcal{G}$ as the Cartesian product $\times_{i \in \mathcal{G}} A_i$ of all the individual group members' sets of actions. We use $a_{\mathcal{G}}$ and $a_{\mathcal{G}}'$ as variables for group actions in the set $A_{\mathcal{G}} (= \times_{i \in \mathcal{G}} A_i)$. Moreover, if $a_{\mathcal{G}}$ is a group action of group $\mathcal{G}$ and if $\mathcal{F} \subseteq \mathcal{G}$, we use $a_{\mathcal{F}}$ to denote the subgroup action that is $\mathcal{F}$'s component subgroup action of the group action $a_{\mathcal{G}}$. Finally, we use $-\mathcal{G}$ to denote the relative complement $N - \mathcal{G}$.

We follow John Horty's (1996, 2001) deontic logic of agency and analyse a group's collective obligations in terms of its optimal group actions that are in turn defined using a dominance ordering of the group actions that are available to the group. Intuitively, group action $a_{\mathcal{G}}$ *weakly dominates* group action $a_{\mathcal{G}}'$ in coordination game S (notation: $a_{\mathcal{G}} \succeq_S a_{\mathcal{G}}'$) if and only if $a_{\mathcal{G}}$ promotes deontic ideality in S at least as well as $a_{\mathcal{G}}'$, regardless of what the group's non-members do:

---

[3] We adopt the notational conventions of Osborne and Rubinstein (1994: sec. 1.7) and omit braces if the omission does not give rise to ambiguities.

[4] We take deontically ideal action profiles to represent a single moral code, similar to deontically ideal worlds in the possible-worlds semantics for standard deontic logic (Hilpinen 1971: 13–15). Our binary ordering of the action profiles in terms of deontic ideality can also be taken to reflect a simple preference ordering of agents who classify action profiles unanimously as 'good' or 'bad'.

	a_j	a'_j	a''_j
a_i	1	1	1
a'_i	1	1	0
a''_i	1	0	0

FIGURE 1. Coordination game S_1

Definition 3 (Dominance). *Let $S = \langle N, (A_i), d \rangle$ be a coordination game. Let $\mathcal{G} \subseteq N$ be a group of agents. Let $a_{\mathcal{G}}, a'_{\mathcal{G}} \in A_{\mathcal{G}}$ be actions available to group $\mathcal{G}$. Then*

$$a_{\mathcal{G}} \succeq_S a'_{\mathcal{G}} \quad \text{iff} \quad \text{for all } a''_{-\mathcal{G}} \in A_{-\mathcal{G}} \text{ it holds that } d(a_{\mathcal{G}}, a''_{-\mathcal{G}}) \geq d(a'_{\mathcal{G}}, a''_{-\mathcal{G}}).$$

Strong dominance is defined in terms of weak dominance: $a_{\mathcal{G}} \succ_S a'_{\mathcal{G}}$ if and only if $a_{\mathcal{G}} \succeq_S a'_{\mathcal{G}}$ and $a'_{\mathcal{G}} \not\succeq_S a_{\mathcal{G}}$.

We define the set of *optimal group actions* that are available to a group $\mathcal{G}$ of agents in a coordination game S in terms of the dominance ordering of $A_{\mathcal{G}}$. A group action $a_{\mathcal{G}}$ in $A_{\mathcal{G}}$ is optimal if and only if it is not strongly dominated by any group action in $A_{\mathcal{G}}$:

Definition 4 (Optimal Actions). *Let $S = \langle N, (A_i), d \rangle$ be a coordination game. Let $\mathcal{G} \subseteq N$ be a group of agents. Then the set of $\mathcal{G}$'s optimal actions in S, denoted by $\text{Optimal}_S(\mathcal{G})$, is given by*

$$\text{Optimal}_S(\mathcal{G}) = \{a_{\mathcal{G}} \in A_{\mathcal{G}} : \text{there is no } a'_{\mathcal{G}} \in A_{\mathcal{G}} \text{ such that } a'_{\mathcal{G}} \succ_S a_{\mathcal{G}}\}.$$

Note that this definition implies that for every non-empty group of agents there is at least one optimal action (see Lemma 1 in the Appendix). Because there are optimal actions for any non-empty group and any of its non-empty subgroups, it makes sense to enquire about the relation between optimal group actions and optimal subgroup actions. First, note that an optimal group action is not necessarily composed of optimal subgroup actions. In the coordination game S_1 of Figure 1 it holds that (a'_i, a'_j) is optimal for the group $\{i, j\}$, although a'_i is not optimal for agent i and a'_j is not optimal for agent j.

Secondly, the composition of optimal subgroup actions is not necessarily an optimal group action. In the coordination game S_2 of Figure 2 it holds that a_i is optimal for agent i and a'_j is optimal for agent j, although (a_i, a'_j) is not optimal for the group $\{i, j\}$.

Finally, we say that a group $\mathcal{G}$ fulfils its *collective obligation* in coordination game S if and only if it performs one of its optimal group actions. (Likewise, an agent i fulfils her *individual obligation* in coordination game S if and only if she performs one of her optimal individual actions.) Our discussion of Figures 1 and 2 shows that collective obligations and individual obligations do not match: the

	a_j	a'_j
a_i	1	0
a'_i	0	1

FIGURE 2. Coordination game S_2

fulfilment of a collective obligation is neither necessary nor sufficient for the fulfilment of individual obligations. To fulfil its collective obligation the group members must ensure that the group action composed of the individual actions performed by them is an optimal group action. The gap between collective obligations and individual agency is filled by *member obligations*. Intuitively, a member obligation is what a group member ought to do in order to help ensure that the group fulfils its collective obligation, that is, in order to help ensure that the group performs an optimal group action. In this paper we consider member obligations as following from a collective obligation via a group plan.

3. GROUP PLANS AND MEMBER OBLIGATIONS

A group that aims to fulfil a collective obligation has a coordination problem: the group members must cooperate to ensure that the composition of their individual actions amounts to an optimal group action. In contrast to most coordination theorists, we assume that group members are able to communicate and to agree on a *group plan* to solve their coordination problem.[5] By agreeing on a group plan, it becomes common knowledge among the members of the group that the plan has been adopted. A group plan, once agreed upon, coordinates the individual actions of the members of the group: it allows them to act *simultaneously* and *unconditionally*, in the full belief that every group member acts according to plan. Group actions that are regulated by a group plan constitute a strong sense of joint action.[6] Raimo Tuomela calls this sense of acting together *proper joint action*:

> Acting together involves sociality in the relatively strong sense that such action must be based on joint intention or shared collective goal. This makes

---

[5] Coordination theorists typically address the following problem: given a one-shot two-player game where no communication between the agents is possible, how are we to explain an agent's ability to anticipate the other agent's actions? Schelling (1960: 54–58), Lewis (1969), Gauthier (1975), Sugden (1993, 1995, 2003) and Bacharach (2006) all study this type of coordination problem, arguing that concepts like *salience*, *focal points* or *framing* are the key to understanding the coordination abilities of non-communicating agents.

[6] Weaker senses of joint action are explored by, for instance, Jackson (1987), Kutz (2000) and Chant (2007).

any case of acting together cooperative at least to the extent that the persons are collectively committed to making true a certain state of affairs. . . . In the strongest sense of acting together we require acting on a joint, agreed-upon plan. This I will call proper joint action. (Tuomela 2000: 7)

To design a group plan the group members must discover which are the best courses of action they can take *as a group*. To do so, group members must abandon their individual perspective and reason from the standpoint of a group and the group actions that are available to it. This kind of reasoning is known as *team reasoning* (Sugden 1993, 2000, 2003; Bacharach 1999, 2006) or *we-reasoning* (Tuomela 2013).[7] We take it that team reasoning permits the group members to come up with and agree on a group plan.[8] Once adopted, a group plan provides 'a filter on options that are potential solutions' (Bratman 1987: 35) to a group's coordination problem. By filtering out group actions, a group plan specifies the remaining group actions and in consequence it highlights the individual actions that are the components of these highlighted group actions. Accordingly, a group plan specifies for each group member what she ought to do.

We intend to understand how a group plan gives rise to member obligations by way of Margaret Gilbert's account of joint commitments. We submit that by adopting a group plan, a group of agents enters a joint commitment to execute the plan.[9] After entering such a joint commitment, the group members 'owe each other conformity to the commitment. Thus, they have obligations towards each other' (Gilbert 2006a: 156). Such obligations of joint commitment are genuine obligations, Gilbert argues: they meet the condition that 'one who has an obligation to perform some action will have sufficient reason for performing it, sufficient reason that is independent of his own inclinations or self-interest and that cannot be eradicated by his own fiat' (Gilbert 2006a: 157). Nonetheless, obligations of joint commitment are social rather than moral obligations.

The relation between a joint commitment and its concomitant social obligations is, Gilbert claims, *a priori*: the mere fact of entering a joint

---

[7] Bacharach (2006: 121) writes: 'Roughly, somebody 'team-reasons' if she *works out the best feasible combination of actions for all the members of her team, then does her part in it*.' Bacharach and Sugden assume that team reasoning leads to a unique group action. In the present setting we do not need this methodological assumption. In Section 5.1 we compare our account of cooperation with Bacharach's and Sugden's team-reasoning account.

[8] Similarly, Gold and Sugden (2007: 126) argue that 'it is natural to regard the intentions that result from team reasoning as collective intentions.'

[9] Compare Tuomela (2000: 8): 'In general, the performance of a joint action can be regarded as agreement-based if the plan has been accepted by the participants and if they have communicated their acceptances appropriately to the others so that a joint commitment to perform the joint action has come about.'

commitment implies social obligations on the members of the group.[10] They 'owe each other actions by virtue of their participation in the joint commitment *and that alone*. Consideration of what bad consequences might flow from violation of the commitment, for instance, is not relevant to the issue' (Gilbert 1999: 151). A publicly adopted plan thus generates member obligations, regardless of whether or not the group members correctly assume that it is a good plan.[11] Robert Sugden concurs:

> To act as a member of the team is to act as a *component* of the team. It is to act on a concerted plan, doing one's allotted part in that plan without asking whether, taking other members' actions as given, one's own action is contributing towards the team's objective....It must be sufficient for each member of the team that the plan itself is designed to achieve the team's objective: the objective will be achieved if everyone follows the plan. (Sugden 1993: 86)[12]

We suggest that a group plan P_G of a group G in coordination game S be thought of as any subset of the set A_G of group actions that are available to G in S. By adopting a group plan, the group members enter a joint commitment to perform one of the group actions in the plan. Given an adopted group plan P_G and an individual group member i of G, agent i fulfils her member obligation specified by the plan P_G if and only if she performs an action that is her component action of one of the group actions in the plan, that is, if and only if she performs an action from the set $\{a_i : a_G \in P_G\}$.

3.1. Updating coordination games by group plans

We have argued that the adoption of a group plan generates member obligations and that these member obligations are specified by the group

---

[10] See also Tuomela (2005: 345): 'If agreement making is in question, there will also be a publicly existing social (or, if you like, '*quasi*-moral') obligation to participate in joint action. This entailment of an obligation can be regarded as a conceptual truth about the notion of agreement'. Bacharach (2006: 64 – notation modified) is not so sure about the a priori character of this entailment relation: '[E]ven granted that an agent should decide that a certain profile *a* should be realized, why does this give her a reason to do her part in *a*?' Scanlon (1998: 317) argues against its a prioricity: 'If a convention or social practice is taken to consist in the fact that people accept certain rules or norms and typically act in accordance with them, then we need a mediating moral principle to explain how such practices can be morally binding and generate specific obligations.'

[11] See again Tuomela (2000: 207): 'If some persons make an agreement to cooperate, that entails the obligation to cooperate, be cooperation rational or not. Accepting such an obligating agreement in a full sense entails for the participants a collective commitment to fulfill it.'

[12] Compare Sugden (1993: 72): 'A cooperative morality enjoins each individual to *do her part* in achieving outcomes that are good for all. ... [T]he individual does not ask whether her own actions, considered in isolation, yield preferred outcomes.'

plan. There is, however, a second way in which the adoption of a group plan generates obligations. The adoption of a group plan can also be understood as inducing a change of the deontic ideality of the action profiles in a coordination game.[13] Because a change in deontic ideality might affect the dominance ordering of an agent's set of actions, such a change might lead to a different set of actions with which an agent, whether individual or group, might fulfil her individual or its collective obligation.

Technically, we model the adoption of a group plan in coordination game S as an *update* of the deontic ideality function of the game's action profiles. We use $S \uparrow P_G$ to denote the coordination game that results from updating coordination game S with a group plan P_G. To keep the update of the deontic ideality function technically manageable, we require that the group plan be adopted *publicly*. We may hence assume that after the public adoption of a group plan it is *common knowledge* among the agents in the game, whether or not they are members of the group that adopts the plan, that the plan has been adopted. The public adoption of a group plan hence changes the decision context not only for group members, but also for agents who are not in the group. After the public adoption of a group plan each agent in the outgroup decides what to do, on the supposition that the group members act according to plan. To implement this in a simple way, we consider the update of the deontic ideality function to be uniform for all agents in the game.[14] Accordingly, updating a game with a group plan amounts to this: if an action profile is forbidden by the plan, its deontic ideality becomes 0; and if an action profile is permitted by the plan, its deontic ideality remains unchanged. Or equivalently, an action profile a is deontically ideal in $S \uparrow P_G$ if and only if (1) the action profile a is deontically ideal in S; and (2) the group's component action a_G of the action profile a is a group action in P_G:

Definition 5 (Plan Updates). *Let $S = \langle N, (A_i), d \rangle$ be a coordination game. Let $G \subseteq N$ and let $P_G \subseteq A_G$. Then the update of S with P_G, notation: $S \uparrow P_G$, is the coordination game $S \uparrow P_G = \langle N, (A_i), d^\uparrow \rangle$, where*

$$d^\uparrow(a) = \begin{cases} d(a), & \text{if } a_G \in P_G \\ 0, & \text{if } a_G \notin P_G. \end{cases}$$

---

[13] Tuomela (2000: 210) makes a similar observation: 'To be sure agreement making can well change more in the game than has been assumed above: Due to its institutional and *quasi-moral* character it can change the payoffs of defection in a way that changes the whole nature of the game.'

[14] These idealizing assumptions are similar to those that underlie the logic of public announcements (Plaza 1989). They were relaxed in the logic of epistemic actions (Baltag *et al.* 1998). Van Ditmarsch *et al.* (2007: chs 4 and 5) provide textbook presentations of these logics.

Because an update of a coordination game with a group plan might change the deontic ideality of the game's action profiles, it might also change the dominance ordering of any group's group actions. There is one exception. Updating a coordination game with a group plan leaves the dominance ordering of the group actions in the plan unaffected (all claims are proved in the Appendix):

Observation 1. *Let $S = \langle N, (A_i), d \rangle$ be a coordination game. Let $\mathcal{G} \subseteq N$ and let $P_\mathcal{G} \subseteq A_\mathcal{G}$. Let $a_\mathcal{G}, a_\mathcal{G}^* \in P_\mathcal{G}$. Then*

$$a_\mathcal{G} \succeq_S a_\mathcal{G}^* \quad \textit{iff} \quad a_\mathcal{G} \succeq_{S \uparrow P_\mathcal{G}} a_\mathcal{G}^*.$$

Although an update with a group plan does not change the dominance ordering of the group actions in the plan, it might change the dominance ordering of the individual actions that are component actions of the group actions in the plan. Ideally, an update with a group plan should change the dominance ordering of these individual actions in such a way that an individual action is a component of one of the group actions in the plan if and only if that individual action is an optimal individual action in the coordination game that results from updating the original coordination game with the plan. Not all group plans have this updating property. We show that if a group plan is a good plan (in the sense defined below), then it has this updating property.

4. GOOD PLANS AND BAD PLANS

There are good plans and bad plans. A good plan guarantees that if every group member fulfils her member obligation specified by the plan, then the group itself fulfils its collective obligation.[15] Or equivalently, a good plan guarantees that if the group itself does *not* fulfil its collective obligation, then at least one group member does *not* fulfil her member obligation specified by the plan.[16] A bad plan fails to do so. By agreeing on a bad plan, even though it thereby becomes common knowledge that the plan has been adopted, the group members do not rule out the possibility that every group member fulfils her member obligation specified by the plan, while the composition of the group members'

---

[15] Compare Regan ([1980](#): 138 – notation adapted): 'If the members of $\mathcal{G}$ all do their part in the best pattern of behaviour for the members of $\mathcal{G}$ given the behaviour of non-members, it is clear that the members of $\mathcal{G}$ produce the best consequences possible as a group.' We argue below that the truth of this conditional depends on the structure of the plan: if the plan is what we call 'optimal' and 'interchangeable', then the conditional is true.

[16] Bratman ([2014](#): 34) claims that in basic cases 'violations of these norms of social rationality will be constituted by violations, by one or more participants, of associated norms of individual planning agency.' See also Jackson ([1987](#): 107), who agrees with us that the conditional 'If a group act is wrong, at least one of its constituent individual acts is wrong' is not a truth of logic.

actions does not amount to a group action that fulfils their collective obligation. Accordingly, there are two conditions on fulfilling a collective obligation by way of a group plan: first, the group members must act according to plan, and secondly, the plan itself must be good. But what makes a group plan a good plan? We submit that a plan is a good plan if and only if it is *optimal* and *interchangeable*.

4.1. Optimal plans

The aim of a group plan is to fulfil a collective obligation. A group fulfils its collective obligation if and only if it performs an optimal group action. A good plan should hence be designed such that if all group members act according to plan, then the composition of their actions amounts to an optimal group action. The first design requirement of a good plan therefore is that the plan only consists of optimal group actions (as we shall see below, this is a necessary but not a sufficient condition). We say that a group plan is *optimal* if and only if it is a non-empty subset of the group's optimal group actions:

Definition 6 (Optimal Plans). *Let $S = \langle N, (A_i), d \rangle$ be a coordination game. Let $\mathcal{G} \subseteq N$ and let $P_{\mathcal{G}} \subseteq A_{\mathcal{G}}$. Then $P_{\mathcal{G}}$ is* optimal *if and only if $P_{\mathcal{G}} \neq \emptyset$ and $P_{\mathcal{G}} \subseteq Optimal_S(\mathcal{G})$.*

What happens if we update a coordination game with such an optimal plan? Ruling out flat coordination games (in which there is no deontically ideal action profile), an update of a coordination game with an optimal plan does not introduce optimal group actions that were not in the plan and does not eliminate optimal group actions that were in the plan. The set of optimal group actions in the coordination game that results from updating a coordination game with an optimal plan equals the set of group actions in the plan:

Observation 2. *Let $S = \langle N, (A_i), d \rangle$ be a non-flat coordination game. Let $\mathcal{G} \subseteq N$ and let $P_{\mathcal{G}} \subseteq A_{\mathcal{G}}$ be optimal. Then*

$$Optimal_{S \uparrow P_{\mathcal{G}}}(\mathcal{G}) = P_{\mathcal{G}}.$$

This observation implies that every group action that is optimal in the coordination game $S \uparrow P_{\mathcal{G}}$ that results from updating a non-flat coordination game S with an optimal plan $P_{\mathcal{G}}$ is an optimal group action in the original coordination game S, that is, $Optimal_{S \uparrow P_{\mathcal{G}}}(\mathcal{G}) \subseteq Optimal_S(\mathcal{G})$. Accordingly, an update with an optimal plan highlights particular optimal group actions in the original coordination game.

However, it is not sufficient to require of a good plan that it be optimal. Indeed, if a plan consists of several optimal group actions and if each group member does her share of an optimal group action in the plan, then the composition of the group members' actions need not be an optimal

group action. Writing on action profiles in games with individual utility functions rather than on group actions in games with a deontic ideality function, David Gauthier (1975: 201) observes: 'If each person performs an action which has a best equilibrium as a possible outcome, and if there are several best equilibria, then the outcome need not be a best equilibrium.' Because of this, coordination theorists standardly assume that a solution to a coordination problem consists in singling out a unique action profile:

> Several best equilibria are too many of a good thing. ... What we need is a way to restructure our conception of the situation so that we are left with but one. We must restrict the possible actions which we consider, in such a way that we convert our representation of the situation into one with but one best equilibrium. (Gauthier 1975: 210)[17]

In the present setting, we do not need this methodological assumption. To tackle our current problem of specifying the conditions under which a group plan successfully coordinates the actions of group members, we do not have to assume that the agents must somehow agree on a unique group action. We show that it suffices to require of good plans not only that they be optimal, but also interchangeable.

4.2. Interchangeable plans

An optimal plan does not guarantee that if all group members act according to plan, then the group itself performs an optimal group action. We therefore need an additional requirement. This second design requirement of a good plan is that the plan be closed under component individual actions. We say that a group plan is *interchangeable* if and only if for any two group actions in the plan it holds that the composition of a group member's contribution to the first group action and the joint contribution of all other group members to the second group action amounts to a group action that is also in the plan:[18]

Definition 7 (Interchangeable Plans). *Let $S = \langle N, (A_i), d \rangle$ be a coordination game. Let $\mathcal{G} \subseteq N$ and let $P_{\mathcal{G}} \subseteq A_{\mathcal{G}}$. Then $P_{\mathcal{G}}$ is* interchangeable *if and only if for all $a_{\mathcal{G}}, a'_{\mathcal{G}} \in P_{\mathcal{G}}$ and all $i \in \mathcal{G}$ it holds that $(a_i, a'_{\mathcal{G}-i}) \in P_{\mathcal{G}}$.*

A few examples: in coordination game S_3 of Figure 3 the plan $\{(a_i, a_j)\}$ is interchangeable (as is every plan consisting of a single group action). Likewise, $\{(a_i, a_j), (a_i, a'_j), (a_i, a''_j)\}$ is interchangeable. The

---

[17] Compare Harsanyi and Selten (1988: 13): 'Clearly a theory telling us no more than that the outcome can be any one of these equilibrium points will not give us much useful information. We need a theory selecting one equilibrium point as the solution of the game.'

[18] The concept of interchangeability goes back to Nash (1951: 290).

	a_j	a'_j	a''_j
a_i	1	1	0
a'_i	0	0	1
a''_i	0	1	0

FIGURE 3. Coordination game S_3

plan $\{(a_i, a_j), (a_i, a'_j), (a'_i, a_j)\}$ is not, because (a'_i, a'_j) is not in it. Finally, $\{(a_i, a_j), (a_i, a''_j), (a'_i, a_j), (a'_i, a''_j)\}$ is interchangeable.

Our definition implies that a plan is interchangeable if and only if for any two group actions in the plan it holds that the composition of a subgroup's contribution to the first group action and the joint contribution of all other group members to the second group action amounts to a group action that is also in the plan:

Observation 3. *Let $S = \langle N, (A_i), d \rangle$ be a coordination game. Let $\mathcal{G} \subseteq N$ and let $P_\mathcal{G} \subseteq A_\mathcal{G}$. Then $P_\mathcal{G}$ is interchangeable if and only if for all $a_\mathcal{G}, a'_\mathcal{G} \in P_\mathcal{G}$ and all $\mathcal{F} \subseteq \mathcal{G}$ it holds that $(a_\mathcal{F}, a'_{\mathcal{G}-\mathcal{F}}) \in P_\mathcal{G}$.*

We have argued above that a group plan does not need to single out a unique group action to successfully coordinate the individual actions of the group members. To make this clear, consider again the coordination game S_3 of Figure 3. To solve their coordination problem, group members i and j must agree on a plan that successfully coordinates their actions. Even though it consists of several group actions, the optimal and interchangeable plan $\{(a_i, a_j), (a_i, a'_j)\}$ does the job. The plan specifies the following member obligations on agents i and j: agent i ought to perform action a_i and agent j ought to perform one of the actions a_j and a'_j. The plan hence gives some leeway to agent j. Nonetheless, the plan ensures that if both group members act according to plan, they perform an optimal group action and hence the group itself fulfils its collective obligation.

This point can be made fully general: an optimal and interchangeable plan *guarantees* that if every group member fulfils her member obligation specified by the plan, then the group itself fulfils its collective obligation. Let us see why this is so. If every group member fulfils her member obligation specified by an optimal and interchangeable plan, then every group member performs an action that is a component action of one of the group actions in the plan. Because the plan is interchangeable, the combination of the actions performed by the group members must also be in the plan. Because the plan is optimal, it only contains optimal group actions, and hence this combination of group member actions is an optimal group action. The group members hence perform an optimal

group action and thus fulfil the group's collective obligation. Therefore, if every group member acts according to an optimal and interchangeable plan, then the group itself fulfils its collective obligation.

4.3. Updates with optimal and interchangeable plans

Our central theorem concerns what happens if we update a coordination game with an optimal and interchangeable group plan. We show that if we update a non-flat coordination game with such a plan, then for any subgroup of this group it holds that an action of that subgroup is an optimal subgroup action in the coordination game that results from updating the original coordination game with the plan if and only if that subgroup action is a component subgroup action of one of the group actions in the plan:

Theorem 1. *Let $S = \langle N, (A_i), d \rangle$ be a non-flat coordination game. Let $\mathcal{F} \subseteq \mathcal{G} \subseteq N$ and let $P_{\mathcal{G}} \subseteq A_{\mathcal{G}}$ be optimal and interchangeable. Then*

$$Optimal_{S \uparrow P_{\mathcal{G}}}(\mathcal{F}) = \{a_{\mathcal{F}} : a_{\mathcal{G}} \in P_{\mathcal{G}}\}.$$

Our theorem shows that updates of coordination games with optimal and interchangeable plans relate collective obligations to individual obligations. Indeed, by setting $\mathcal{F} = \{i\}$ for any individual member i of $\mathcal{G}$ we obtain $Optimal_{S \uparrow P_{\mathcal{G}}}(i) = \{a_i : a_{\mathcal{G}} \in P_{\mathcal{G}}\}$. Accordingly, if we update a non-flat coordination game with an optimal and interchangeable group plan, then for every group member it holds that she fulfils her member obligation specified by the plan if and only if she fulfils her individual obligation in the coordination game that results from updating the original coordination game with the plan. Consequently, under the given assumptions, a group member acts according to plan if and only if she performs an optimal individual action in the changed decision context. It follows that, under the given assumptions, if every group member performs an optimal individual action in the changed decision context, then the group itself performs an optimal group action.

If we think of acting rationally as performing those actions that best promote deontic ideality, then we could say that an individual action is *individually rational* if and only if it is an optimal individual action, and that a group action is *collectively rational* if and only if it is an optimal group action. Within the narrow confines of our idealizing assumptions, Theorem 1 would then relate collective rationality to individual rationality and thereby give a tentative answer to Bratman's questions:

> Are there norms that are in some way fundamental to shared agency? If so, how precisely are they related to shared agency, and how are they related to norms of individual rationality? (Bratman 2014: 4)

The answer runs as follows. A group action is collectively rational if and only if it is an optimal group action. To perform an optimal group action, the group members must adopt a good plan to ensure that the combination of their actions is an optimal group action. After the public adoption of such a plan, a group member's action is rational with respect to the group's objective of performing an optimal group action if and only if this action is permitted by the plan.[19] Moreover, when the plan has been publicly adopted and thereby changed the context in which agents make a decision about what to do, an individual action is individually rational if and only if it is an optimal individual action in the changed decision context. Our theorem shows that, under the given assumptions, a group member's action is rational with respect to the group's objective of performing an optimal group action if and only if that action is individually rational in the decision context that results from the public adoption of the plan. It follows that, under the given assumptions, if every group member performs an individually rational action in the changed decision context, then the group itself performs a collectively rational group action. Therefore, if we think of acting rationally as performing those actions that best promote deontic ideality, our theorem would establish a connection between collective rationality and individual rationality.

5. ZERO-PRESERVING COOPERATION GAMES

Thus far, we have used coordination games involving a single deontic ideality function to study the relation between collective obligations, group plans, and individual actions. In this section we show that almost all our findings on coordination games transfer to 'cooperation games' involving group-relative utility functions, provided these utility functions are 'zero-preserving'. The only exception is the right-to-left inclusion of Theorem 1. In the context of zero-preserving cooperation games this property is undesirable. To show all of this let us first introduce cooperation games.

A cooperation game involves a finite set of individual agents. Each individual agent has a non-empty, finite set of available individual actions. Each group of agents assigns to each action profile a non-negative group utility:

---

[19] Compare Gold (2012: 185): 'The basic idea is that, when an individual reasons as a member of a team, she considers which *combination* of actions by members of the team would best promote the team's objective, and then performs her part of that combination. The rationality of each individual's action derives from the rationality of the joint action of the team.' Sugden (1993: 87), Anderson (2001: 28–30) and Bacharach (2006: 136) defend similar notions of group member rationality.

Definition 8 (Cooperation Game). *A cooperation game S is a triple $\langle N, (A_i), (u_{\mathcal{H}}) \rangle$, where N is a finite set of individual agents, for each agent i in N it holds that A_i is a non-empty and finite set of actions available to agent i, and for each group of agents $\mathcal{H}$ from N it holds that $u_{\mathcal{H}}$ is a utility function that assigns to each action profile a in $A (= \times_{i \in N} A_i)$ a utility $u_{\mathcal{H}}(a) \in \mathbb{R}_{\geq 0}$.*

To prove our claims about cooperation games we rule out cooperation games in which a group assigns to every action profile a utility of 0. Such games are *flat with respect to that group*:

Definition 9. *Let $S = \langle N, (A_i), (u_{\mathcal{H}}) \rangle$ be a cooperation game. Let $\mathcal{G} \subseteq N$. Then S is $\mathcal{G}$-flat if and only if for every action profile $a \in A$ it holds that $u_{\mathcal{G}}(a) = 0$.*

To define the $\mathcal{G}$-optimal group actions that are available to a group $\mathcal{G}$, we first define a dominance ordering: group action $a_{\mathcal{G}}$ *weakly $\mathcal{G}$-dominates* group action $a'_{\mathcal{G}}$ in cooperation game S (notation: $a_{\mathcal{G}} \succeq_S a'_{\mathcal{G}}$) if and only if $a_{\mathcal{G}}$ promotes the utility of group $\mathcal{G}$ in S at least as well as $a'_{\mathcal{G}}$, regardless of what $\mathcal{G}$'s non-members do:

Definition 10 ($\mathcal{G}$-Dominance). *Let $S = \langle N, (A_i), (u_{\mathcal{H}}) \rangle$ be a cooperation game. Let $\mathcal{G} \subseteq N$. Let $a_{\mathcal{G}}, a'_{\mathcal{G}} \in A_{\mathcal{G}}$. Then*

$$a_{\mathcal{G}} \succeq_S a'_{\mathcal{G}} \quad iff \quad for\ all\ a''_{-\mathcal{G}} \in A_{-\mathcal{G}}\ it\ holds\ that\ u_{\mathcal{G}}(a_{\mathcal{G}}, a''_{-\mathcal{G}}) \geq u_{\mathcal{G}}(a'_{\mathcal{G}}, a''_{-\mathcal{G}}).$$

Strong $\mathcal{G}$-dominance is defined in terms of weak $\mathcal{G}$-dominance: $a_{\mathcal{G}} \succ_S a'_{\mathcal{G}}$ if and only if $a_{\mathcal{G}} \succeq_S a'_{\mathcal{G}}$ and $a'_{\mathcal{G}} \not\succeq_S a_{\mathcal{G}}$.

A group action is $\mathcal{G}$-optimal in cooperation game S if and only if it is not strongly $\mathcal{G}$-dominated by any group action that is available to the group in S:

Definition 11 ($\mathcal{G}$-Optimal Actions). *Let $S = \langle N, (A_i), (u_{\mathcal{H}}) \rangle$ be a cooperation game. Let $\mathcal{G} \subseteq N$. Then*

$$Optimal_S(\mathcal{G}) = \{a_{\mathcal{G}} \in A_{\mathcal{G}} : there\ is\ no\ a'_{\mathcal{G}} \in A_{\mathcal{G}}\ such\ that\ a'_{\mathcal{G}} \succ_S a_{\mathcal{G}}\}.$$

Just as in coordination games, a group $\mathcal{G}$ fulfils its collective obligation in cooperation game S if and only if it performs one of its $\mathcal{G}$-optimal actions. Likewise, an agent i fulfils her individual obligation in cooperation game S if and only if she performs one of her i-optimal actions.

As before, a group plan $P_{\mathcal{G}}$ of group $\mathcal{G}$ in a cooperation game S can be any subset of the set $A_{\mathcal{G}}$ of group actions that are available to group $\mathcal{G}$ in S. Again, given an adopted group plan $P_{\mathcal{G}}$ and a group member i of $\mathcal{G}$, agent i fulfils her member obligation specified by the plan if and only if she performs an action that is her component action of one of the group actions in the plan.

In analogy to the plan updates of Section 3.1, we now model the public adoption of a group plan in a cooperation game as an update of the game's utility functions. Updating a cooperation game with a group

plan amounts to the following: if an action profile is forbidden by the plan, then for every group the group utility of that action profile becomes 0; and if an action profile is permitted by the plan, then for every group the group utility of that action profile remains unchanged:

Definition 12 (Plan Updates). *Let $S = \langle N, (A_i), (u_{\mathcal{H}}) \rangle$ be a cooperation game. Let $\mathcal{G} \subseteq N$ and let $P_{\mathcal{G}} \subseteq A_{\mathcal{G}}$. Then the update of S with $P_{\mathcal{G}}$, notation: $S \uparrow P_{\mathcal{G}}$, is the cooperation game $S \uparrow P_{\mathcal{G}} = \langle N, (A_i), (u_{\mathcal{H}}^{\uparrow}) \rangle$, where*

$$u_{\mathcal{H}}^{\uparrow}(a) = \begin{cases} u_{\mathcal{H}}(a), & \text{if } a_{\mathcal{G}} \in P_{\mathcal{G}} \\ 0, & \text{if } a_{\mathcal{G}} \notin P_{\mathcal{G}}. \end{cases}$$

As in coordination games, an update of a cooperation game with a group plan leaves the dominance ordering of the group actions in the plan unaffected:

Observation 4. *Let $S = \langle N, (A_i), (u_{\mathcal{H}}) \rangle$ be a cooperation game. Let $\mathcal{G} \subseteq N$ and let $P_{\mathcal{G}} \subseteq A_{\mathcal{G}}$. Let $a_{\mathcal{G}}, a_{\mathcal{G}}^{*} \in P_{\mathcal{G}}$. Then*

$$a_{\mathcal{G}} \succeq_S a_{\mathcal{G}}^{*} \quad \text{iff} \quad a_{\mathcal{G}} \succeq_{S \uparrow P_{\mathcal{G}}} a_{\mathcal{G}}^{*}.$$

$\mathcal{G}$-optimal plans and interchangeable plans are formally defined as in Definitions 6 and 7 above. A group plan is $\mathcal{G}$-*optimal* if and only if it is a non-empty subset of the group's $\mathcal{G}$-optimal actions. A group plan is *interchangeable* if and only if for any two group actions in the plan it holds that the composition of a group member's contribution to the first group action and the joint contribution of all other group members to the second group action amounts to a group action that is also in the plan. Again, a $\mathcal{G}$-optimal and interchangeable plan guarantees that if every group member fulfils her member obligation specified by the plan, then the group itself fulfils its collective obligation. For an informal proof of this conditional, see the end of Section 4.2.

After updating a non-$\mathcal{G}$-flat cooperation game with a $\mathcal{G}$-optimal group plan $P_{\mathcal{G}}$ it holds that a group action is a $\mathcal{G}$-optimal group action in the cooperation game that results from the update if and only if it is one of the group actions in the plan:

Observation 5. *Let $S = \langle N, (A_i), (u_{\mathcal{H}}) \rangle$ be a non-$\mathcal{G}$-flat cooperation game. Let $\mathcal{G} \subseteq N$ and let $P_{\mathcal{G}} \subseteq A_{\mathcal{G}}$ be $\mathcal{G}$-optimal. Then*

$$Optimal_{S \uparrow P_{\mathcal{G}}}(\mathcal{G}) = P_{\mathcal{G}}.$$

To transfer our findings on coordination games to cooperation games, we must relate group utilities to subgroup utilities. To do this, we require the relation between subgroup utilities and group utilities to be *zero-preserving*. A cooperation game is zero-preserving if and only if there is no action profile in the game which has a non-zero utility for some group, while it does have zero utility for some subgroup of that group:

Definition 13 (Zero-Preservation). *Let $S = \langle N, (A_i), (u_{\mathcal{H}}) \rangle$ be a cooperation game. Then S is* zero-preserving *if and only if for all non-empty groups $\mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq N$ and for all action profiles $a \in A$ it holds that if $u_{\mathcal{F}_1}(a) = 0$, then $u_{\mathcal{F}_2}(a) = 0$.*

Although this might seem a rather strong condition, any cooperation game can be transformed into a zero-preserving one by linearly transforming the game's utility functions: adding 1 to each utility in the game makes it zero-preserving. Apart from using linear transformations to ensure that a cooperation game is zero-preserving, some social welfare functions always ensure zero-preservation. The Rawlsian social welfare function $u_{\mathcal{H}}(a) = \min_{i \in \mathcal{H}}(u_i(a))$ is but one example. Other social welfare functions, such as summing or averaging, only meet this condition in some cooperation games – the Hi-Lo game that we discuss in Section 5.1 is a case in point.[20]

Plan updates sustain zero-preservation:

Observation 6. *Let $S = \langle N, (A_i), (u_{\mathcal{H}}) \rangle$ be a cooperation game. Let $\mathcal{G} \subseteq N$ and let $P_{\mathcal{G}} \subseteq A_{\mathcal{G}}$. Then*

> *If S is zero-preserving, then $S \uparrow P_{\mathcal{G}}$ is zero-preserving.*

Finally, the left-to-right inclusion of Theorem 1 also holds for cooperation games. We show that if we update a zero-preserving non-$\mathcal{G}$-flat cooperation game with a $\mathcal{G}$-optimal group plan $P_{\mathcal{G}}$, then for any subgroup $\mathcal{F}$ of $\mathcal{G}$ it holds that if an action of that subgroup is an $\mathcal{F}$-optimal subgroup action in the cooperation game that results from updating the original cooperation game with the plan, then that subgroup action is a component subgroup action of one of the group actions in the plan:

Observation 7. *Let $S = \langle N, (A_i), (u_{\mathcal{H}}) \rangle$ be a zero-preserving non-$\mathcal{G}$-flat cooperation game. Let $\mathcal{F} \subseteq \mathcal{G} \subseteq N$ be non-empty and let $P_{\mathcal{G}} \subseteq A_{\mathcal{G}}$ be $\mathcal{G}$-optimal. Then*

$$Optimal_{S \uparrow P_{\mathcal{G}}}(\mathcal{F}) \subseteq \{a_{\mathcal{F}} : a_{\mathcal{G}} \in P_{\mathcal{G}}\}.$$

Analogous to Theorem 1, our observation shows also that updates of cooperation games with $\mathcal{G}$-optimal plans relate collective obligations to individual obligations. Indeed, by setting $\mathcal{F} = \{i\}$ for any individual

---

[20] Team-reasoning theorists are not in favour of imposing conceptual constraints on social welfare functions. Because his theory of team reasoning does not solely concern cooperation for mutual advantage, Bacharach 'allowed in principle that the group objective might be welfare decreasing for some members' (Gold 2012: 195). According to Sugden (2000: 176), 'the preferences of a team are not necessarily reducible to, or capable of being constructed out of, the preferences that govern the choices that the members of the team make as individuals.'

	a'_j	a''_j
a'_i	$3, 4, 3$	$0, 1, 0$
a''_i	$4, 3, 3$	$2, 5, 2$

FIGURE 4. Cooperation game S_4

	a'_j	a''_j
a'_i	$3, 4, 3$	$0, 0, 0$
a''_i	$4, 3, 3$	$0, 0, 0$

FIGURE 5. Cooperation game $S_4 \uparrow P_{i,j}$

member i of $\mathcal{G}$ we obtain $Optimal_{S \uparrow P_{\mathcal{G}}}(i) \subseteq \{a_i : a_{\mathcal{G}} \in P_{\mathcal{G}}\}$. Accordingly, if we update a zero-preserving non-$\mathcal{G}$-flat cooperation game with a $\mathcal{G}$-optimal plan, then for every group member it holds that she fulfils her member obligation specified by the plan if she fulfils her individual obligation in the cooperation game that results from updating the original cooperation game with the plan. Consequently, under the given assumptions, a group member i acts according to plan if she performs an i-optimal individual action in the changed decision context. (Notice that, until now, we have not required the plan to be interchangeable.) It follows that, under the given assumptions, a $\mathcal{G}$-optimal and interchangeable plan guarantees that if each group member i performs an i-optimal individual action in the changed decision context, then the group itself performs a $\mathcal{G}$-optimal group action.

The right-to-left inclusion does not hold, because plan updates do not affect utilities of action profiles that are permitted by the plan. This means that if we update a zero-preserving non-$\mathcal{G}$-flat cooperation game with a $\mathcal{G}$-optimal and interchangeable group plan, then it might be that a group member acts according to plan without performing an i-optimal individual action in the changed decision context. To see this, consider the zero-preserving non-$\mathcal{G}$-flat cooperation game S_4 of Figure 4, where each action profile a is assigned a triple of utilities $\langle u_i(a), u_j(a), u_{i,j}(a) \rangle$. (The utility function of the empty group is suppressed.)

Updating S_4 with the $\mathcal{G}$-optimal and interchangeable plan $P_{i,j} = \{(a'_i, a'_j), (a''_i, a'_j)\}$ results in the cooperation game $S_4 \uparrow P_{i,j}$ of Figure 5, where it holds that $a'_i \in \{a_i : a_{i,j} \in P_{i,j}\}$ and $a'_i \notin Optimal_{S_4 \uparrow P_{i,j}}(i)$.

5.1. The team-reasoning account of cooperation

Our discussion of cooperation games permits us to compare our account of the relation between collective obligations, group plans and individual

	a'_j	a''_j
a'_i	$5, 5, 10$	$0, 0, 0$
a''_i	$0, 0, 0$	$1, 1, 2$

FIGURE 6. Cooperation game S_5

actions and the team-reasoning account of cooperation. Disregarding the differences in their accounts of cooperation, Bacharach and Sugden both argue that if we wish to give a game-theoretical explanation of cooperation in single-shot games, then an individualist perspective will not do.[21] Therefore, they propose to extend traditional game theory with group notions. Given a group of agents, team reasoning describes the reasoning process by which a group member first identifies the best combination of actions that the group members can perform and then decides to perform the individual action that is her part in that combination. To see how this works, consider the Hi-Lo cooperation game S_5 of Figure 6, where each action profile a is assigned a triple of utilities $\langle u_i(a), u_j(a), u_{i,j}(a) \rangle$ such that $u_{i,j}(a)$ is the sum of $u_i(a)$ and $u_j(a)$.

Agent i is engaged in team reasoning only if she reaches her decision about what to do by reasoning from the standpoint of the team consisting of i and j. First, it will then be clear to her that, of the four courses of action that are available to the group, the combination (a'_i, a'_j) is the best. She therefore decides to do her part in that combination and perform action a'_i. Reasoning in an analogous fashion, agent j decides to perform action a'_j. Accordingly, if both agents engage in team reasoning, they will perform the best group action. In contrast to traditional game theory, the team-reasoning account explains cooperation in Hi-Lo games even if communication is impossible.

Contrary to the team-reasoning account of cooperation, we assume that communication is possible and that agreement is unproblematic. Our research agenda therefore differs from that set by Bacharach and Sugden. Our account of cooperation starts from the assumption that the agents i and j are able to communicate and make agreements about what to do. If they decide to act as a group, they should abandon their individual perspective and evaluate the actions available to them from the standpoint of the group.[22] Because it will be clear to them that (a'_i, a'_j) is the only

---

[21] Gold (2012) provides a detailed comparison of Bacharach's and Sugden's accounts of team reasoning.

[22] Note that pre-play communication does not automatically lead to group identifying or team reasoning. Gold (2012: 200) observes: 'For both Bacharach and Sugden there are potential gaps between noticing the group and group identifying, and between group identifying and team reasoning.'

	a'_j	a''_j
a'_i	$5, 5, 10$	$0, 0, 0$
a''_i	$0, 0, 0$	$0, 0, 0$

FIGURE 7. Cooperation game $S_5 \uparrow \{(a'_i, a'_j)\}$

$\{i, j\}$-optimal group action, they should agree on the group plan $\{(a'_i, a'_j)\}$. At the moment they have publicly adopted this group plan, agent i has the member obligation to perform action a'_i and agent j has the member obligation to perform a'_j. Moreover, by publicly adopting this group plan, the decision context is changed into cooperation game $S_5 \uparrow \{(a'_i, a'_j)\}$ of Figure 7. In this changed decision context, agent i has the individual obligation to perform action a'_i and agent j has the individual obligation to perform action a'_j.

Our analysis hence suggests that, given a decision context in which the agents are able to communicate and make agreements, it holds that if they decide to act as a group $\mathcal{G}$ of agents, and if there is a unique $\mathcal{G}$-optimal group action, then the group members should agree on the group plan that only consists of this unique $\mathcal{G}$-optimal group action. The member obligations specified by the plan and the individual obligations in the decision context that results from publicly adopting the plan coincide with the recommendations that issue from the team-reasoning account. Therefore, if there is a unique $\mathcal{G}$-optimal group action, our account and the team-reasoning account of cooperation give similar recommendations to group members. Things are different, however, if there are several $\mathcal{G}$-optimal group actions: whereas the team-reasoning account has nothing much to offer, our analysis still applies.

We have argued that if the group members design and adopt a group plan to solve their coordination problem, and if we allow a group plan to consist of *several* group actions (thereby giving some leeway to at least some group members), the plan might fail to guarantee that if every group member acts according to plan, then the combination of their actions is a $\mathcal{G}$-optimal group action. Some such plans do guarantee this, however. The focus of our study has therefore been on the structural conditions that a group plan must meet in order to successfully coordinate the individual actions of the group members, and on what happens to a decision context when such a group plan is publicly adopted.

6. CONCLUSION

What, if anything, does our account of collective obligations, group plans and individual actions tell us about the concept of collective

blameworthiness, that is, backward-looking collective moral responsibility? The complexity of this concept has led some to argue that formal methods are hardly any use in the study of collective moral responsibility. Margaret Gilbert, for example, writes:

> What does the blameworthiness of the collective's act imply about the personal blameworthiness of any one member of that collective? From a logical point of view, the short answer is: *nothing*. Everything depends on the details of a given member's particular situation. (Gilbert 2006b: 109)[23]

Gilbert notes that some members of a blameworthy collective might have authorized others to decide and that this may make a difference to the attribution of blame to those members. At present, our formal analysis is too coarse to capture these structural aspects of collective moral responsibility. Nonetheless, even our relatively simple formalization of the relation between collective obligations, group plans and individual actions brings to the fore some central aspects of the logic of backward-looking collective responsibility.

    We take it that the concept of backward-looking individual responsibility supports the claim that not fulfilling an obligation is a necessary condition for being individually blameworthy: if an individual agent is individually blameworthy, then she has failed to fulfil an obligation. The converse does not hold, because she might have a plausible excuse for not doing what she ought to do. Analogously, we submit that not fulfilling a collective obligation is a necessary, but not a sufficient condition for being collectively blameworthy. We thus obtain the following implication: if a group is collectively blameworthy, then it does not fulfil a collective obligation.

    Our account of the relation between collective obligations, group plans and individual actions brought to light that a good plan has two key characteristics. First, a good plan guarantees that if every group member fulfils her member obligation specified by the plan, then the group itself fulfils its collective obligation. Secondly, if the group members, aiming to fulfil a collective obligation, publicly adopt a good plan to coordinate their individual actions, then for any group member it holds that if she fulfils her individual obligation in the decision context that results from publicly adopting the plan, then she fulfils her member obligation specified by the plan. We thus obtain a second implication (which holds for both coordination games and cooperation games): given that the group members, aiming to fulfil a collective obligation, publicly adopt a good plan to coordinate their individual actions, it holds that their failure to fulfil the collective obligation *logically implies* that at least one of them

---

[23] Likewise, Isaacs (2011: 24) argues that 'claims about collective moral responsibility neither entail nor are derivable from claims about individual moral responsibility.'

failed to fulfil her member obligation specified by the plan, which in turn entails that at least one group member failed to fulfil her individual obligation in the changed decision context.

Combining the two implications, we conclude: given that the group members, aiming to fulfil a collective obligation, publicly adopt a good plan to coordinate their actions, it holds that the collective blameworthiness of the group *logically implies* that at least one of the group members failed to act according to plan, which in turn entails that at least one group member failed to fulfil her individual obligation in changed decision context.[24] If, however, for some reason or other, the group members fail to agree on a group plan or if they agree on a bad plan, then it is not so clear how being collectively blameworthy relates to the failure of individuals to fulfil their obligations. The study of these cases must be left to another occasion.

REFERENCES

Anderson, E. 2001. Unstrapping the straitjacket of 'preference': a comment on Amartya Sen's contributions to philosophy and economics. *Economics and Philosophy* 17: 21–38.

Bacharach, M. 1999. Interactive team reasoning: a contribution to the theory of cooperation. *Research in Economics* 53: 117–147.

Bacharach, M. 2006. *Beyond Individual Choice: Teams and Frames in Game Theory*, ed. N. Gold and R. Sugden. Princeton, NJ: Princeton University Press.

Baltag, A., L.S. Moss and S. Solecki. 1998. The logic of public announcements, common knowledge, and private suspicions. In *Proceedings of the Seventh Conference on Theoretical Aspects of Rationality and Knowledge (TARK 1998)*, ed. I. Gilboa, 43–56. San Francisco, CA: Morgan Kaufmann Publishers.

Bratman, M.E. 1987. *Intention, Plans, and Practical Reason*. Cambridge, MA: Harvard University Press.

Bratman, M.E. 2014. *Shared Agency: A Planning Theory of Acting Together*. Oxford: Oxford University Press.

---

[24] Again, a group member's failure to fulfil her obligations does not entail that she is personally blameworthy – she may have had good reasons not to fulfil her obligations.

Chant, S.R. 2007. Unintentional collective action. *Philosophical Explorations* 10: 245–256.

Gauthier, D. 1975. Coordination. *Dialogue* 14: 195–221.

Gilbert, M. 1999. Obligation and joint commitment. *Utilitas* 11: 143–163.

Gilbert, M. 2006a. *A Theory of Political Obligation: Membership, Commitment, and the Bonds of Society*. Oxford: Clarendon Press.

Gilbert, M. 2006b. Who's to blame? Collective moral responsibility and its implications for group members. *Midwest Studies in Philosophy* 30: 94–114.

Gold, N. 2012. Team reasoning, framing and cooperation. In *Evolution and Rationality: Decisions, Co-operation and Strategic Behaviour*, ed. S. Okasha and K. Binmore, 185–212. Cambridge: Cambridge University Press.

Gold, N. and R. Sugden. 2007. Collective intentions and team agency. *Journal of Philosophy* 104: 109–137.

Harsanyi, J.C. and R. Selten. 1988. *A General Theory of Equilibrium Selection in Games*. Cambridge, MA: MIT Press.

Hilpinen, R., ed. 1971. *Deontic Logic: Introductory and Systematic Readings*. Dordrecht: D. Reidel Publishing Company.

Horty, J.F. 1996. Agency and obligation. *Synthese* 108: 269–307.

Horty, J.F. 2001. *Agency and Deontic Logic*. New York, NY: Oxford University Press.

Isaacs, T. 2011. *Moral Responsibility in Collective Contexts*. New York, NY: Oxford University Press.

Jackson, F. 1987. Group morality. In *Metaphysics and Morality*, ed. P. Pettit, R. Sylvan and J. Norman, 91–110. Oxford: Basil Blackwell.

Kooi, B. and A. Tamminga. 2008. Moral conflicts between groups of agents. *Journal of Philosophical Logic* 37: 1–21.

Kutz, C. 2000. *Complicity: Ethics and Law for a Collective Age*. Cambridge: Cambridge University Press.

Lewis, D.K. 1969. *Convention: A Philosophical Study*. Cambridge, MA: Harvard University Press.

Nash, J. 1951. Non-cooperative games. *Annals of Mathematics* 54: 286–295.

Osborne, M.J. and A. Rubinstein. 1994. *A Course in Game Theory*. Cambridge, MA: MIT Press.

Plaza, J. 1989. Logics of public communications. In *Proceedings of the Fourth International Symposium on Methodologies for Intelligent Systems (ISMIS 1989): Poster Session Program*, ed. M.L. Emrich, M.S. Pfeifer, M. Hadzikadic and Z.W. Ras, 201–216. Charlotte, NC: Oak Ridge National Laboratory.

Regan, D. 1980. *Utilitarianism and Co-operation*. New York, NY: Oxford University Press.

Roy, O. 2009a. Intentions and interactive transformations of decision problems. *Synthese* 169: 335–349.

Roy, O. 2009b. A dynamic-epistemic hybrid logic for intentions and information changes in strategic games. *Synthese* 171: 291–320.

Scanlon, T.M. 1998. *What We Owe to Each Other*. Cambridge, MA: Belknap Press.

Schelling, T.C. 1960. *The Strategy of Conflict*. Cambridge, MA: Harvard University Press.

Sugden, R. 1993. Thinking as a team: towards an explanation of non-selfish behavior. *Social Philosophy and Policy* 10: 69–89.

Sugden, R. 1995. A theory of focal points. *Economic Journal* 105: 533–550.

Sugden, R. 2000. Team preferences. *Economics and Philosophy* 16: 175–204.

Sugden, R. 2003. The logic of team reasoning. *Philosophical Explorations* 6: 165–181.

Tuomela, R. 2000. *Cooperation: A Philosophical Study*. Dordrecht: Kluwer Academic Publishers.

Tuomela, R. 2005. We-intentions revisited. *Philosophical Studies* 125: 327–369.

Tuomela, R. 2013. *Social Ontology: Collective Intentionality and Group Agents*. New York, NY: Oxford University Press.

van Ditmarsch, H., W. van der Hoek and B. Kooi. 2007. *Dynamic Epistemic Logic*. Dordrecht: Springer.

van Hees, M. and O. Roy. 2008. Intentions and plans in decision and game theory. In *Reasons and Intentions*, ed. B. Verbeek, 207–226. Aldershot: Ashgate.

APPENDIX

COORDINATION GAMES: PROOFS

Lemma 1 (Horty 2001: 74). *Let $S = \langle N, (A_i), d \rangle$ be a coordination game. Let $\mathcal{G} \subseteq N$ be a non-empty group of agents. Then*

$$Optimal_S(\mathcal{G}) \neq \emptyset.$$

Proof. Assume that $Optimal_S(\mathcal{G}) = \emptyset$. By definition of S, it holds that $A_{\mathcal{G}}$ is finite and non-empty. Then for all $a'_{\mathcal{G}} \in A_{\mathcal{G}}$ there is an $a_{\mathcal{G}} \in A_{\mathcal{G}}$ such that $a_{\mathcal{G}} \succ_S a'_{\mathcal{G}}$. Because $A_{\mathcal{G}}$ is finite, there must be a cycle $a^n_{\mathcal{G}} \succ_S a^{n-1}_{\mathcal{G}} \succ_S \ldots \succ_S a^2_{\mathcal{G}} \succ_S a^1_{\mathcal{G}} = a^n_{\mathcal{G}}$ with all members $a^k_{\mathcal{G}}$ in $A_{\mathcal{G}}$. Because $\succ_S$ is transitive, it holds that $a^n_{\mathcal{G}} \succ_S a^n_{\mathcal{G}}$, that is, $a^n_{\mathcal{G}} \succeq_S a^n_{\mathcal{G}}$ and $a^n_{\mathcal{G}} \not\succeq_S a^n_{\mathcal{G}}$. Contradiction. Therefore, $Optimal_S(\mathcal{G}) \neq \emptyset$. $\square$

Lemma 2 (Kooi and Tamminga 2008: 9). *Let $S = \langle N, (A_i), d \rangle$ be a coordination game. Let $\mathcal{F} \subseteq \mathcal{G} \subseteq N$. Let $a_{\mathcal{F}}, a'_{\mathcal{F}} \in A_{\mathcal{F}}$ and let $a''_{\mathcal{G}-\mathcal{F}} \in A_{\mathcal{G}-\mathcal{F}}$. Then*

$$\text{If } a_{\mathcal{F}} \succeq_S a'_{\mathcal{F}}, \text{ then } (a_{\mathcal{F}}, a''_{\mathcal{G}-\mathcal{F}}) \succeq_S (a'_{\mathcal{F}}, a''_{\mathcal{G}-\mathcal{F}}).$$

Proof. Assume $a_{\mathcal{F}} \succeq_S a'_{\mathcal{F}}$. Let $a''_{\mathcal{G}-\mathcal{F}} \in A_{\mathcal{G}-\mathcal{F}}$. Suppose $a'''_{-\mathcal{G}} \in A_{-\mathcal{G}}$. Then $(a''_{\mathcal{G}-\mathcal{F}}, a'''_{-\mathcal{G}}) \in A_{-\mathcal{F}}$. Hence, by our assumption it must be that $d(a_{\mathcal{F}}, a''_{\mathcal{G}-\mathcal{F}}, a'''_{-\mathcal{G}}) \geq d(a'_{\mathcal{F}}, a''_{\mathcal{G}-\mathcal{F}}, a'''_{-\mathcal{G}})$. Therefore, for all $a'''_{-\mathcal{G}} \in A_{-\mathcal{G}}$ it holds that $d(a_{\mathcal{F}}, a''_{\mathcal{G}-\mathcal{F}}, a'''_{-\mathcal{G}}) \geq d(a'_{\mathcal{F}}, a''_{\mathcal{G}-\mathcal{F}}, a'''_{-\mathcal{G}})$. Hence, $(a_{\mathcal{F}}, a''_{\mathcal{G}-\mathcal{F}}) \succeq_S (a'_{\mathcal{F}}, a''_{\mathcal{G}-\mathcal{F}})$. $\square$

Lemma 3. *Let $S = \langle N, (A_i), d \rangle$ be a non-flat coordination game. Let $\mathcal{G} \subseteq N$. Then*

$$\text{If } a_{\mathcal{G}} \in Optimal_S(\mathcal{G}), \text{ then there is an } a'_{-\mathcal{G}} \in A_{-\mathcal{G}} \text{ such that } d(a_{\mathcal{G}}, a'_{-\mathcal{G}}) = 1.$$

Proof. Suppose that $a_{\mathcal{G}} \in Optimal_S(\mathcal{G})$. Suppose that for every $a'_{-\mathcal{G}} \in A_{-\mathcal{G}}$ it holds that $d(a_{\mathcal{G}}, a'_{-\mathcal{G}}) = 0$. Because S is non-flat, there is an $a^* \in A$ such that $d(a^*) = 1$. Consider $a^*_{\mathcal{G}}$. It holds that $a^*_{\mathcal{G}} \succ_S a_{\mathcal{G}}$, that is (1) $a^*_{\mathcal{G}} \succeq_S a_{\mathcal{G}}$ and (2) $a_{\mathcal{G}} \not\succeq_S a^*_{\mathcal{G}}$. We have (1), because for every $a'_{-\mathcal{G}} \in A_{-\mathcal{G}}$ it holds that $d(a_{\mathcal{G}}, a'_{-\mathcal{G}}) = 0$. We have (2), because $d(a_{\mathcal{G}}, a^*_{-\mathcal{G}}) = 0$ and $d(a^*_{\mathcal{G}}, a^*_{-\mathcal{G}}) = d(a^*) = 1$. Hence, $a_{\mathcal{G}} \notin Optimal_S(\mathcal{G})$. Contradiction. $\square$

Observation 1. *Let $S = \langle N, (A_i), d \rangle$ be a coordination game. Let $\mathcal{G} \subseteq N$ and let $P_{\mathcal{G}} \subseteq A_{\mathcal{G}}$. Let $a_{\mathcal{G}}, a^*_{\mathcal{G}} \in P_{\mathcal{G}}$. Then*

$$a_{\mathcal{G}} \succeq_S a^*_{\mathcal{G}} \quad \text{iff} \quad a_{\mathcal{G}} \succeq_{S \uparrow P_{\mathcal{G}}} a^*_{\mathcal{G}}.$$

Proof. It is given that $a_{\mathcal{G}}, a^*_{\mathcal{G}} \in P_{\mathcal{G}}$. By construction of $S \uparrow P_{\mathcal{G}}$, for every $a'_{-\mathcal{G}} \in A_{-\mathcal{G}}$ it holds that $d(a_{\mathcal{G}}, a'_{-\mathcal{G}}) = d^{\uparrow}(a_{\mathcal{G}}, a'_{-\mathcal{G}})$ and $d(a^*_{\mathcal{G}}, a'_{-\mathcal{G}}) = d^{\uparrow}(a^*_{\mathcal{G}}, a'_{-\mathcal{G}})$. Then the following four statements are equivalent: (1) $a_{\mathcal{G}} \succeq_S a^*_{\mathcal{G}}$; (2) for every $a'_{-\mathcal{G}} \in A_{-\mathcal{G}}$ it holds that $d(a_{\mathcal{G}}, a'_{-\mathcal{G}}) \geq d(a^*_{\mathcal{G}}, a'_{-\mathcal{G}})$; (3) for every $a'_{-\mathcal{G}} \in A_{-\mathcal{G}}$ it holds that $d^{\uparrow}(a_{\mathcal{G}}, a'_{-\mathcal{G}}) \geq d^{\uparrow}(a^*_{\mathcal{G}}, a'_{-\mathcal{G}})$; and (4) $a_{\mathcal{G}} \succeq_{S \uparrow P_{\mathcal{G}}} a^*_{\mathcal{G}}$. $\square$

Lemma 4. *Let $S = \langle N, (A_i), d \rangle$ be a coordination game. Let $\mathcal{G} \subseteq N$ and let $P_{\mathcal{G}}$ be optimal. Then*

$$\text{If } S \text{ is non-flat, then } S \uparrow P_{\mathcal{G}} \text{ is non-flat.}$$

Proof. Suppose that S is non-flat. Because $P_{\mathcal{G}}$ is optimal, there is an $a_{\mathcal{G}} \in P_{\mathcal{G}}$ such that $a_{\mathcal{G}} \in Optimal_S(\mathcal{G})$. By Lemma 3 there is an $a'_{-\mathcal{G}} \in A_{-\mathcal{G}}$ such that $d(a_{\mathcal{G}}, a'_{-\mathcal{G}}) = 1$. By the construction of $S \uparrow P_{\mathcal{G}}$, it must be that $d^{\uparrow}(a_{\mathcal{G}}, a'_{-\mathcal{G}}) = 1$. Therefore, $S \uparrow P_{\mathcal{G}}$ is non-flat. $\qquad\square$

Observation 2. *Let $S = \langle N, (A_i), d \rangle$ be a non-flat coordination game. Let $\mathcal{G} \subseteq N$ and let $P_{\mathcal{G}} \subseteq A_{\mathcal{G}}$ be optimal. Then*

$$Optimal_{S \uparrow P_{\mathcal{G}}}(\mathcal{G}) = P_{\mathcal{G}}.$$

Proof. ($\subseteq$) Suppose that $a_{\mathcal{G}} \in Optimal_{S \uparrow P_{\mathcal{G}}}(\mathcal{G})$. By Lemma 4, $S \uparrow P_{\mathcal{G}}$ is non-flat. By Lemma 3, it must be that there is an $a'_{-\mathcal{G}} \in A_{-\mathcal{G}}$ such that $d^{\uparrow}(a_{\mathcal{G}}, a'_{-\mathcal{G}}) = 1$. By construction of $S \uparrow P_{\mathcal{G}}$, it must be that $a_{\mathcal{G}} \in P_{\mathcal{G}}$.

($\supseteq$) Suppose that $a_{\mathcal{G}} \in P_{\mathcal{G}}$. It is given that $a_{\mathcal{G}} \in Optimal_S(\mathcal{G})$ and that S is non-flat. Suppose that $a_{\mathcal{G}} \notin Optimal_{S \uparrow P_{\mathcal{G}}}(\mathcal{G})$. Then there is an $a^{*}_{\mathcal{G}} \in Optimal_{S \uparrow P_{\mathcal{G}}}(\mathcal{G})$ such that $a^{*}_{\mathcal{G}} \succ_{S \uparrow P_{\mathcal{G}}} a_{\mathcal{G}}$. By the first part of this proof, $a^{*}_{\mathcal{G}} \in P_{\mathcal{G}}$. By Observation 1, it must be that $a^{*}_{\mathcal{G}} \succ_S a_{\mathcal{G}}$. Hence, $a_{\mathcal{G}} \notin Optimal_S(\mathcal{G})$. Contradiction. Therefore, $a_{\mathcal{G}} \in Optimal_{S \uparrow P_{\mathcal{G}}}(\mathcal{G})$. $\qquad\square$

Observation 3. *Let $S = \langle N, (A_i), d \rangle$ be a coordination game. Let $\mathcal{G} \subseteq N$ and let $P_{\mathcal{G}} \subseteq A_{\mathcal{G}}$. Then $P_{\mathcal{G}}$ is interchangeable if and only if for all $a_{\mathcal{G}}, a'_{\mathcal{G}} \in P_{\mathcal{G}}$ and all $\mathcal{F} \subseteq \mathcal{G}$ it holds that $(a_{\mathcal{F}}, a'_{\mathcal{G}-\mathcal{F}}) \in P_{\mathcal{G}}$.*

Proof. By straightforward induction on the number of members of $\mathcal{F}$. $\qquad\square$

Theorem 1. *Let $S = \langle N, (A_i), d \rangle$ be a non-flat coordination game. Let $\mathcal{F} \subseteq \mathcal{G} \subseteq N$ and let $P_{\mathcal{G}} \subseteq A_{\mathcal{G}}$ be optimal and interchangeable. Then*

$$Optimal_{S \uparrow P_{\mathcal{G}}}(\mathcal{F}) = \{a_{\mathcal{F}} : a_{\mathcal{G}} \in P_{\mathcal{G}}\}.$$

Proof. ($\subseteq$) Suppose that $a_{\mathcal{F}} \in Optimal_{S \uparrow P_{\mathcal{G}}}(\mathcal{F})$. By Lemma 3, there is an $a'_{-\mathcal{F}} \in A_{-\mathcal{F}}$ such that $d^{\uparrow}(a_{\mathcal{F}}, a'_{-\mathcal{F}}) = 1$. By the construction of $S \uparrow P_{\mathcal{G}}$, it must be that $(a_{\mathcal{F}}, a'_{\mathcal{G}-\mathcal{F}}) \in P_{\mathcal{G}}$. Therefore, $a_{\mathcal{F}} \in \{a_{\mathcal{F}} : a_{\mathcal{G}} \in P_{\mathcal{G}}\}$.

($\supseteq$) Suppose that $a_{\mathcal{F}} \in \{a_{\mathcal{F}} : a_{\mathcal{G}} \in P_{\mathcal{G}}\}$. Then there is an $a'_{\mathcal{G}-\mathcal{F}} \in A_{\mathcal{G}-\mathcal{F}}$ such that $(a_{\mathcal{F}}, a'_{\mathcal{G}-\mathcal{F}}) \in P_{\mathcal{G}}$. Suppose that $a_{\mathcal{F}} \notin Optimal_{S \uparrow P_{\mathcal{G}}}(\mathcal{F})$. Then there is an $a^{*}_{\mathcal{F}} \in Optimal_{S \uparrow P_{\mathcal{G}}}(\mathcal{F})$ such that $a^{*}_{\mathcal{F}} \succ_{S \uparrow P_{\mathcal{G}}} a_{\mathcal{F}}$. Then there must be an $a^{**}_{-\mathcal{F}} \in A_{-\mathcal{F}}$ such that $d^{\uparrow}(a^{*}_{\mathcal{F}}, a^{**}_{-\mathcal{F}}) > d^{\uparrow}(a_{\mathcal{F}}, a^{**}_{-\mathcal{F}})$. Let $a'' = (a^{*}_{\mathcal{F}}, a^{**}_{-\mathcal{F}})$. Then $d^{\uparrow}(a'') = 1$. By construction of $S \uparrow P_{\mathcal{G}}$, it must be that $a''_{\mathcal{G}} = (a^{*}_{\mathcal{F}}, a''_{\mathcal{G}-\mathcal{F}}) = (a^{*}_{\mathcal{F}}, a''_{\mathcal{G}-\mathcal{F}}) \in P_{\mathcal{G}}$. By Observation 3, it must be that $(a_{\mathcal{F}}, a''_{\mathcal{G}-\mathcal{F}}) \in P_{\mathcal{G}}$. Because $a^{*}_{\mathcal{F}} \succeq_{S \uparrow P_{\mathcal{G}}} a_{\mathcal{F}}$ and Lemma 2, it must be that $(a^{*}_{\mathcal{F}}, a''_{\mathcal{G}-\mathcal{F}}) \succeq_{S \uparrow P_{\mathcal{G}}} (a_{\mathcal{F}}, a''_{\mathcal{G}-\mathcal{F}})$. Because it holds that $d^{\uparrow}(a^{*}_{\mathcal{F}}, a''_{\mathcal{G}-\mathcal{F}}, a'_{-\mathcal{G}}) = d^{\uparrow}(a^{*}_{\mathcal{F}}, a^{**}_{-\mathcal{F}}) = 1$ and $d^{\uparrow}(a_{\mathcal{F}}, a''_{\mathcal{G}-\mathcal{F}}, a'_{-\mathcal{G}}) = d^{\uparrow}(a_{\mathcal{F}}, a^{**}_{-\mathcal{F}}) = 0$, we have $(a_{\mathcal{F}}, a''_{\mathcal{G}-\mathcal{F}}) \not\succeq_{S \uparrow P_{\mathcal{G}}} (a^{*}_{\mathcal{F}}, a''_{\mathcal{G}-\mathcal{F}})$. Then it must be that $(a^{*}_{\mathcal{F}}, a''_{\mathcal{G}-\mathcal{F}}) \succ_{S \uparrow P_{\mathcal{G}}} (a_{\mathcal{F}}, a''_{\mathcal{G}-\mathcal{F}})$. By Observation 1, it must be that $(a^{*}_{\mathcal{F}}, a''_{\mathcal{G}-\mathcal{F}}) \succ_S (a_{\mathcal{F}}, a''_{\mathcal{G}-\mathcal{F}})$. Then $(a_{\mathcal{F}}, a''_{\mathcal{G}-\mathcal{F}}) \notin Optimal_S(\mathcal{G})$ which contradicts the fact that $(a_{\mathcal{F}}, a''_{\mathcal{G}-\mathcal{F}}) \in P_{\mathcal{G}}$. Therefore, $a_{\mathcal{F}} \in Optimal_{S \uparrow P_{\mathcal{G}}}(\mathcal{F})$. $\qquad\square$

COOPERATION GAMES: PROOFS

Lemma 5. *Let $S = \langle N, (A_i), (u_{\mathcal{H}}) \rangle$ be a non-$\mathcal{G}$-flat cooperation game. Let $\mathcal{G} \subseteq N$. Then*

If $a_{\mathcal{G}} \in Optimal_S(\mathcal{G})$, then there is an $a'_{-\mathcal{G}} \in A_{-\mathcal{G}}$ such that $u_{\mathcal{G}}(a_{\mathcal{G}}, a'_{-\mathcal{G}}) > 0$.

Proof. Suppose that $a_{\mathcal{G}} \in Optimal_S(\mathcal{G})$. Suppose that for every $a'_{-\mathcal{G}} \in A_{-\mathcal{G}}$ it holds that $u_{\mathcal{G}}(a_{\mathcal{G}}, a'_{-\mathcal{G}}) = 0$. Because S is non-$\mathcal{G}$-flat, there is an $a^* \in A$ such that $u_{\mathcal{G}}(a^*) > 0$. Consider $a^*_{\mathcal{G}}$. It holds that $a^*_{\mathcal{G}} \succ_S a_{\mathcal{G}}$, that is (1) $a^*_{\mathcal{G}} \succeq_S a_{\mathcal{G}}$ and (2) $a_{\mathcal{G}} \not\succeq_S a^*_{\mathcal{G}}$. We have (1), because for every $a'_{-\mathcal{G}} \in A_{-\mathcal{G}}$ it holds that $u_{\mathcal{G}}(a_{\mathcal{G}}, a'_{-\mathcal{G}}) = 0$. We have (2), because $u_{\mathcal{G}}(a_{\mathcal{G}}, a^*_{-\mathcal{G}}) = 0$ and $u_{\mathcal{G}}(a^*_{\mathcal{G}}, a^*_{-\mathcal{G}}) = u_{\mathcal{G}}(a^*) > 0$. Hence, $a_{\mathcal{G}} \notin Optimal_S(\mathcal{G})$. Contradiction. $\square$

Observation 4. *Let $S = \langle N, (A_i), (u_{\mathcal{H}}) \rangle$ be a cooperation game. Let $\mathcal{G} \subseteq N$ and let $P_{\mathcal{G}} \subseteq A_{\mathcal{G}}$. Let $a_{\mathcal{G}}, a^*_{\mathcal{G}} \in P_{\mathcal{G}}$. Then*

$$a_{\mathcal{G}} \succeq_S a^*_{\mathcal{G}} \quad iff \quad a_{\mathcal{G}} \succeq_{S \uparrow P_{\mathcal{G}}} a^*_{\mathcal{G}}.$$

Proof. It is given that $a_{\mathcal{G}}, a^*_{\mathcal{G}} \in P_{\mathcal{G}}$. By construction of $S \uparrow P_{\mathcal{G}}$, for every $a'_{-\mathcal{G}} \in A_{-\mathcal{G}}$ it holds that $u_{\mathcal{G}}(a_{\mathcal{G}}, a'_{-\mathcal{G}}) = u^{\uparrow}_{\mathcal{G}}(a_{\mathcal{G}}, a'_{-\mathcal{G}})$ and $u_{\mathcal{G}}(a^*_{\mathcal{G}}, a'_{-\mathcal{G}}) = u^{\uparrow}_{\mathcal{G}}(a^*_{\mathcal{G}}, a'_{-\mathcal{G}})$. Then the following four statements are equivalent: (1) $a_{\mathcal{G}} \succeq_S a^*_{\mathcal{G}}$; (2) for every $a'_{-\mathcal{G}} \in A_{-\mathcal{G}}$ it holds that $u_{\mathcal{G}}(a_{\mathcal{G}}, a'_{-\mathcal{G}}) \geq u_{\mathcal{G}}(a^*_{\mathcal{G}}, a'_{-\mathcal{G}})$; (3) for every $a'_{-\mathcal{G}} \in A_{-\mathcal{G}}$ it holds that $u^{\uparrow}_{\mathcal{G}}(a_{\mathcal{G}}, a'_{-\mathcal{G}}) \geq u^{\uparrow}_{\mathcal{G}}(a^*_{\mathcal{G}}, a'_{-\mathcal{G}})$; and (4) $a_{\mathcal{G}} \succeq_{S \uparrow P_{\mathcal{G}}} a^*_{\mathcal{G}}$. $\square$

Lemma 6. *Let $S = \langle N, (A_i), (u_{\mathcal{H}}) \rangle$ be a cooperation game. Let $\mathcal{G} \subseteq N$ and let $P_{\mathcal{G}} \subseteq A_{\mathcal{G}}$ be $\mathcal{G}$-optimal. Then*

If S is non-$\mathcal{G}$-flat, then $S \uparrow P_{\mathcal{G}}$ is non-$\mathcal{G}$-flat.

Proof. Suppose that S is non-$\mathcal{G}$-flat. Because $P_{\mathcal{G}}$ is $\mathcal{G}$-optimal, there is an $a_{\mathcal{G}} \in P_{\mathcal{G}}$ such that $a_{\mathcal{G}} \in Optimal_S(\mathcal{G})$. By Lemma 5, there is an $a'_{-\mathcal{G}} \in A_{-\mathcal{G}}$ such that $u_{\mathcal{G}}(a_{\mathcal{G}}, a'_{-\mathcal{G}}) > 0$. By the construction of $S \uparrow P_{\mathcal{G}}$, it must be that $u^{\uparrow}_{\mathcal{G}}(a_{\mathcal{G}}, a'_{-\mathcal{G}}) > 0$. Therefore, $S \uparrow P_{\mathcal{G}}$ is non-$\mathcal{G}$-flat. $\square$

Observation 5. *Let $S = \langle N, (A_i), (u_{\mathcal{H}}) \rangle$ be a non-$\mathcal{G}$-flat cooperation game. Let $\mathcal{G} \subseteq N$ and let $P_{\mathcal{G}} \subseteq A_{\mathcal{G}}$ be $\mathcal{G}$-optimal. Then*

$$Optimal_{S \uparrow P_{\mathcal{G}}}(\mathcal{G}) = P_{\mathcal{G}}.$$

Proof. ($\subseteq$) Suppose that $a_{\mathcal{G}} \in Optimal_{S \uparrow P_{\mathcal{G}}}(\mathcal{G})$. By Lemma 6, $S \uparrow P_{\mathcal{G}}$ is non-$\mathcal{G}$-flat. By Lemma 5, there is an $a'_{-\mathcal{G}} \in A_{-\mathcal{G}}$ such that $u^{\uparrow}_{\mathcal{G}}(a_{\mathcal{G}}, a'_{-\mathcal{G}}) > 0$. By construction of $S \uparrow P_{\mathcal{G}}$, it must be that $a_{\mathcal{G}} \in P_{\mathcal{G}}$.

($\supseteq$) Suppose that $a_{\mathcal{G}} \in P_{\mathcal{G}}$. It is given that $a_{\mathcal{G}} \in Optimal_S(\mathcal{G})$ and that S is non-$\mathcal{G}$-flat. Suppose that $a_{\mathcal{G}} \notin Optimal_{S \uparrow P_{\mathcal{G}}}(\mathcal{G})$. Then there is an $a^*_{\mathcal{G}} \in Optimal_{S \uparrow P_{\mathcal{G}}}(\mathcal{G})$ such that $a^*_{\mathcal{G}} \succ_{S \uparrow P_{\mathcal{G}}} a_{\mathcal{G}}$. By the first part of this proof, $a^*_{\mathcal{G}} \in P_{\mathcal{G}}$. By Observation 4, it must be that $a^*_{\mathcal{G}} \succ_S a_{\mathcal{G}}$. Hence, $a_{\mathcal{G}} \notin Optimal_S(\mathcal{G})$. Contradiction. Therefore, $a_{\mathcal{G}} \in Optimal_{S \uparrow P_{\mathcal{G}}}(\mathcal{G})$. $\square$

Observation 6. *Let $S = \langle N, (A_i), (u_{\mathcal{H}}) \rangle$ be a cooperation game. Let $\mathcal{G} \subseteq N$ and let $P_{\mathcal{G}} \subseteq A_{\mathcal{G}}$. Then*

If S is zero-preserving, then $S \uparrow P_{\mathcal{G}}$ is zero-preserving.

Proof. Suppose that S is zero-preserving. Suppose that $\mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq N$ are non-empty and suppose that $a \in A$. There are two cases: (1) $a_{\mathcal{G}} \in P_{\mathcal{G}}$ and (2) $a_{\mathcal{G}} \notin P_{\mathcal{G}}$. If we have (1), then by construction of $S \uparrow P_{\mathcal{G}}$ it holds that $u^{\uparrow}_{\mathcal{F}_1}(a) = u_{\mathcal{F}_1}(a)$ and $u^{\uparrow}_{\mathcal{F}_2}(a) = u_{\mathcal{F}_2}(a)$. It is given that if $u_{\mathcal{F}_1}(a) = 0$, then $u_{\mathcal{F}_2}(a) = 0$. Hence, if $u^{\uparrow}_{\mathcal{F}_1}(a) = 0$, then $u^{\uparrow}_{\mathcal{F}_2}(a) = 0$. If

we have (2), then by construction of $S \uparrow P_{\mathcal{G}}$ it holds that $u_{\mathcal{F}_1}^{\uparrow}(a) = 0$ and $u_{\mathcal{F}_2}^{\uparrow}(a) = 0$. Hence, if $u_{\mathcal{F}_1}^{\uparrow}(a) = 0$, then $u_{\mathcal{F}_2}^{\uparrow}(a) = 0$. Therefore, $S \uparrow P_{\mathcal{G}}$ is zero-preserving. □

Lemma 7. *Let $S = \langle N, (A_i), (u_{\mathcal{H}}) \rangle$ be a zero-preserving cooperation game. Let $\mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq N$ be non-empty. Then*

> *If S is $\mathcal{F}_1$-flat, then S is $\mathcal{F}_2$-flat.*

Proof. Suppose that S is $\mathcal{F}_1$-flat. Then for every $a \in A$ it holds that $u_{\mathcal{F}_1}(a) = 0$. Because S is zero-preserving, for every $a \in A$ it holds that $u_{\mathcal{F}_2}(a) = 0$. Therefore, S is $\mathcal{F}_2$-flat. □

Observation 7. *Let $S = \langle N, (A_i), (u_{\mathcal{H}}) \rangle$ be a zero-preserving non-$\mathcal{G}$-flat cooperation game. Let $\mathcal{F} \subseteq \mathcal{G} \subseteq N$ be non-empty and let $P_{\mathcal{G}} \subseteq A_{\mathcal{G}}$ be $\mathcal{G}$-optimal. Then*

$$Optimal_{S \uparrow P_{\mathcal{G}}}(\mathcal{F}) \subseteq \{a_{\mathcal{F}} : a_{\mathcal{G}} \in P_{\mathcal{G}}\}.$$

Proof. Suppose that $a_{\mathcal{F}} \in Optimal_{S \uparrow P_{\mathcal{G}}}(\mathcal{F})$. By Lemma 7, S is non-$\mathcal{F}$-flat. By Lemma 5, there is an $a'_{-\mathcal{F}} \in A_{-\mathcal{F}}$ such that $u_{\mathcal{F}}^{\uparrow}(a_{\mathcal{F}}, a'_{-\mathcal{F}}) > 0$. By the construction of $S \uparrow P_{\mathcal{G}}$, it must be that $(a_{\mathcal{F}}, a'_{\mathcal{G}-\mathcal{F}}) \in P_{\mathcal{G}}$. Therefore, $a_{\mathcal{F}} \in \{a_{\mathcal{F}} : a_{\mathcal{G}} \in P_{\mathcal{G}}\}$. □

BIOGRAPHICAL INFORMATION

Allard Tamminga is a senior lecturer in Theoretical Philosophy at the University of Groningen and a researcher in Theoretical Philosophy at Utrecht University. He is the author of 'Deontic logic for strategic games' (*Erkenntnis*, 2013), and, with Barteld Kooi, the author of 'Moral conflicts between groups of agents' (*Journal of Philosophical Logic*, 2008). His current research interests focus on formal aspects of collective moral responsibility and on proof theory for many-valued logics.

Hein Duijf is a PhD student in Theoretical Philosophy at Utrecht University. His formal-philosophical studies embrace collective agency (including collective intentionality and collective moral responsibility), practical reasoning and interactive epistemology.