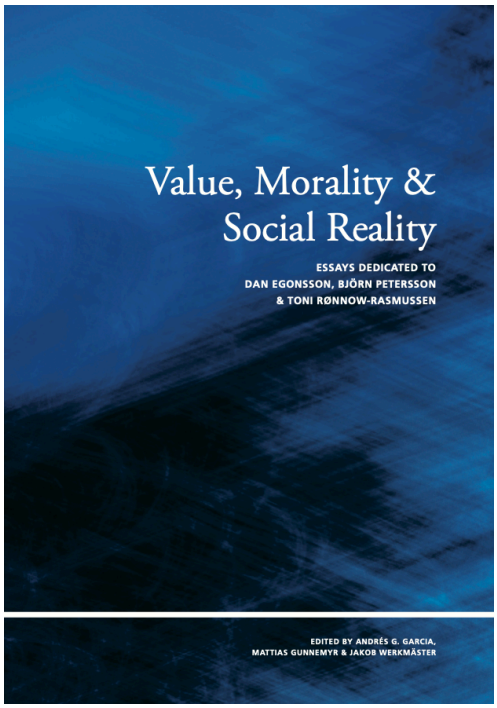
Collective Obligations and the Moral Hi-Lo Game

Kirk Ludwig

In: Garcia, A., Gunnemyr, M. & Werkmäster, J. (2023) *Value, Morality & Social Reality: Essays dedicated to Dan Egonsson, Björn Petersson & Toni Rønnow-Rasmussen.* Lund: Department of Philosophy, Lund University. DOI: https://doi.org/10.37852/oblu.189

ISBN: 978-91-89415-65-2 (print), 978-91-89415-66-9 (digital)



Published by the Department of Philosophy, Lund University.
Edited by: Andrés Garcia, Mattias Gunnemyr, and Jakob Werkmäster
Cover image by Fabian Jones. Cover layout by Gunilla Albertén.

DOI: https://doi.org/10.37852/oblu.189.c526

This text is licensed under a Creative Commons Attribution-NonCommercial license.
This license allows reusers to distribute, remix, adapt, and build upon the material in any medium or format, so long as attribution is given to the creator. The license does not allow for commercial use.
(License: http://creativecommons.org/licenses/by-nc/4.0/)

Text © Andrés Garcia, Mattias Gunnemyr, and Jakob Werkmäster 2023.
Copyright of individual chapters is maintained by the chapters' authors.

Collective Obligations and the Moral Hi-Lo Game

Kirk Ludwig

Introduction

Olle Blomberg and Björn Petersson (2023) argue that collective moral obligations, at least in some cases, are irreducibly collective. By this they mean the subject of the obligation is a group and their having a moral obligation collectively cannot be analyzed into individual obligations of its members to do their parts in what the group has an obligation to do. The main argument focuses on a choice situation that looks like a moral Hi-Lo game, in which we have the intuition that the group is responsible for pursuing the best moral outcome. Blomberg and Petersson argue that we cannot account for this intuition by deriving it from individual obligations of the parties to do their parts in bringing about the best moral outcome. In contrast, I will argue that the case has not been made and that we can plausibly account for the intuition that the group has a moral obligation while seeing it as grounded in the independently derived obligations of the members to do their parts.

We typically attribute obligations to informal groups with plural referring terms as in [1].

> [1] We ought to save the children

[1] is ambiguous between a distributive and a collective reading. On the distributive reading, [1] is understood as equivalent to [1d].

> [1d] Each *x* of us: *x* ought to save the children.

On the collective reading, in contrast, the group is, in some sense, the locus of the obligation, as in [1c].

> [1c] We are such that we ought to save the children (working together).

Just as an individual obligation requires of its subject, *ceteris paribus*, action to fulfill the individual obligation, so group obligation requires of its "subject", *ceteris paribus*, collective action to fulfill the group obligation. I put "subject" in scare quotes in the second clause because at this point we do not want to assume that [1c] involves commitment to the group per se being the bearer of the obligation rather than it distributing obligations to its members *to contribute to their jointly saving the children*. We make attributions of collective obligation when

> (a) group action is necessary in order to bring about a moral good or avoid a moral harm or
>
> (b) group action, even if not necessary, nonetheless will be more effective or carry less risk of failure in pursuing a moral good or in avoiding of a moral harm.

One can take two different stances on collective obligations. First, one can regard the collective obligations of groups as grounded in their individual obligations. This is the bottom up approach and entails that collective obligations are reducible to individual obligations to contribute to collective action. Second, one can regard collective obligations of groups as primary and any individual obligations, when present, as derived from them. This leaves open that collective obligations may not always entail or be accompanied by individual obligations. This is a top down approach. I endorse the bottom up approach.

The Case Against Reduction

Blomberg and Petersson make a case for the top down approach by appeal to cases. Here is the central example (2023, 1; page number citations are to the online first version of the paper).

> *Burning Building*: Three children are trapped in a burning building. One of them is in one room, and the other two are in a second room some distance away. The neighbours Agnetha and Benny see each other approaching the building from opposite sides. Agnetha breaks in and has enough time to do her part of rescuing either the child in the first room or the two children in the second room. The rescue can succeed only if Benny heads straight for the same room with his fire extinguisher. If both go to the first room, they will only rescue the first child. If both go to the second room, they will only rescue the two other children. If each goes to a different room, no child will be rescued. Suppose that Agnetha and Benny can make these choices without any significant risk to their own or each other's life or health. All this is common knowledge between them, but they do not have any opportunity to communicate with each other—each must choose which room to head for independently of the other. (Adapted from (Colman, Pulford, and Lawrence 2014, 36))

Blomberg and Petersson share the intuition that Agnetha and Benny have an obligation jointly to save the two children. They call the basic intuition, following

Schwenkenbecher (2019, 28). The reason Agnetha and Benny have a collective moral obligation is that: "Only together do they have the ability to rescue the two children, and rescuing the two children is the best they can do, morally speaking" (2023, 2). Blomberg and Petersson aim to vindicate the basic intuition in the face of a puzzle about how it is possible. The puzzle is that it seems that for the group of Agnetha and Benny to have an obligation, they must be together an agent. But the only agents present are Agnetha and Benny, who can't communicate and must act independently. Blomberg and Petersson argue that it suffices for them to have a collective obligation that they are each able to ask not just "What ought I to do?" but "What ought we to do?" More specifically, what is necessary for this is that Agnetha and Benny are able to (i) identify with the group, (ii) to "we-frame" the situation, and (iii) deliberate about what the group ought to do. They develop this account drawing on Michael Bacharach's development of the team reasoning framework in decision theory (Bacharach 2006). The argument that collective obligations are not analyzable into individual obligations of its members rests on the claim that the latter are answers to the question "What ought I to do?" and that that starting point is insufficient to recover the basic intuition in cases like Burning Building.

Blomberg and Petersson note that Stephanie Collins (2019, 140) has argued that the individual agents can reason from individual obligations to participating in group action. Blomberg and Petersson claim that this cannot vindicate the basic intuition.

> We will now explain why Agnetha and Benny would not be able to have a collective deliberative obligation if they could only ask and answer the question: "What ought I do?" We claim that if Agnetha starts by asking what she ought to do in Burning Building, then she cannot rationally settle on any determinate answer. The same is true of Benny. They could therefore not have any collective deliberative obligation to save the two children. But why is this so? (2023, 5)

The core of the argument is that Agnetha and Benny face a moral Hi-Lo game. In a Hi-Lo game, there are two Nash equilibria (if the other(s) maintain their choices, no one has an incentive to change theirs), but one of the two has a higher payoff for both. In Burning Building, the structure is represented in Figure 1 (the first number in each box represents Benny's payout, the second Agnetha's).

Moral Hi-Lo Game		Agnetha	
		Room 1	Room 2
Benny	Room 1	1, 1	0, 0
	Room 2	0, 0	2, 2

Figure 1: Hi-Lo moral dilemma. Agnetha and Benny can save 1 or 2 children if both go to room 1 or room 2, but none otherwise.

The Hi-Lo game is a problem for classical game theory, which treats agents as individual strategic reasoners who make choices in the light of the choices that others make or are likely to make. In a Hi-Lo game, if you do not have any evidence bearing on what choice the other will make, it seems you can only engage in conditional best response reasoning. Room 1 is best for Benny if Agnetha goes to room 1, but room 2 is best for Benny if Agnetha goes to room 2. *Mutatis mutandis* for Agnetha. Thus, in the moral Hi-Lo game, if Benny asks "What ought I to do?", it seems that his answer will depend on how Agnetha answers the question "What ought I to do?" For each should aim for the morally best outcome, it seems, *given what the other does*. But then they seem to be at an impasse, and neither can reach a rational decision about what he or she ought to do.[1]

The argument that an individualist account of collective obligation cannot accommodate the basic intuition can be put as follows.

1. Agnetha and Benny have a collective deliberative obligation to save the two children in room 2 only if Agnetha and Benny first have individual obligations to contribute to their saving the two children in room 2.
2. Agnetha and Benny have individual obligations to contribute to their saving the two children in room 2 only if each can reason correctly that the answer to the question 'What should I do?' in Burning Building is to go to room 2.
3. Neither Agnetha nor Benny is in a position to reason correctly that the answer to the question 'What should I do?" in Burning Building is to go to room 2.
4. Therefore (from 1-3), Agnetha and Benny do not have a collective deliberative obligation to save the two children in room 2.

Premises 1 and 2 express the individualistic approach to collective moral obligation. Premise 3 is supported by the claim that they cannot move past conditional best response reasoning in Burning Building if they are focused on the question "What should I do?" The conclusion 4 contradicts the basic intuition. To hold onto the basic intuition, then, we must give up the conjunction of 1 and 2, that is, the individualistic approach to collective moral obligation. I will call this the moral Hi-Lo argument. If this is correct, then the basic intuition cannot be accounted for by the bottom up approach, which takes individual obligations to be explanatorily basic.

In cases of decision making under uncertainty, the indifference principle suggests giving each option equal weight. Assume this is a rational strategy. (Surely it is better than doing nothing when doing nothing is guaranteed to have no payoff and there is no cost to action.) In the Hi-Lo game, this gives choosing Hi (room 2 in our case) a higher expected utility ($.5 \times 2 > .5 \times 1$). This would provide a way of moving past conditional best response reasoning which leaves Agnetha and Benny at an impasse and thereby show that premise 3 is false.

[1] For development of this idea see (Sugden 2000, 179-182, Bacharach 2006, 35-68).

However, Blomberg and Petersson argue that we should doubt that the indifference principle is a sound general principle for reasoning in the absence of information about probabilities, on the grounds that its application to some Stag Hunt games leads to the wrong result. In a Stag Hunt game, to take a simple case, two hunters have the option of hunting rabbits or stags. There's one stag in the hunting range and a number of rabbits. One stag yields as much nutrition as six rabbits. But a successful stag hunt requires the hunters to cooperate. We suppose they can't communicate and must make a choice of proper hunting tools before going on the hunt. If one chooses stag and the other rabbit, the one who chooses rabbit will be able to trap two rabbits in the time available but the other can hunt neither stags nor rabbits. If both choose rabbit, there are enough rabbits in the range for each to trap two. This game structure is shown in Figure 2.

Stag Hunt		Player 1	
		Stag	Rabbit
Player 1	Stag	3, 3	0, 2
	Rabbit	2, 0	2, 2

Figure 2: Stag Hunt. There are two Nash equilibria, Stag-Stag and Rabbit-Rabbit.

In this game, the indifference principle doesn't help to choose the payoff dominant Nash equilibrium because the expected payout is higher for each in choosing Rabbit (.5 x 2 + .5 x 2 > .5 x 3 + .5 x 0, that is, 2 > 1.5). Here Blomberg and Petersson assume that it is, contrary to the result of applying the indifference principle, not irrational[2] for the hunters to play (stag, stag) rather than (rabbit, rabbit).[3]

[2] While I will not appeal to the indifference principle below in defending a reductive account of group obligation, it is worthwhile asking whether it really delivers the wrong result in this case. Why is it supposed to be rational for them to hunt stag rather than rabbit? The intuitive idea is that they'd be better off, so surely if they both choose stag, they cannot be charged with being irrational. But doesn't this depend on what they think about the other? If player 1 thinks player 2 is risk averse, then it would not be rational for player 1 to choose stag. If player 2 thinks player 1 thinks player 2 is risk averse, it would not be rational for player 2 to choose stag. So intuitions about whether it is rational for them to choose stag depend on how we fill in the picture about what they think about the other. If each thinks the other is likely to choose stag, then it is certainly rational for each to do so. But what about the case in which they have no idea what the other thinks or is disposed to do? Is it rational to assume that the other is more likely to choose stag than rabbit or vice versa? Neither, evidently. It is in this circumstance that the indifference principle looks like a reasonable basis for a decision. Blomberg and Petersson think that this gives the wrong verdict, and that it is in fact rational for the players to choose (stag, stag). But why? They think group identification and team reasoning can make sense of it, for if both team reason, they are better off. But it is reasonable for each to team reason only if they each have good reason to think the other is doing so as well. But in this case they do not.

[3] These are not the only options. Perhaps it is a reasonable principle, for example, to choose the Pareto optimal equilibrium in cases in which there is more than one Nash equilibrium and they are strictly ordered by Pareto dominance. An outcome *x* Pareto dominates outcome *y* iff it *x* is strictly better for at

Blomberg and Petersson also argue that the indifference principle and the common knowledge of rationality requirement would lead Agnetha and Benny to have contradictory beliefs. If each knows the other applies the principle, each can then reason that the probability of the other going to room 2 is higher than 50%. But this, they say, leads Agnetha and Benny to have contradictory beliefs: each thinks the probability that each will go to room 2 is 50% and also higher than 50%.

A second objection to a reductive account of collective obligation focuses on whether starting from the question "What should I do?" Agnetha and Benny could have normative reasons to save the two children.

> Plausibly, the subject of an obligation to Φ must not only have the ability to Φ, but also the ability to Φ for the normative reasons that make Φ-ing morally obligatory (see (Lord 2015); cf. Collins 2019: ch. 3). Agnetha and Benny would lack this ability if each of them were restricted to asking and answering the question: "What ought I do?" In its first-person singular form, the deliberative question concerns what to do solely on the basis of the person's own agentive abilities. Hence, if Agnetha were limited to "I-reasoning," then she would not have a normative reason to do her part in saving the two children. Nor would Benny if he were limited in the same way. Only together can they have a normative reason to save the two children (cf. (Dietz 2016, 960-963)); and only if they together have this normative reason will each of them have a normative reason to do their part. … they can therefore at most each have an obligation to do their part in a collective endeavour such as a joint rescue. (2023, 8-9)

I take the central assumption here to be

> *I-Reasoning*[4]: Reasoning in the first person always concerns only what the agent can do by herself with no contributions from others.

A secondary assumption is a cousin of the Ought Implies Can principle:

> *Having a Normative Reason to* φ *Implies that You Can* φ (*NORIC*):
> If one cannot do something, then one cannot have a normative reason to do it.

Then, since neither Benny nor Agnetha can alone save any children, I-Reasoning would preclude them from seeing anything they can do that would save any children, and, hence, by NORIC, they cannot have a normative reason to save any children. Hence, they cannot have an obligation to do so, since they can have an obligation to

---

least one player than *y* and no worse for any. An equilibrium is Pareto optimal iff no change can make anyone better off without making someone worse off. This principle would select (Hi, Hi) in the Hi-Lo game. See (Harsanyi and Selton 1988).

[4] Blomberg (personal communication) states that he would not endorse this principle in general but maintains it is true about I-reasoning in Hi-Lo. Substituting 'in Hi-Lo' for 'always' will not affect the critical remarks below.

do so only if they can act on a normative reason to do it. I'll call this the argument from I-reasoning.

Blomberg and Petersson argue further that the key to Benny and Agnetha having a collective obligation is their having the capacity to we-reason, to adopt the team reasoning perspective in which each thinks in terms of maximizing the (moral payoff) for group action, not for individual action, and then derives from that what they ought to do as part of the team. This makes obligation relative to agential perspective. They argue that in some situations the "I"-relative oughts and "we"-relative oughts require or permit different responses, and this yields a kind of moral incommensurability. The capacity to we-reason is itself context relative because the capacity for group identification, which is the key to we-framing their decisions, will be affected by the context.

These further conclusions rest on the idea that the basic intuition that Benny and Agnetha have a collective obligation in Burning Building cannot be accommodated by starting with a standard picture of obligations that individuals have. In the following, I will argue that the arguments against reduction are not successful.

Reply to the Hi-Lo Argument

The basic intuition is that in such cases as Burning Building agents can be morally obligated to do something as a group, something we express by saying that *they* have an obligation to do something together. In particular, Benny and Agnetha have an obligation to save the two children by their going to room 2 because that is the morally best outcome. This cannot be taken as a judgment that they have an irreducibly collective obligation without begging the question. Our question is whether the basic intuition can be vindicated without accepting that the attribution of collective responsibility is irreducible.

It will be helpful to begin with a reductive account of collective obligation. We may put it this way (for pro tanto obligation add 'pro tanto' before 'collective obligation' on the left hand side of the biconditional and before 'moral obligation' in (b)):

> *Collective Obligation as Distributed Obligation to Engage (CODE):*
> A group *G* has a collective obligation to *J* in context C if and only if in C:
>
> (a) *G J*-ing is necessary in order to bring about a moral good or avoid a moral harm
> or
> *G J*-ing will be more effective or carries less risk of failure in bringing about a moral good or in avoiding of a moral harm, and
>
> (b) each member *x* of *G* individually has a moral obligation to contribute to *G J*-ing which is not derived from *G* having independently a collective obligation to *J*.

The Hi-Lo Argument aims to show that our intuition in Burning Building cannot be vindicated by appeal to the CODE conception of collective responsibility because, the argument goes, condition (b) is not met in that case.

The moral Hi-Lo game is supposed to present a problem for Benny and Agnetha if they are thinking in terms of their individual moral responsibilities. The problem is supposed to be that what each should morally do, if asking 'What should I do?', depends on what the other in fact does. Each is able, the claim is, only to reason, "If the other goes to room 2, I ought to go to room 2; if the other goes to room 1, I ought to go to room 1."

The scenario stipulates that they can't communicate and that the basic setup is common knowledge between them. But in many other ways the case is under described. Therefore, the intuition that they have a collective obligation to save the two children may rely on how we fill in the case. This will be relevant also to what resources they have for reasoning in the first person about what they should do. In the case of Burning Building, how do we fill in our understanding of what Benny and Agnetha are likely to know or believe about each other? We likely think that if they could communicate, they would confirm with each other their understanding of the situation and what it is best to do—which is obvious in this case—and then do it. That is why we stipulate that they can't communicate.[5] But that already shows that we are thinking that both of them have pro moral attitudes and expect that the other does as well. If this is our default view of their characters and attitudes, then we will also suppose that each of them assigns a relatively high likelihood to the other *being motivated to achieve the best moral outcome in the situation*. Let us say that each assigns a probability of .8 to the other aiming for the best moral outcome (the exact number turns out not to be material). We can assume that each knows or has a subjective probability close to 1 (let's say 1) that he or she wants the best moral outcome. Then each will assign a probability of .8 that they both want the best moral outcome. They know that the best outcome is saving the two children in room 2 and that working together is necessary to bring it about. So each will know, or have good reason to believe: each of us wants to cooperate on saving the two children in room 2 because that is the best moral outcome. Neither will think: we want to cooperate on saving the one child in room 1. The result of their communicating would be to confirm this. But communication is not necessary for them to have good reason to believe it. They will thus be able to identify in the circumstances a goal that each knows that each wants to promote over other things that they can do and that it is what they would agree to do if they could communicate. Thus, each has a *reasonable expectation* that his or her going to room 2 will promote what is morally the best outcome in the circumstances. (The reasoning here does not start by adopting the perspective of the group and so is not team reasoning; rather it draws on background

[5] Of course, one of them, Agnetha, say, might simply announce that she is going to room 2, counting on Benny to follow suit since it settles what he should do. But that shows also that she already thinks he wants what is morally best as she does.

information about the circumstances to increase confidence in both that the other will go to room 2.) Each has an individual obligation to contribute to the best moral outcome they can achieve in the circumstances. Therefore, each has an obligation to do his or her part in their rescuing 2 children in room 2. Thus, they have a collective obligation to rescue 2 children in room 2. This bit of reasoning vindicates the basic intuition, but it does so by deriving individual obligations from their expectations about the other sharing the goal of achieving the best moral outcome, which is a natural way to fill in the background in Burning Building, and so via satisfying (b) in CODE. This shows that premise 3 in the argument above from a reductive account of collective obligation to Agnetha and Benny not having a collective obligation in Burning Building is false.

We standardly expect, when we have common interests with others which we can only achieve by working together and there are no costs in cooperating, that the other will cooperate with us, that is, that we will cooperate. Imagine a Burning Building scenario where children are in only one room. There is still a question about what one should do when it requires two to rescue them, but you cannot communicate. Your expectation is that the other will contribute to bringing about the best moral outcome and you act on that assumption. Burning Building differs from this case in having not just two possible outcomes (save the children or don't) ranked in terms of their moral desirability but three (save two, save one, save none). Yet on the assumption that you both will act from moral motives, it is obvious that the morally right thing to do is to rescue as many children as you can.

The methodological point is that our judgements about a group being collectively responsible in a scenario is very likely to be conditioned by our making default assumptions, which reflect a natural way of filling in the background, when nothing contrary is explicitly stated in the description of the scenario (and sometimes even when it is). The natural background for Burning Building (and one implied by how it is framed in terms of a matrix that represents only moral goods and the exclusion of communication) is that each of Agnetha and Benny want to do the right thing and believe this about each other. If the intuition rests on this way of filling in the background, then the basic intuition can be recovered from the individual perspective without difficulty.

Blomberg and Petersson might level the objection that whatever assumptions Agnetha and Benny make about each other's tendency to go to room 2 will lead to their having contradictory beliefs about probabilities when combined with the assumption of common knowledge of rationality and common knowledge of the circumstances. The argument has the same form as in the case of the appeal to the principle of indifference. If each knows the other has reason to assign a probability p to the other's going to room 2 and that that yields a higher expected payoff than going to room 1, then the probability that the other will go to room 2 should be higher than p. In fact, this style of objection, if successful, would work no matter what the probability (other than 1 or 0) they each assign to the other's doing something, as long as it selected one of the options as superior to the other. The

problem, if there is one, is also independent of the Hi-Lo game. It arises even if there is just one room with children in a burning building. Yet, surely in this case they need not have contradictory beliefs.

There are a number of threads to disentangle here.

(1) The reasoning I sketched above did not in the first instance assign a probability to the other going to room 2 but instead to the other aiming for the morally best outcome. If they each have reason to assign .8 to the other aiming at the best outcome, and this is common knowledge, then they can each reason that there is a .8 probability that they both aim at the best moral outcome. As they can each see that the other will reason in this way, it will be common knowledge among them that there is a .8 probability that they both aim at the best moral outcome. On the assumption they both aim at the best moral outcome, they should each go to room 2. It is more reasonable for each to adopt that assumption than any other. Given that they are rational, they will adopt the more reasonable assumption. Assume this is all common knowledge. Then each knows the other assumes they both aim at the best moral outcome and so knows that each believes that he/she should go to room 2. If they each believe that he or she should go to room 2, then each of them will. Assume each believes this. Then each of them believes that each of them will go to room 2 and each of them will. I have not here represented the reasoning in the last stages as involving a calculation of an expected value for going to room 2, but as a matter of adopting an assumption for the purposes of determining what to do. Subsequent reasoning takes it as a fixed point.

(2) Assume, however, that at some point each assigns a probability to the other going to room 2 that yields going to room 2 as having the higher moral expected value. If the reasoning proceeds in this way, each has sufficient normative reason to go to room 2. Further reasoning is superfluous. What matters for action is not the size of the difference in expected value between going to room 2 and other options, but whether it is higher. If it is, then each can just act without further reflection, and they would not be involved in any incoherence, if there is any threat of it, from further reflection.

(3) However, even if they do engage in further reflection, it should not result in their holding inconsistent beliefs. Either the reasoning that is supposed to generate a greater confidence level takes some time or it takes no time. If it takes some time, then inference to a higher probability than initially assigned will require a simple update of the probability and the agent does not need to hold contradictory beliefs.[6] If it takes no time, it is not an inference at all. All we could have in mind is that the agent's beliefs are from the beginning where they should be on working out the consequences of all their basic beliefs and assumptions. But this would not involve any contradictory beliefs.

[6] Blomberg and Petersson suggest that the reasoning would amount to assuming the probability that the other will choose one option is *p* and then inferring from that that the probability is different from *p*, which might be taken to be a reductio of the original assumption. But this need not be how it

One might grant that the intuition *in Burning Building* can be generated by thinking of each of Benny and Agnetha having individual obligations to do their parts in saving two children based on calculation of the expected value of going to room 2 versus room 1 given the background assumptions sketched above, but deny that this generalizes on the grounds that we can always find a *Moral Stag Hunt* which can't be solved no matter how confident (short of certainty) the hunters are of the other(s) hunting stag. (I am not attributing this response to Blomberg and Petersson, but it will be instructive to consider it.)

First, let's assume that the reasoning of Agnetha and Benny must proceed by assigning a probability of .8 that the other will choose the morally best option. Then it seems that there is a set of payoffs in the Stag Hunt game that will make it less rational morally to hunt stag than rabbit even if one knows that the others are .8 likely to hunt stag. For example, if you know that the payoff for hunting stag if the others do is 10 and 0 otherwise, but 9 if you hunt rabbits, then the expected value for hunting stag is .8 x 10 = 8 but for hunting rabbit, which is a sure thing, it is 9. See Figure 3.

Moral Stag Hunt		Player 1	
		Stag	Rabbit
Player 1	Stag	10, 10	0, 9
	Rabbit	9, 0	9, 9

Figure 3: Moral Stag Hunt. Even with a .8 probability that the other will choose Stag, hunting Rabbit appears better for each.

In general, for any probability assigned to the others' hunting stag, we can find a ratio of payoffs for hunting stag versus rabbit in the stag hunt scenario that will have the result that hunting rabbit is better than hunting stag. Let s be the utility of hunting stag when all parties do, r be the utility of hunting rabbit alone, and p be the probability that the others will hunt stag. If $r > s \times p$, then the expected value for hunting rabbit will be greater than hunting stag. For any $p < 1$, for any s, there is an r such that $r < s$ and $r > s \times p$. Thus, the argument goes, no matter the probability assigned to the other hunting stag, we can find a payoff structure in which the

proceeds. One can reason as follows. The antecedent probability, without taking into account how the other will reason, that the other will go to room 2 is .5. If it is .5, then going to room 2 is best. Therefore, I should go to room 2. But the other will reason in the same way that I just did. If the other reasons in the same way I just did, then the other will reach the conclusion that going to room 2 is what he should do. Given that the other is rational, that is what he will do. Therefore, taking into account additionally how the other will reason, the probability that the other will go to room 2 is 1. This is a revision of the assignment on the basis of taking into account additional evidence relevant to the assignment of the probability, namely, how the other will reason. There is nothing incoherent in assigning a probability of .5 to **α** given background knowledge K and assigning a probability of 1 to **α** given K + **δ.**

morally best outcome is not chosen from the standpoint of individual moral reasoning.

But *not so fast!* There are two points we need to attend to. The first is what we mean by saying the utilities are interpreted as moral outcomes—we need a story that makes sense of this. The second is how we are to think of the utilities for the two players given that they represent moral outcomes.

On the first point, it is hardly clear that the group is morally required to maximize the amount of meat summed across all hunters or to maximize the amount each takes home from the hunt. To construct a case where it seems that there is a clear forward looking moral obligation to work together, let us suppose that the purpose of hunting is to procure food to save children from dying of starvation. Let's suppose that each unit of value in the matrix represents enough meat to save one child from starvation. Then if Agnetha and Benny both hunt stag, 20 children are saved. If they each hunt rabbit, 18 children are saved. If either hunts stag and the other rabbit, then nine children are saved.

On the second point, the good for each contributor is the total number saved, just as in Burning Building, for they aim for the moral good, and so all should share equally in the good that results from the intersection of their choices. Given this, for each the value of the intersection of any pair of actions is the total number of children saved. This alters the payouts as shown in Figure 4.

Moral Stag Hunt Revised		Player 1	
		Stag	Rabbit
Player 1	Stag	20, 20	9, 9
	Rabbit	9, 9	18, 18

Figure 4: Moral Stag Hunt revised. 20 children are saved if both hunt stag, 18 if both hunt rabbit, and 9 if one hunts rabbit and the other stag.

But now the altered payoff structure gives us a different result. In effect, this transforms it into a Hi-Lo game.[7] Give the payoffs, as long as the probability that

[7] Blomberg and Petersson agree with this conclusion (see p. 30), which is reached on the basis of the assumption that moral payoffs are agent-neutral in the sense that it is irrelevant who brings them about. They suggest, however, that we can generate a genuine moral Stag Hunt scenario if morality is, at least in some cases, agent-relative (see p. 31) in the sense that one can assign a higher moral value to saving someone oneself rather than another doing it. The common-sense argument for agent-relativity is that we can assign a higher moral value to helping someone to whom we have a special relation, such as a child or spouse, than to someone else doing so. Blomberg and Petersson do not take a stand on whether morality can be agent-relative. They aim rather to illustrate how the I-perspective and team-perspective may ground different moral judgments about what one ought to do, if morality is agent-relative. However, even if we grant morality can be, in some cases, agent-relative, in the sense that an agent may place greater moral value on saving someone herself than another doing so (saving your child for example), it is not clear that this avoids the collapse into a Hi-Lo game. For if it is better for a mother to save her child than a stranger because of her special relation to her child, then plausibly that is better

the other hunts stag is greater than 9/20 (.45) it will be best to hunt stag. Thus, for example, simply applying the principle of indifference would weigh the options equally, and then the payout for choosing to hunt stag is 20 x .5 + 9 x .5 = 14.5 which is greater than the payout, 9 x .5 + 18 x .5 = 13.5, for choosing to hunt rabbit. If we can assume, as we did above, the likelihood is higher than .5, the gap in the expected values will be greater. So the result is not that *individually* they morally ought each to hunt rabbit when they are asking 'What ought I to do?' but instead that they should each hunt stag with the other. The central point is that if we think of the payoffs as representing the moral good, then positive outcomes are the same for everyone (see note 8 on whether agent-relative morality makes a difference). When *N* children are saved, that is not less morally good when I don't save them. Thus, we don't get the standard structure of the stag hunt when we think about the outcomes as in terms of what is morally best. I will say more about the significance of this below.

I'll return to some other ways of filling in the scenario in Burning Building where Agnetha and Benny may have more reason to think the other *will not* help at all or not go to room 2. We will see that this does not help the argument. But before that it will be useful to turn to the argument from I-reasoning. For what I have just described may sound as if it is already we-reasoning, and that if Benny and Agnetha were really restricted to thinking about their individual obligations and reasoning about what they could do, they could never arrive at the conclusion that they were each obligated to contribute to their saving two children.

Reply to the Argument from I-reasoning

The two assumptions of this argument were I-reasoning and NORIC, repeated here.

> *I-Reasoning:* Reasoning in the first person always concerns only what the agent can do by herself with no contributions from others.
>
> *Having a Normative Reason to* φ *Implies that You Can* φ *(NORIC)*:
> If one cannot do something, then one cannot have a normative reason to do it.

---

from everyone's perspective, not just the mother's. But then that extra value will simply be added to the sum total good brought about by the combination of actions, leading to a transformation of the matrix into a Hi-Lo game (preserving symmetry for the players). That is, everyone sees the value added when someone helps those they have a special relation to. Agent relativity would have to be interpreted as meaning not just that an agent may assign a greater value to helping someone she has a special relation to, but also that that fact is morally irrelevant to everyone else. But *that* does not seem to be of a piece with common sense morality. We surely do all agree that, other things being equal, it is better for parents to feed their children than strangers, for example, because of their special relation to them.

I think that both of these assumptions are mistaken.

I-Reasoning is false because in thinking about what I can bring about, I do not need to exclude calculations about what others are doing or would do given what I do. Given any goal I have, I can ask myself what I can do to achieve it. This is reasoning in the first person, but it precludes no means to the end. Sometimes I can get what I want by getting others to do things whether they know that is my intention or not. I can clear a building by setting off a fire alarm. Clearing the building requires contributions from others. But what I do gets them to make their contributions. I can also get things done by working with others. If I want to move a large table from one room to another, one answer to the question 'What should I do?' is to get some help. So reasoning in the first person does not always concern only what the agent can do by herself with no contributions from others. So when someone asks, "What ought I to do?" in circumstances like Burning Building, the answer can be that I should work with someone else to achieve the best moral outcome. That is a perfectly fine answer to the question about individual obligation. That gives one a normative reason to contribute to saving the two children.

NORIC is controversial.[8] I believe that it is false. There is not space here to do justice to the issues, but here is a quick brief. We have both positive and negative moral duties. A positive moral duty is a duty to do something, either to bring about a good or prevent a harm, construed broadly. A negative moral duty is a duty not to do something because of the harm it will bring about. These general duties are standing requirements. If I am not in a position to fulfill a duty, it does not disappear. What the lack of ability to fulfill a duty pertains to is whether it is proper to blame me for not fulfilling it. The duty I have to prevent harm gives me a normative reason to save children in a burning building even if I am the only one around. I have that reason even if I cannot act on it because I do not have the ability alone to do it. This is true of all our reasons. A reason to do something is a consideration that speaks in favor of it (it need not speak decisively in favor of it because we can have conflicting reasons). That I cannot do something does not entail that I have no reason to do it, that is, that nothing speaks in favor of it. If I develop a craving for water in the desert, I have a reason to drink some water. If there is none around, I can't act on it. But I am motivated to look for some because I have a reason to drink some water.

[8] Bart Streumer (2007) and (Lord 2015) argue it is true; Ulrike Heuer (2010) and Kimberley Brownlee (2010) argue it is false; Streumer replies to Heuer in his (2010). One could *insist* that a consideration that speaks in favor of something isn't sufficient for there to be a reason in favor of it (here we do not mean all-in or everything considered) but that there must also be some prospect of acting on it successfully. Suppose we call these c-reasons. By definition you don't have c-reasons if you can't do what they are putative reasons for. But now we should reject the claim that you can't have an obligation to do something if you don't have a normative c-reason for it. For sometimes we fail to fulfil duties we indisputably have, such as caring for our children, because we can't do it, for example, because we suffer from a debilitating fatal disease. In this case, we would have the duty but no normative c-reasons to do it. We are not blamable in this case, but it is not because we fail to have any duties to our children.

If I find an oasis, then my reason becomes practically relevant, but it was there all along, as evidenced by its motivating me to look.

Whatever the verdict on NORIC, the argument from I-reasoning fails because in Burning Building Agnetha and Benny are not restricted in first person reasoning to thinking about what they can do alone.

Burning Building Redux

Return to the case of Burning Building. Surely all that is needed is a set of circumstances, even if it is not the default way of filling in the case, in which we judge that Agnetha and Benny are collectively obligated to save the two children, but it is not the case that each can arrive at an obligation to do their parts by answering the question "What ought I do?" Suppose that Agnetha and Benny live in a dystopian future in which there is extreme scarcity and almost everyone has developed a reflexive tendency to mind their own business. Agnetha and Benny know that most people would not make any attempt to save the children even if they could do it alone. They know nothing about each other. They can't get any information about what the other is doing before they actually undertake to arrive in room 1 or room 2. Thus, each (Agnetha/Benny) has more reason than not to think the other (Benny/Agnetha) will not contribute even knowing the other (Agnetha/Benny) wanted to save the children. We'll call this the Dystopian Burning Building case.

Let's grant that they collectively have an obligation to save the children, that is, the answer to the question 'What should they do?' is 'They should save the two children'. They also then collectively have a normative reason to save the children. However, given what I said above about the NORIC principle, we should also grant that they have a normative reason to do their parts in saving the children. Perhaps this then supports also the judgment that they, in some sense, still have an obligation to do their parts in saving the children.

Still, each of them has *most* reason to think that *they can't fulfill that obligation*. So we might focus on the question whether there is a contrast between what they have most normative reason to do assuming they act together and what each has most normative reason to do individually. Perhaps each of them has most reason not to do anything, given the high probability that the other will not do anything. The group perspective might be said to be expressed by the question, 'What do we have most reason to do together?' If the answer to this question is to save the two children, then we apparently get a contrast in judgments about what the group has most normative reason to do and what its members, reasoning from the first person, have the most normative reason to do.

In response, first, although it is clear that each has most reason to think *they cannot fulfill* an obligation to contribute to their saving the children in Dystopian Burning Building, it does not follow that each does not have *most reason to (undertake to) do their part in their saving two children*. Benny believes it is unlikely that Agnetha will make any effort to save any of the children. Still, it is not out of the question for Benny that she will. Since Benny's going into the Burning Building does not involve any risk to himself, it still makes sense for him to go to one of the rooms on the chance that Agnetha will go to one of them. But which one? While from Benny's point of view, it is epistemically unlikely that Agnetha will act to save any of the children, if she does, it is because she is motivated by moral considerations. In that case, she is more likely to go to room 2 because that is morally the best outcome, as we argued above. So Benny should act to go to room 2, even if the chances of success are low. Mutatis mutandis for Agnetha. The decision matrix in this case includes three options: go to room 1, go to room 2, and don't do anything. This matrix is represented in Figure 5.

Moral Hi-Lo Game 2		Agnetha		
		Room 1	Room 2	Nothing
Benny	Room 1	1, 1	0, 0	0, 0
	Room 2	0, 0	2, 2	0, 0
	Nothing	0, 0	0, 0	0, 0

Figure 5: Hi-Lo with the option of doing nothing.

Given that, from the point of view of Benny/Agnetha there is some chance that Agnetha/Benny will go to room 1 or room 2, doing nothing will have a lower expected value that going to one of the rooms. Once the choice is made to set aside doing nothing, this reduces the decision matrix to the original. So in fact Dystopian Burning Building does not introduce any essentially new elements.

When Agnetha Knows that Benny Will Go to Room 1

It might still be argued that we can get a divergence in what obligations Agnetha and Benny have together and individually by considering a case in which it is common knowledge among them that one will go to room 1, no matter what.[9] The

[9] These arguments are not the ones that Blomberg and Petersson offer for individual and collective obligations coming apart, but theirs depend on the idea that team reasoning may generate obligations independently of individual obligations to do their part. This rests in part on the idea that we can't recover the basic intuition about Burning Building, which I have disputed. I'll also argue in the last section that team reasoning has to be justified before we engage in it, and so can't function as an

thought is that since they clearly can together save two children and not just one, and that is strictly morally better, they ought to save the two children. But if, for example, Agnetha knows with certainty that Benny will go to room 1, then her moral obligation is to go to room 1. Surely if Agnetha has an obligation to do her part in their saving the child in room 1, then she does *not* have an obligation to do her part in their saving two children. But as *they* have an obligation to do that, group obligation comes apart from individual obligation, and group obligation cannot be reduced to obligations of the members to do their parts.

The argument here goes as follows:

(1) Agnetha and Benny have an obligation collectively (together) to save two children by their going to room 2 and cooperating in removing the two children from the burning building.

(2) It is common certain knowledge among Agnetha and Benny that Benny will go to room 1 to try to save the child there.

(3) If Benny will go to room 1, then the morally best outcome that Agnetha can contribute to in that circumstance is saving the child in room 1 by going to room 1 and helping Benny to rescue the child there.

(4) If the morally best outcome that Agnetha can contribute to in the circumstance that Benny will go to room 1 is saving the child in room 1 by going to room 1 and helping Benny to rescue the child there and this is known to Agnetha, then Agnetha has an obligation to go to room 1 to help Benny save the child there.

(5) Therefore (2-4), Agnetha has an obligation to go to room 1 to help Benny save the child there.

(6) If Agnetha has an obligation to go to room 1 to do her part in saving one child with Benny, and that is incompatible with her going to room 2 to do her part in saving two children with Benny, then Agnetha does not have an obligation to go to room 2 to do her part in saving two children.

(7) Therefore (5, 6), Agnetha does not have an obligation to go to room 2 to do her part in saving two children.

(8) Therefore (1, 7), while Agnetha and Benny have an obligation collectively (together) to save two children by their going to room 2 and cooperating in removing the two children from the burning building, they do not both have an obligation to go to room 2 to do their parts in saving two children.

We have an obligation to prevent harm from happening through action or inaction. In the latter case, we are called on to act. In practice, we are limited by what harm

---

independent standpoint from which to generate collective moral obligations. The arguments I respond to in this section help to bring out another source of the intuition that collective and individual obligations come apart but I will show that in fact they do not come apart with respect to the same background circumstances.

we can foresee. Even then, we can't in general prevent all the harm that we can foresee. If what I have been saying is correct, then we still have the obligation, but we can't be blamed for what we cannot prevent (excepting non-cooperators who can contribute). But then what should we do when we cannot prevent all the harm we can foresee, that is, fulfill the strict requirements of duty? We should prevent as much harm as we can, taking into account what options are open to us. If I alone can save 1 or 2 children by going to room 1 or 2, but not all, then I should go to room 2 in those circumstances. If I cannot save the children in room 2 because it is inaccessible, I should go to room 1 in those circumstances. These are obligations of the form: I should **φ** given C. 'C' is replaced by a specification of the circumstances which delimit what you can expect reasonably to be able to bring about.

I want to allow that there is a sense in which (1) is true. And I want to allow that there is a sense in which (5) is true. But I will argue that the circumstances in which we understand (1) to be true are not the same as the circumstances that make (5) true.

Why is (1) true? Here we are thinking that each of them can do their part in their saving the children in room 2 and each of them knows (or ought to know) this and that it is the morally best thing to do *in the circumstances.* Here we do not include in the circumstances *what they intend or are willing to do and what they will or are likely to do*, that is, we bracket their intentions and likely actions. Call these circumstances C_1. So in C_1 they ought to save the two children and each of them ought to do their parts in that. (Here I assume that they can, relative to C_1, rationally arrive at the conclusion that they morally ought each to do their parts in their saving the children.) A crucial fact about these circumstances is that we have not included in them anything about the actual inclinations or intentions of either Agnetha or Benny. In thinking of what each *should do*, and so of what they should do, we treat their wills as *compliant to* the requirements of morality. For otherwise simply not intending or being willing to do something would alter the circumstances so that there was nothing they could do in the circumstances. We ask, bracketing what they intend or will actually do, given the circumstances, what are the requirements of morality on each of them: they each should go to room 2 and cooperate with the other in saving two children. So they ought to both go to room 2 and save the two children there.

Now consider Agnetha's position when she knows that Benny will not go to room 2 but will instead go to room 1. This adds an additional fact to the circumstances against which she evaluates what it would be best for her to do. Call these circumstances C_2. Fixing that Benny will go to room 1, the morally best outcome that Agnetha can contribute to is saving one child by going to room 1 and cooperating with Benny in doing that. Thus, in (5) the obligation Agnetha has is relative to C_2. In (1) in contrast the obligation Agnetha has is relative to C_1. When we then relativize the conclusion to the different circumstances, we get

> While Agnetha and Benny have an obligation collectively (together) in C_1 to save two children by their going to room 2 and cooperating in removing the two children from the burning building, they do not both have an obligation to go to room 2 to do their parts in saving two children in C_2.

But this is compatible with Agnetha having an obligation relative to C_1 to go to room 2 to cooperate with Benny in saving two children. So the argument does not show after all that they can have an obligation to save two children when Agnetha does not, relative to the same circumstances, which is what is needed to show that their obligation together can come apart from their obligations to do their parts.

What should we say about Benny in these circumstances? Benny is violating his duty in going to room 1. He is blamable for not choosing the best option, while Agnetha is not. Agnetha is not because she is responding to what he does, and she would go to room 2 conditional on Benny doing so. Benny's will is out of line with the requirements of morality; Agnetha's is not.

Can we answer the question what Benny and Agnetha should do together relative to C_2? There is a difficulty with this. As noted above, when we ask what a group of people should do when they can intervene to prevent some harm, we bracket what they intend or will actually do. We imagine that each will expect others to focus on what they can do together to achieve the best end. Each expects that the others, where the circumstances are public, will be willing to do her part in the ensemble of acts that will have the best outcome. Then we conclude each has an obligation to do her part in that. But if we ask about Agnetha and Benny in C_2, then the question is what each should do, given as a fixed point that one of them will go to room 1, full stop. The oddness of the question of what they should do lies in the fact that Benny has already decided what he will do. There is no question left for Benny to consider about what he should do in the circumstances in which he goes to room 1. It is a presupposition of asking what he should do regarding some matter that we don't include what he does. This then is also a presupposition of the question of what they should do in the circumstances. Thus, the question of what they should do in C_2 presupposes it is open what Benny does while the circumstances determine that it is not open.

In sum, the difficulty is that to show that a group can have an obligation when its members do not have obligations to do their parts, we need to relativize the obligations to the same circumstances, for here we are thinking about what they have most moral reason to do, and that depends on the circumstances. When we ask what an individual should do, we bracket what they intend to or will do on the matter. We do this because if we were to treat as part of the circumstance against which we evaluate what she should do *what she will in fact do*, then it is not open to her to do otherwise than she will *in those circumstances*, and the question is moot. The question of compliance to what duty requires in an agent's circumstances must focus only on the circumstances exclusive of the will of the agent and what she will do in response to the circumstances. We assume a compliant will and ask what the

agent would do to satisfy the requirements of duty. This same constraint applies when we ask what a group of individuals should do in some circumstances. Here we have in view the wills of each of the members. We then bracket, in asking what the group should do, what the individual members intend, and what they will or are likely to do, and ask essentially what they would do on the assumption that their wills are compliant to duty. But for each of them then the circumstances include the assumption of the wills of the others being compliant to their duties. When we focus not on the group, but on an individual member of the group, then the requirement that we bracket the wills of the others is lifted. We can then take into account what others intend, or will or are likely to do. This then can alter what it is morally best to do from the point of view of the individual, but this is also an evaluation relative to different circumstances. Thus, we do not get the verdict that relative to fixed circumstances the group obligation to do something and its members obligations to do their parts come apart.

Takeaways

First, we can see that one reason the basic intuition seems compelling is because a presupposition of asking what a group should do is that their wills are compliant to the demands of morality. So if we then ask what each should do, we think about this in the light of each of the others being compliant to doing what is best in those circumstances. Then the best that each can do is to make their contribution to the best outcome the group can achieve by all doing their parts in that. But this, I suggest, is a matter of the presupposition of the question, namely, that the parties' wills are not determined and are compliant to the demands of morality. When we focus on the question 'What should I do?' asked from the perspective of one member of the group, we presuppose only that the agent's will is compliant. We can then in principle get different answers to what *they* should do (given the presuppositions of the question) and what *individual agents* should do (given what additional facts can enter into the calculation given the different presuppositions). But this does not show that group obligations are not reducible to or derivable from individual obligations.

Second, when we ask what individual agents ought morally to do, we are restricting our attention to moral considerations. We are not asking all-in what the agent should do from the standpoint of individual rationality. If moral considerations are not overriding, then it can be rational for an agent to do something other than what is morally best or required. When we are evaluating outcomes with respect to moral considerations alone, however, differences in individual perspective on the value of outcomes disappears. The best moral outcome is by its nature best for everyone when we are restricting attention to moral considerations (see note 8 on agent-relative morality). A consequence of this is that what the

individual engaged in moral reasoning aims at in the context of group action is the best outcome for the group. There is no question about whether the best outcome from the group could come apart from the best outcome for the individual, as long as we are focusing on moral considerations alone. If participants can reasonably assume that it is public that all the potential participants are acting from moral considerations and informed about the outcomes, then except in cases in which there is a tie for the best moral outcome, there is a unique answer to the question 'What should I do?' which is determined by what the best outcome the group can achieve is, namely, my part (perhaps to be determined) in our achieving the morally best outcome.

This shows that moral reasoning by its nature requires something akin to group benefactor reasoning, in Bacharach's terms. In group benefactor reasoning, agents facing a decision about how to act in concert with others prioritize, in their individual preferences, group utility (however defined). If we think of maximizing group utility in terms of the best moral outcome, then reasoning from moral considerations alone is group benefactor reasoning.

Bacharach argued that team reasoning, which involves agency transformation, is not equivalent to group benefactor reasoning, which involves preference transformation (Bacharach 2006). The cash value of this distinction is that in team reasoning the individual starts with the question what is best for the team (if the team were an agent and I were the team, what would I do?), and then derives his part in that. If there is a strict ordering of the action ensembles in terms of group utility, team reasoning chooses the ensemble that is ranked highest. In group benefactor reasoning, the individual aims for the best team outcome, but still must think about it in terms of what the other players do. This allows that team reasoning, the argument goes, may resolve (by transforming) certain Prisoner's Dilemma games when group benefactor reasoning does not. For example, in a standard Prisoner's Dilemma such as that shown in Figure 6 (Petersson 2017),

Prisoner's Dilemma		Player 1	
		Cooperate	Defect
Player 1	Cooperate	4, 4	0, 5
	Defect	5, 0	3, 3

Figure 6: Prisoner's Dilemma

both team reasoning and group benefactor reasoning will look at a transformed matrix, as shown in Figure 7, which treats each square's value as the sum of the values of the players.[10]

[10] Donald Regan (1980, chapter 2, esp. p. 62) observed that Prisoner's Dilemmas are not possible for act utilitarians (AU) because there can be no agent-relative differences in the payoffs of combinations

Prisoner's Dilemma Transformed		Player 1	
		Cooperate	Defect
Player 1	Cooperate	8	5
	Defect	5	6

Figure 7: How the Prisoner's Dilemma in Figure 6 looks from the perspective of team reasoning and group benefactor reasoning.

Team reasoning recommends the combination of (Cooperate, Cooperate) because that is the highest payout for the group. For the group benefactor, that is clearly best, but the claim is that since each agent is engaged in individual strategic reasoning, what she should do depends on what the other does, for this is a Hi-Lo game.[11]

If we are focusing on moral outcomes, and the values in Figure 6 represent summable goods, then we will arrive at the payouts in figure 7. But I argued above that in this situation moral agents reasoning from the first-person perspective can rationally arrive at the decision to cooperate if it is common knowledge (or just reasonably assumed) that the players aim (or are likely to aim) for the best moral outcome or, equivalently, are paying attention only to moral considerations. In fact, in presenting the values in the decision matrix as moral values we presuppose each is reasoning from moral considerations, and if each knows the matrix that is relevant each knows that is true of the other. So team reasoning is not necessary to arrive at the correct decision in these cases.

Moreover, in many cases, even if you have very good reason to think the other will not pursue the morally best outcome, the highest expected value may favor

---

of actions. The result is the same for any theory (AU or not) on which the moral values realized by any combination of actions are the same for all those acting. As pointed out in note 8, this is compatible with agent-relative moral value based on special relations as long as it adds to the total for everyone. Wlodzimierz Rabinowicz (1989), in a very interesting paper, has argued that one can construct Prisoner's Dilemmas for AU. The arguments are too involved to go into here in detail. One depends on construing AU as requiring agents to maximize the utility of their own acts (which is a form of agent-relativity that introduces a bias toward bringing about a good oneself) and to ignore the past (even if relevant to the total value contributed to the universe by one's present actions) and then involves considering temporal sequences of acts by different agents and an organic value principle (the whole may be better than the sum of its parts). Another depends on adopting preference utilitarianism and future orientation and considering preference changes from the perspective of a single agent at different times. There are assumptions here which I would want to raise some questions about, but, in any case, the scenarios don't impinge on the discussion of the decision matrices considered in the text, which involve simultaneous choices by distinct agents.

[11] Paul Weirich (2018) responds to this argument by noting that what act one chooses can carry information about what act the other will choose when we assume "each player can predict his counterpart's response to his choices" (218). Player 1 choosing Hi "supplies a reason for its performance because it carries information that the other player, predicting his act, will choose High in response" (218). Choosing Hi is self-supporting. The same goes for choosing Lo, but Hi has greater self-conditional utility, so Hi is the rational choice.

cooperation. In Figure 7, given the values, the probability that the other will defect must be greater than .75 for it to be best to defect. So cooperating will be reasonable unless the prospects are quite dim.

Finally, say that in the game in Figure 7, player 1 knows player 2 will defect with probability .8. In these circumstances, player 1 should defect as well because that promotes the better outcome. As noted above, these are not the circumstances in which we would say that they ought to cooperate, so we do not get a mismatch between individual obligations and collective obligations. But what does team reasoning recommend? If player 1 engages in team reasoning, he will choose to cooperate, despite knowing this does not lead to the morally best outcome. Thus, team reasoning is inappropriate in the sense that it leads to what are known to be suboptimal outcomes. It would be morally wrong for player 1, given what he knows, to engage in team reasoning. It would be morally permissible to engage in team reasoning only if one independently had reason to think that would lead to the best outcome given how the other is likely to reason. This generalizes. When one ought not to act on the assumption that the other aims at the morally best outcome, one ought not to engage in team reasoning. It is permissible to engage in team reasoning only when one has independent reason to think the expected value of doing one's part in what would be the best outcome objectively is higher than other choices. It is therefore not appropriate to decide what one ought to do by first adopting the strategy of team reasoning and then deriving what one ought to do from that. The fundamental standpoint from which to derive one's obligations remains that of the individual agent thinking about what is best given the actual facts about what it is reasonable to think others will or are likely to do.[12]

References

Bacharach, Michael. 2006. *Beyond Individual Choice: Teams and Frames in Game Theory*. Princeton: Princeton University Press.

Blomberg, Olle, and Björn Petersson. 2023. "Team Reasoning and Collective Moral Obligation." *Social Theory and Practice*. doi: https://doi.org/10.5840/soctheorpract2023120177.

Brownlee, Kimberley. 2010. "Reasons and ideals." *Philosophical Studies* 151 (3):433-444.

Collins, Stephanie. 2019. *Group Duties: Their Existence and Their Implications for Individuals*. Oxford: Oxford University Press.

Colman, Andrew M., Briony D. Pulford, and Catherin L. Lawrence. 2014. "Explaining Strategic Coordination: Cognitive Hierarchy Theory, Strong Stakelberg Reasoning, and Team Reasoning." *Decision* 1 (1):35-58.

[12] My thanks for very helpful comments from Olle Blomberg and Mattias Gunnemyr.

Dietz, Alexander. 2016. "What We Together Ought to Do." *Ethics* 126 (4):57-69.

Harsanyi, John C. , and Reinhard Selton. 1988. *A General Theory of Equilibrium Selection in Games*. Cambridge, Mass.: MIT Press.

Heuer, Ulrike. 2010. "Reasons and impossibility." *Philosophical Studies* 147 (2):235 - 246.

Lord, Errol. 2015. "Acting for the Right Reasons, Abilities, and Obligation." *Oxford Studies in Metaethics* 10:26-52.

Petersson, Björn. 2017. "Team Reasoning and Collective Intentionality." *Review of Philosophy and Psychology* 8 (2):199-218.

Rabinowicz, Wlodzimierz. 1989. "Act-utilitarian prisoner's dilemmas." *Theoria* 55 (1):1-44.

Regan, Donald. 1980. *Utilitarianism and Co-operation*. Oxford: Oxford University Press.

Schwenkenbecher, Anne. 2019. "Collective Moral Obligations: 'We-Reasoning' and the Perspective of the Deliberating Agent." *The Monist* 102 (2):151-171.

Streumer, Bart. 2007. "Reasons and Impossibility." *Philosophical Studies* 136 (3):351-384.

Streumer, Bart. 2010. "Reasons, impossibility and efficient steps: reply to Heuer." *Philosophical Studies* 151 (1):79 - 86.

Sugden, Robert. 2000. "Team Preferences." *Economics and Philosophy* 16:175-204.

Weirich, Paul. 2018. "Rationality and Cooperation." In *The Handbook of Collective Intentionality*, edited by Marija Jankovic and Kirk Ludwig, 209-220. New York: Routledge.