

Group Agency and Overdetermination

David Killoren · Bekka Williams

Accepted: 10 January 2012 / Published online: 3 February 2012
© Springer Science+Business Media B.V. 2012

Abstract A morally objectionable outcome can be overdetermined by the actions of multiple individual agents. In such cases, the outcome is the same regardless of what any individual does or does not do. (For a clear example of such a case, imagine the execution of an innocent person by a firing squad.) We argue that, in some of these types of cases, (a) there exists a group agent, a moral agent constituted by individual agents; (b) the group agent is guilty of violating a moral obligation; however, (c) none of the individual agents violate any of their moral obligations. We explicate and defend this view, and consider its applications to problems generated by anthropogenic climate change and electoral politics.

Keywords Group agency · Overdetermination · Climate change · Electoral politics

1 Introduction

Consider the following overdetermination case. A torture machine is equipped with two switches. An innocent victim is attached to the machine. If no switches are flipped, then the victim feels no pain. But if either switch is flipped, then the victim feels tremendous pain. The victim feels the same tremendous pain whether just one switch is flipped, or just the other switch is flipped, or both switches are flipped. Torturer A controls one switch, and torturer B controls the other switch. If either torturer fails to throw his or her switch, then that torturer will be mildly punished, but will suffer no other bad consequences.

Suppose, for the sake of illustration, that if the victim feels tremendous pain, then this is equivalent to -1000 units of disutility, while the punishment of either torturer is

D. Killoren · B. Williams (✉)
University of Wisconsin, Madison, WI, USA
e-mail: rlwilliams1@wisc.edu

D. Killoren
e-mail: killoren@wisc.edu

equivalent to -10 units of disutility. Then we can represent the torturers' situation with this matrix:

	B throws his switch	B does not throw his switch
A throws her switch	The victim feels tremendous pain (-1000); neither A nor B is punished (0)	The victim feels tremendous pain (-1000); B is mildly punished (-10)
A does not throw her switch	The victim feels tremendous pain (-1000); A is mildly punished (-10)	The victim feels no pain (0); A and B are mildly punished (-20)

Now, consider this example from the perspective of standard maximizing act-utilitarianism. Suppose that A is justifiably certain that B will throw his switch. In that case, if A throws her switch, then -1000 units of disutility will result, whereas if A does *not* throw her switch, then -1010 units of disutility will result. So, given that B will throw his switch, A would maximize utility if she throws her switch. Thus A is *required* to throw her switch, according to act-utilitarianism. And precisely the same reasoning applies to B, if B is justifiably certain that A will throw her switch. Thus we get the counterintuitive result that A and B are each morally required to flip their switches just as long as each knows the other will flip his or her switch.

Of course, neither A nor B can be fully certain of what the other will do. Given this, perhaps each ought to play it safe and refrain from throwing his or her switch, in the hope that the other will do the same. This suggestion gains traction given a *subjective* version of act-utilitarianism, according to which A and B are each required to maximize *expected* utility. However, as long as the probability that B will throw his switch exceeds 99%, and the probability that A will throw her switch exceeds 99%, then the expected utility of throwing one's switch (whether one is A or B) exceeds the expected utility of not doing so.¹ Thus, even for the subjective act-utilitarian, as long as A and B are each sufficiently confident that the other will throw his or her switch, then each is morally required to do so.

Let us try to be precise about the nature of the difficulty here. How might A and B each be sure that the other will throw his or her switch? Perhaps A and B are simply very bad people who want to torture others, and each knows this feature of the other's character; so each knows that the other will commit torture when given the opportunity to do so. Then, from an act-utilitarian point of view, A and B are not *entirely* immune from moral criticism. The act-utilitarian can criticize A and B for being the sorts of people who desire to torture, since being that sort of person tends to cause vast avoidable harm.

So the problem raised by this case is *not* that the act-utilitarian is unable to criticize A and B. Instead, the problem is that once the act-utilitarian has criticized A's and B's *character*, then the act-utilitarian has no more criticism left to dish out. So when A and B simultaneously throw their switches, causing immense pain to their victim, their *actions* are not subject to moral criticism. Indeed, as we have seen, the act-utilitarian has to say that their actions are morally required.

This result seems paradoxical, to say the least. And the sense of paradox is heightened by the fact that this result seems *especially* unacceptable from an act-utilitarian perspective. After all, if A and B would both simply refrain from throwing their switches, then only -20 units of disutility would result, as compared to -1000 units if they both throw their switches.

¹ That is, if $EU(A \text{ throws A's switch}) = -1000 \cdot \text{pr}(B \text{ throws B's switch}) + -1010 \cdot \text{pr}(B \text{ does not throw B's switch})$, and $EU(A \text{ does not throw A's switch}) = -1010 \cdot \text{pr}(B \text{ throws B's switch}) + -20 \cdot \text{pr}(B \text{ does not throw B's switch})$, then $EU(A \text{ throws A's switch}) > EU(A \text{ does not throw A's switch})$ when $\text{pr}(B \text{ throws B's switch}) > 0.99$. And something similar is true for $EU(B \text{ throws B's switch})$.

The latter outcome is far worse *from an act-utilitarian point of view*. Thus the act-utilitarian's commitments seem to be at odds with act-utilitarianism itself.

The fact that overdetermination raises such problems for act-utilitarianism is not news.² But some philosophers continue to neglect these problems. For instance, in a recent paper, Shelly Kagan ignores the possibility of overdetermination; as a result of this oversight, he defends a false conclusion.³ Moreover, the *scope* of overdetermination problems seems to remain underappreciated. It is often assumed that overdetermination is mainly a problem for act-utilitarians (and other maximizing consequentialists), and that we are able to avoid such problems simply by shifting to a non-consequentialist view. That assumption is mistaken.

To see why, consider the following scenario. Moe and Larry have promised to carry a piano upstairs by noon. This is a two-person job (neither stooge can carry the piano alone), and will require a half hour of time. Suppose that fulfilling this promise would result in greater utility than violating it, so act-utilitarians will say that Moe and Larry have a clear moral reason to carry the piano upstairs. Typical non-consequentialists will say that Moe and Larry have at least one additional moral reason to carry the piano upstairs. That reason is provided by the fact that Moe and Larry have promised to do so. In light of these points, it seems that a wide range of theorists, including non-consequentialists as well as consequentialists, will be tempted to say that Moe and Larry are *obligated* to carry the piano upstairs.

However, it is now 11:29 am, and both stooges are simply relaxing on the couch, each too lazy to put forth any effort. They have decided to let noon pass without even touching the piano. Here is a matrix that represents the stooges' situation:

	Larry does nothing	Larry puts effort toward carrying the piano
Moe does nothing	The stooges' promise goes unsatisfied	The stooges' promise goes unsatisfied
Moe puts effort toward carrying the piano	The stooges' promise goes unsatisfied	The stooges fulfill their promise

As we have suggested, one need not be an act-utilitarian to be inclined to say that the stooges violate a moral obligation by failing to carry the piano upstairs. But notice that the breaking of the stooges' promise is ensured by the fact that Moe refuses to keep it, *and* is ensured by the fact that Larry refuses to keep it. Thus, neither stooge is individually able to bring it about that their promise is kept—just as neither torturer in our previous example is individually able to bring it about that the victim is not tortured. Of course, in the stooges' case, the breaking of their promise is brought about by their *failure* to act, rather than by their positive actions, as in the case of the torturers. Thus we might call this an "overdetermination by omission" case. But it is an overdetermination case all the same.

So Moe and Larry present a problem. Neither Moe nor Larry is *able* to keep their promise, given that both Moe and Larry are *unwilling* to do so. Thus, it is unclear whether

² See, for instance, the careful discussion found in Regan (1980).

³ Kagan (2011). In that paper, Kagan considers problems for consequentialism raised by what he calls (following Parfit) "imperceptible difference" cases. These are cases in which "when enough of us perform the act in question, the results will be bad, but for all that, my individual act will make no morally relevant difference at all" (p. 117). Kagan argues, at some length, that such cases are *logically impossible* given consequentialism (pp. 129–134). To support his view, he shows that one apparent example of such a case turns out not to be such a case after all; from this, he draws the general conclusion that *all* cases that seem to be imperceptible difference cases cannot be *genuine* imperceptible difference cases. But in this argument, Kagan mistakenly assumes that all imperceptible difference cases are of the same type as his example. Overdetermination cases (like the case of the two torturers above) meet his criteria for an imperceptible difference case, and are logically possible, so his claim that imperceptible difference cases are logically impossible is false.

we can justifiably say that either of them is (as an individual) morally *obligated* to do so. Nevertheless, many of us—consequentialists and non-consequentialists alike—still *want* to say that Moe and Larry are obligated to keep their promise.

There are only a few possible solutions to overdetermination problems like this one. We will review these solutions and will argue in favor of one of them: in overdetermination cases, there is a *group agent*—a moral agent composed of individual agents—that bears a moral obligation.⁴ According to this view, as it is applied to the case of the stooges, neither stooge bears an obligation to carry the piano upstairs, although the two stooges *together* comprise a group agent that *is* obligated to carry the piano upstairs. We will defend this view in sections 1-3. Then, in the Conclusion, we will briefly discuss the significance of this view for problems in practical ethics.

2 The Argument from Overdetermination for Group Agency

We want to argue that group agents exist, that group agents bear moral obligations, and that the obligations of group agents can be significantly different from the obligations of their members. To make that argument, we will focus on the case of the stooges. As we suggested above, the following claim seems attractive:

Deontic Judgment about the Stooges (DJS): Moe and Larry are morally obligated to get off the couch and carry the piano upstairs.

However, if one affirms DJS, it is still not entirely clear *what* one affirms. According to a straightforward individualistic reading, DJS ascribes two separate obligations to two separate individuals: Moe is morally obligated to get off the couch and carry the piano upstairs, and Larry is morally obligated to get off the couch and carry the piano upstairs. But consider the following principle:

Obligations Imply Can (OIC): If an agent is morally obligated to do X, then the agent can do X.

If Moe is morally obligated to carry the piano upstairs, then he is morally obligated to do something that he cannot do. (Moe cannot carry the piano alone, and Larry is so lazy that he will refuse to help.) Likewise, if Larry is obligated to carry the piano upstairs, then he, too, is obligated to do something he cannot do. (Moe is just as lazy as Larry.) Thus DJS on a straightforward individualistic reading conflicts with OIC. But OIC is plausible.⁵

So consider a *revisionist* individualistic reading of DJS, according to which Moe and Larry are each individually obligated to *put effort toward* carrying the piano upstairs. On this reading, Moe is (for instance) obligated to go to the piano as if he intends to carry it upstairs, to urge his fellow stooge to join him, and in general to *try* to bring it about that the piano is carried upstairs. And Larry is obligated to do the same. This reading of DJS is not literal (since DJS does not literally say anything about what Moe and Larry are obligated to *try* to do). But this reading still seems *close* (perhaps close enough) to literal.

The revisionist individualistic reading is consistent with OIC—Moe and Larry are each fully capable of *trying* to carry the piano upstairs—but it faces a different difficulty. According to DJS on this reading, each stooge has a moral obligation to do something that

⁴ A “group agent,” in our sense, is fundamentally a *moral agent*, i.e., a bearer of moral obligations. Therefore, our position is that multiple agents can, as a group, bear moral obligations. This position comes with various commitments, which we will discuss in Section 3. For now, the important point is that, fundamentally, we understand group agents as *moral* agents.

⁵ For a useful discussion of OIC, see Vranas (2007).

he arguably has no *reason* to do. Moe cannot succeed in carrying the piano upstairs alone, and Moe knows that Larry will not help him. So even if Moe *were* to make an effort to carry the piano upstairs, the piano would remain right where it is. Given this, it is unclear why we should think that Moe has any reason to try to carry the piano upstairs. And of course, the same seems true for Larry.

Here is a plausible claim about reasons and obligations:

Moral Rationalism (MR): If an agent is morally obligated to do X, then the agent has a reason to do X.

And here is a plausible principle regarding reasons for effort:

Reasons for Effort (RE): If an agent knows that she cannot do X, then the agent has no reason to try to do X.

Given MR and RE, neither Moe nor Larry is obligated to try to carry the piano upstairs. So if we accept DJS on the individualistic revisionist reading, then we need to either deny MR or deny RE. But both MR and RE are plausible.⁶ Thus, DJS on the revisionist individualistic reading is objectionable.⁷

So consider the group agency reading, according to which DJS posits the existence of a group agent—an agent composed of agents—and ascribes to that agent an obligation. Let us use brackets to denote group agents, like so: [Moe, Larry]. According to this reading, DJS is silent about what Moe is obligated to do, and is silent about what Larry is obligated to do; DJS is instead an assertion about what [Moe, Larry] is obligated to do. In particular, DJS asserts that [Moe, Larry] is obligated to get off the couch and carry the piano upstairs.

The group agency reading seems to avoid the costs we mentioned a moment ago. First, the group agency reading seems compatible with OIC. Although neither Moe nor Larry is able to carry the piano upstairs, the *pair* of them—i.e., [Moe, Larry]—*is* able to carry the piano upstairs.⁸ Second, the group agency reading seems compatible with MR and RE. It seems that [Moe, Larry] has a clear reason to try to carry the piano upstairs; that reason is provided by the fact that Moe and Larry *promised* to carry the piano upstairs.

⁶ Any defense of MR and RE will depend on what reasons turn out to be, and is thus beyond the scope of this paper. So here we must rely simply on the intuitive appeal of these two principles. For deeper discussion, see Shafer-Landau (2005), ch. 7.

⁷ Here it is useful to compare our argument to an argument previously developed by Frank Jackson (1987). Jackson's argument appeals to what he calls the "difference principle," according to which "the morality of an action depends on the difference it makes; it depends, that is, on the relationship between what would be the case were the act performed and what would be the case were the act not performed." He considers a case analogous to the case of the two torturers above, and appeals to the difference principle in order to argue that neither torturer's action is wrong (pp. 94–96). Similarly, in the case of the stooges, Jackson might argue that neither stooge is obligated to try to carry the piano, simply because neither stooge would make a difference to the outcome by so doing. But this line of argument implies that neither Moe nor Larry would be obligated to try to carry the piano *even if* each stooge were unaware that the other will refuse to help. Similarly, in the torturer example, Jackson is committed to say that A can permissibly throw her switch *even if* A does not know that B will throw hers. Thus Jackson's reasoning controversially presupposes that an agent's knowledge is irrelevant to the agent's obligations. By contrast, our argument relies on MR and RE, rather than Jackson's difference principle; so it is open to us—but not to Jackson—to allow that each stooge's obligations might depend on his knowledge of the other's intentions and behavior.

⁸ Or so it seems. That is, it seems that Moe can truthfully say "I am unable to carry the piano upstairs (given that Larry is too lazy to help)," and it seems that Larry can truthfully say "I am unable to carry the piano upstairs (given that Moe is too lazy to help)," but it does *not* seem that Moe and Larry can truthfully say in unison, "We are unable to carry the piano upstairs (given that each of us is too lazy to help the other)." The fact that Moe and Larry are both lazy seems to incapacitate them *as individuals* but not *as a group*. We will provide support for this view in our discussion of joint capacities below, in Section 3.

According to the group agency reading, the phrase “Moe and Larry” (in the sentence “Moe and Larry are morally obligated to get off the couch and carry the piano upstairs”) does not actually refer to Moe, and does not refer to Larry; it refers instead to [Moe, Larry]. Given this reading, the following view becomes attractive: (1) [Moe, Larry] is obligated to carry the piano upstairs, but (2) neither Moe nor Larry, as an individual, is obligated to carry the piano upstairs. Call this the *group agency solution* to the overdetermination problem raised by the case of Moe and Larry. (1) is the group agency solution’s positive component, and (2) is its negative component.

The group agency solution provides us with a way to consistently affirm OIC, MR, and RE while *also* affirming DJS. We believe that this is a major advantage. But the group agency solution is not without problems. We will first discuss problems for the negative component of the group agency solution; then we will discuss problems for its positive component.

3 Problems for the Negative Component of the Group Agency Solution

Suppose that Susan is the person to whom Moe and Larry have promised to carry the piano upstairs. When Susan discovers that Moe and Larry have failed to fulfill their promise, she is likely to blame them for this. And presumably, her blame will be directed at two *individuals*, Moe and Larry. It is not clear that Susan would (or ought to) be satisfied instead to blame a third entity, [Moe, Larry]. At any rate, it seems that Moe and Larry cannot legitimately shift the blame for their inaction away from themselves onto [Moe, Larry]. Yet if neither Moe nor Larry has violated an obligation, as the negative component of the group agency solution says, then it is unclear how either of them can *deserve* any blame. This looks like a problem for the negative component of the group agency solution.⁹

Similarly, after Moe and Larry have failed to fulfill their promise, it seems that they ought to feel guilty for not doing what they promised to do. But it would be strange to think that [Moe, Larry] ought to feel guilty. [Moe, Larry] is constituted by a pair of separate individuals who do not share a mind, and thus cannot share feelings. So, apparently, [Moe, Larry] cannot feel anything; only the *individuals*, Moe and Larry, can feel guilty. So if anyone ought to feel guilty, it would seem to be Moe and Larry as individuals. But again, it is unclear that such feelings would be appropriate if neither individual has violated an obligation.

Faced with this problem, the defender of the group agency solution has to bite one or the other of two bullets. First, she can say that neither Moe nor Larry as individuals deserve any blame, and neither ought to feel guilty. Call this the *blameless stooges view*. Second, she can say that Moe and Larry as individuals *do* deserve blame, and ought to feel guilty, even though neither one violates an obligation.¹⁰ Call this the *blameworthy stooges view*. We will consider each of these options in turn.

We admit that the blameless stooges view is counterintuitive. Intuitively, Moe and Larry *are* deserving of blame, and ought to feel guilty, about their failure to carry the piano. But this intuition might be due simply to a failure to fully appreciate the situation faced by Moe and Larry. As we have emphasized, in the case as described, Moe and Larry are each *unable* to bring it about that the piano is carried upstairs. Ordinarily, it is unreasonable to blame

⁹ Relatedly, Michael Zimmerman (1996) acknowledges the initial attractiveness of a group agency solution to difficulties posed by the sort of case exemplified by the stooges, but he finds the solution unacceptable. He writes: “the group wrongdoing is not attributable to any of the members of the group [...] But this is too much to swallow; the scenario still smacks of two (individual) wrongs making a right” (pp. 262–63).

¹⁰ This possibility is not considered by Zimmerman.

someone for failing to do what he is *unable* to do. Moreover, given that Moe and Larry are unable to carry the piano upstairs, each has no reason to *try* to do so (given MR and RE). Ordinarily, it is unreasonable to blame someone for failing to do what he has no *reason* to do. Given these points, if Moe and Larry (as individuals) are to blame, it is not at all obvious what they might be to blame *for*.¹¹ So, it is at least arguable that the blameless stooges view is correct, even though it is counterintuitive.

Furthermore, proponents of the group agency solution who want to reject the blameless stooges view might be able to accept the blameworthy stooges view—the view that Moe and Larry (as individuals) deserve blame, and ought to feel guilty, even though neither one violates an obligation. It is true that blame and guilt are typically appropriate when an agent violates an obligation. But these might *also* be appropriate when an individual agent is part of a group agent that has violated an obligation, even if the individual agent (*qua* individual) has not violated any obligation.¹² If this were so, then it would support the blameworthy stooges view.

However, a problem for the blameworthy stooges view is raised by the fact that Moe and Larry are each *unable to avoid* being part of a group agent that violates an obligation. Suppose that Moe tries his best to bring it about that the piano is carried upstairs. He gets off the couch and pushes on the piano, but to no effect; it is just too heavy for one person. He begs Larry to help, but Larry refuses. He even offers to pay Larry in exchange for help, but Larry *still* refuses. Noon passes, and the piano remains where it is. In this case, it may seem that only Larry ought to feel guilty. After all, Moe did all he could. But Moe is still a part of a group agent that violates an obligation, so—according to the blameworthy stooges view—it would still be appropriate for him to feel guilt, just as if he had remained on the couch and done nothing at all.

But note two things. First, when Moe gets off the couch and shows his willingness to carry the piano upstairs, *Larry's* moral situation changes. It is no longer the case that the piano would remain unmoved no matter what Larry does. So now we can reasonably say (consistent with OIC, MR, and RE) that Larry violates an obligation when he refuses to carry the piano upstairs. This means that, unlike in the original case, Larry now has *two* grounds for guilt: that he is part of an obligation-violating group agent; and that he violates an obligation as an *individual* agent.

Second, when Moe gets off the couch and shows his willingness to carry the piano upstairs, he reveals that he is not *the type of person* who would refuse to do what he has promised to do. In other words, Moe's demonstration of willingness to carry the piano speaks well of his character. By contrast, if he were unwilling to carry the piano upstairs (as in the original case), then Moe would be, ipso facto, the sort of person who refuses to do what he has promised to do. Arguably, one ought to feel a certain kind of guilt just for being the sort of person who refuses to do what he has promised to do.

So it seems that when Moe gets off the couch and exerts (futile) effort to carry the piano upstairs, he *eliminates* one reason for him to feel guilty, and *adds* one reason for Larry to feel guilty. The upshot is that Larry ought to feel *more* guilty than Moe—whereas, in the original case (in which both Moe and Larry are unwilling to carry the piano), Moe and Larry ought to feel guilty to the same degree.

¹¹ As noted above, Zimmerman (1996) would object that if the stooges' group action was wrong, then surely at least one stooge's individual act must have been wrong as well (p. 255). But, contrary to Zimmerman, this entailment appears not to hold. Consider Jackson's "Morning Traffic Example": To ensure optimal safety, the group of morning commuters ought to adhere to the 60-kilometers-per-hour speed limit, and the group acts wrongly by driving at 80. But it is nonetheless true of each individual commuter that, *given that everyone else is driving at 80*, he ought to drive at 80 as well. So, arguably, while the group's speeding is wrong, no individual commuter's speeding is wrong (Jackson 1987, p. 102).

¹² See Margaret Gilbert's discussion of "membership guilt" (2002, pp. 134–35).

But it remains the case that, according to the blameworthy stooges view, Moe ought to feel some degree of guilt, whatever he does—since, whatever he does, he is a constituent part of a group agent that violates an obligation. Perhaps there will be intuitive resistance to this idea.¹³ If so, then the blameworthy stooges view is not without some intuitive cost.

Even so, the blameworthy stooges view succeeds in significantly mitigating the intuitive problem raised by the negative component of the group agency solution. When Moe and Larry both refuse to get off the couch in order to fulfill their promise, this position allows us to say that both Moe and Larry ought to feel guilty, and deserve some amount of blame, in virtue of their membership in a group agent that violates an obligation. And we are able to say that the *extent to which* Moe and Larry ought to feel guilty, and deserve blame, is to *some* degree under their individual control.

So, defenders of the negative component of the group agency solution have two options. They can endorse the blameless stooges view (according to which neither Moe nor Larry as an individual deserves any blame for his failure to carry the piano upstairs), or they can endorse the blameworthy stooges view (according to which Moe and Larry as individuals *do* deserve blame for their failure to carry the piano upstairs, *even though* neither one violates an obligation). Each option carries certain difficulties; nevertheless, we have argued, these difficulties are not insurmountable. We conclude that the negative component of the group agency solution can achieve a degree of compatibility with our intuitions about individual guilt and blame.

4 Problems for the Positive Component of the Group Agency Solution

According to the positive component of the group agency solution, if Moe and Larry are the only individuals in the room, then there are precisely two *individuals* in the room, but there are (at least) three *agents* in the room.¹⁴ That seems odd. However, the *possibility* of group agency is shown by a famous thought experiment due to Ned Block (1980). Imagine that everyone in China is given a set of instructions on how to behave like a neuron, and they are all somehow linked together in a network that is functionally equivalent to the neural network of a human brain. It seems conceptually possible that the resulting “nation-brain” would be able to make decisions and bear obligations. For instance, if the nation-brain were presented with the choice whether to launch an unprovoked nuclear attack on South Korea, it would seem obligated to refrain. Thus, the nation-brain seems to constitute a genuine agent; the number of agents in China seems to exceed the number of individuals in China.

¹³ However, if there is such resistance, it probably should not be founded in the notion that an individual cannot find himself in a situation in which he cannot avoid having some sort of a reason to feel guilty for what he does, no matter what he does. That notion is deservedly controversial. Suppose, for instance, that Moe promised to meet Lisa for lunch at noon uptown, but also promised to meet Rita for lunch at noon downtown. He cannot avoid breaking a promise in this case. So it is not implausible to think that he will have a reason to feel guilty for what he does, whatever he does. Thus it looks as if there might well be *some* cases in which an agent cannot avoid having reasons to feel guilty.

¹⁴ Of course, there might be more than three agents in the room. Imagine, for instance, a view according to which [Moe, Larry] combines with Moe to form a fourth agent, [Moe, [Moe, Larry]]. But our position does not require this. We do not hold that *just any* collection of agents constitutes a group agent. In that respect, our view differs from that endorsed by Frank Jackson, who counts as group actions all sums of individual actions (and group actions). According to Jackson’s view, “[m]y last eye-blink together with Nero’s burning of Rome is a group action, a highly heterogeneous one of no particular interest to anyone, but a group action nevertheless” (1987, p. 93).

The logical possibility of the nation-brain scenario suggests that groups of agents can conceivably form group agents. But it is still an open question whether it is ever *actually* the case that groups of individuals form group agents. And crucially, it is an open question whether Moe and Larry, in particular, form a group agent.

The group agency solution involves commitments regarding *moral* agency, i.e., the type of agency that is required in order to bear an obligation. As advocates of the group agency solution, we are committed to the view that Moe and Larry constitute a *moral* agent. This is not the same as the view that Moe and Larry constitute a moral *person*. We take it that moral persons are typically moral agents, but are also (perhaps more fundamentally) *patients*. That is, persons are worthy of moral concern in their own right. We do not need to say that [Moe, Larry] is worthy of moral concern in its own right, apart from moral concern due to Moe and Larry as individuals. We need to say only that [Moe, Larry] can bear obligations, apart from the obligations of Moe and Larry as individuals.

In this connection, we are not only, or even primarily, worried about Moe and Larry. We want the argumentative strategy that we pursued in Section 1 to be generalizable; we want it to have implications for real-life overdetermination problems. There is a range of cases that have the following features: (I) A number of different individual agents perform actions (or fail to perform actions) of a given type, and thereby produce a morally objectionable outcome; (II) in light of the morally objectionable outcome of the individuals' choices, we want to say that they violate a moral obligation; yet (III) given OIC, MR, and RE, the individual agents, *as individuals*, do *not* violate a moral obligation. We hope to be able to say that in *many* or *all* cases that have these three features, the individual agents whose actions produce the morally objectionable outcome constitute an obligation-violating group agent. Thus we want to endorse a *general* version of the group agency solution.

This ends up committing us to a fairly permissive conception of moral agency. To see this, imagine that Moe and Larry employ the following decision procedure. Whenever they have to decide what (as a pair) to do, they make a list of their options, and then use a random draw to pick an option from the list. Suppose that each is inflexibly committed to the use of this procedure; so, once they have used the procedure to arrive at a verdict about what to do, each stands by that verdict, no matter what it is.

Now suppose that they use this procedure in order to decide whether or not to keep their promise to carry the piano upstairs. The result is not to keep their promise; so that is what each resolves to do. Apparently, this case satisfies conditions (I), (II), and (III). So we want to say that Moe and Larry constitute a group agent, [Moe, Larry], and [Moe, Larry] is obligated to carry the piano upstairs.

However, since [Moe, Larry] makes decisions by random draw, [Moe, Larry]'s behavior across time would be rampantly irrational. [Moe, Larry] willingly trades \$50 in exchange for \$25 about half the time. [Moe, Larry] chooses opera rather than rock about half the time, and chooses rock rather than opera about half the time. If [Moe, Larry] prefers bacon over ham, and prefers ham over sausage, [Moe, Larry] might nevertheless prefer sausage over bacon. And so on.

It seems that there could be two main classes of conditions for agency: psychological conditions and behavioral (or "functionalist") conditions. On the one hand, it might be thought that any genuine agent must have psychological states such as beliefs and desires. But we are willing to grant (at least for the sake of argument) that [Moe, Larry] lacks beliefs and desires; we accept that all psychological states belong exclusively to individuals. On the other hand, perhaps a genuine agent must exhibit rationally explicable *behavior*—that is,

behavior explicable in terms of reasons for action. But we are willing to grant that [Moe, Larry] fails to exhibit such behavior, given its rampant irrationality.¹⁵

Despite all this, we want to call [Moe, Larry] a moral agent that has various obligations. So we need a conception of moral agency that is consistent with the view that a genuine moral agent can lack both inner psychology *and* rationally explicable behavior. This might be a tall order.

But consider the following pair of cases. In the first case, Moe and Larry are alongside a pond, and a child is drowning in the pond. The child's rescue requires the efforts of two people—neither stooge can do it alone, but they can easily do it together. To decide what to do, the stooges use their usual random-draw procedure, and thus arrive at the decision not to save the child.

The second case is much like the first: Moe and Larry are alongside a pond, a child is drowning, and the child's rescue requires the efforts of two people. But in this second case, Moe and Larry are in two different locations alongside the pond, and neither is aware (or able to become aware) of the other's presence, so they cannot work together to save the child.

In the second case, the stooges seem to be *unable* to save the child. And, not coincidentally, the stooges seem not to be morally culpable for the boy's death, either as individuals or as a group. By contrast, in the first case, it does seem that the stooges violate a moral obligation when they fail to save the child. So, although the first case satisfies condition (II) (and conditions (I) and (III), too), the second case does not satisfy condition (II).

In the first case, the stooges seem to have what we will call a *joint capacity* to save the child, but they lack such joint capacity in the second case. This suggests a view according to which group agency (which, recall, is for us a form of *moral* agency) occurs in the presence of multiple individual agents who are jointly capable of acting as a group in response to reasons.

For the present purpose, we do not need an analysis of the notion of joint capacity. That notion is, we think, as intuitive as the notion of *individual* capacity. Perhaps a given individual can (or cannot) defeat Garry Kasparov in chess. Likewise, perhaps a given group of individuals, working together as a team, can defeat Garry Kasparov, even if none of the individuals can do so. Similarly, in many cases, it seems that a group of individuals can act *in response to reasons*, even if the group rarely (or never) does so in fact. The stooges in fact use a random-draw group decision-making procedure, but they are *able* to change to a more reasons-responsive group decision-making procedure. Thus, for instance, in the first drowning case, [Moe, Larry] seems able to appropriately respond to the child's plight. Given that [Moe, Larry] *can* do this, it *ought* to do so, even if there is no reason to expect that it will in fact do so.

¹⁵ Thus, our position here is a bit more extreme than the position ingeniously developed at length by Philip Pettit and Christian List (2011). Pettit and List begin with a metaphysical account of agency, according to which agents have representational and motivational states, together with the ability to act on such states (p. 20). They then proceed to argue that, given this account of agency, certain types of groups are agents, and can thus bear moral responsibilities. By contrast, our argument begins with intuitively plausible moral judgments, such as DJS; we argue that these judgments provide reasons to endorse the metaphysical claim that certain obligation-bearing group agents, such as [Moe, Larry], exist. So Pettit and List move from metaphysics to morality, whereas we move from morality to metaphysics. This move, we believe, commits us to a significantly more permissive conception of agency than the conception endorsed by Pettit and List. For instance, we have to say that a group that exhibits thoroughgoing irrationality can be an agent, whereas Pettit and List can deny this.

Our suggestion, then, is that a group agent capable of bearing obligations exists when a group of individuals are jointly capable of acting in response to reasons.¹⁶ In such cases, according to the present view, it can make sense to ascribe an obligation to the group as a whole. This view implies that a group can be a moral agent even when it fails to exhibit rationally explicable behavior—so long as it *can* do so.

This position has a certain degree of intuitive plausibility. After all, much the same seems true of *individual* agents. A person who makes all of her decisions by coin-toss does not exhibit rationally explicable behavior; but if she remains *able* to exhibit such behavior, then she would seem to count as a genuine moral agent. This suggests that a conception of moral agency that requires the exhibition of rationally explicable behavior would be far too demanding. The *capacity* for rationally explicable behavior seems ordinarily to be sufficient for moral agency.¹⁷

Let us sum up. In the present section and the previous one, we have discussed a few main costs of the group agency solution. In the previous section, we saw that a satisfactory version of the group agency solution seems to require that *either* membership in an obligation-violating group agent is a basis for some degree of individual blameworthiness and guilt, even in cases where the individual members of the group agent do not violate any of *their own* obligations (the *blameworthy stooges* view)—or, contrary to intuition, joint laziness can absolve individuals of blame and guilt (the *blameless stooges* view). In the present section, we saw that the group agency solution seems to require a very permissive conception of moral agency, according to which the joint capacity of individuals to act in response to reasons establishes that those individuals constitute a group agent. On this view, a group can sometimes count as a moral agent even when it lacks inner psychology and fails to exhibit rationally explicable behavior.

In our view, these costs are acceptable. But we must acknowledge that they *are* costs. The question is whether these costs outweigh the benefits associated with them. The primary benefit of the group agency solution is that it allows us to affirm OIC, MR, and RE, while also justifying us in making assertions like “Moe and Larry are morally obligated to get off the couch and carry the piano upstairs.” We believe that this benefit is large.

5 Conclusion: Practical Applications

We have argued that hypothetical overdetermination cases, like the case of Moe and Larry, provide some substantial motivation for the view that group agents exist and that their obligations can differ from the obligations of their members. Suppose that view is correct. What then?

¹⁶ Note that this position does not commit us to the view that *every* group of agents constitutes a group agent. For instance, the current inhabitants of Earth (taken as a whole) probably do not count as a group agent on our account, because the current inhabitants of Earth are probably not *jointly capable* of acting in response to reasons.

¹⁷ Here it ought to be noted that many of the groups that qualify as group agents on our account are *extremely unlikely* to act in accordance with their obligations. For instance, in the piano case, it is virtually certain that [Moe, Larry] will fail to carry the piano upstairs; given this, there seems to be no *injunctive point* to saying that [Moe, Larry] bears an obligation to carry the piano upstairs. This looks like a problem for our view. However, there are in fact many cases in which a moral judgment is correct and appropriate despite having no injunctive point. For instance, one can correctly and appropriately judge that Jeffrey Dahmer’s killings were morally wrong, even though this judgment will be neither heard nor heeded—since the killings have already been done, and Dahmer is already dead. Similarly, if someone is now planning to commit a murder, it ought to be said that he is morally obligated not to do so, *even if* it is virtually certain that he *will* proceed with his plan. In general, moral judgments can be worth making even when there is no injunctive point in making them.

Consider the example of anthropogenic climate change. According to Walter Sinnott-Armstrong (2005), it is morally permissible to drive a gas-guzzling car on a Sunday afternoon just for fun—even if the widespread use of automobiles is a contributor to global warming, and even if global warming is harmful. Let us focus on one aspect of his discussion, which concerns the issue of whether going on a Sunday drive is harmful.

According to Sinnott-Armstrong, “global warming will still occur even if I do not drive just for fun,” and “global warming will not occur unless lots of other people also expel greenhouse gases.” So driving one’s car just for fun is “neither necessary nor sufficient for global warming” (p. 297). Sinnott-Armstrong is not a climate skeptic; he accepts that global warming is occurring as a result of human activity, and will cause harm. But that harm, he thinks, does not *counterfactually depend* on whether one person decides to go on one drive on a given Sunday afternoon.

So, consider the following argument. Going on a Sunday drive is, in most respects, an obviously harmless activity (assuming the driver is sober, has good vision, is careful, etc.). The only way in which driving on a Sunday *might* be harmful is in virtue of its contribution to global warming. But global warming will occur whether or not a particular individual goes on a Sunday drive. So, a Sunday drive is *not* harmful in virtue of any contribution that it might make to global warming. Thus, going on a Sunday drive is harmless. And if going on a Sunday drive is harmless, then it is morally permissible.¹⁸

Perhaps this argument succeeds; perhaps it does not. There are a number of different points at which one might object to it.¹⁹ Nevertheless, many people will be convinced by some version of this argument. We want to offer some advice to these people.

These people face a problem. It is similar to the problem generated by the case of the two torturers. When we consider that case, we want to say: “The two torturers are *morally obligated* not to torture their victim.” After all, torturing is much more harmful than not torturing. Likewise, many of us will want to say, “*We*—i.e., greenhouse gas-emitting members of industrialized societies—are morally obligated to decrease our collective emissions of greenhouse gases to the extent required in order to prevent global warming.” Indeed, this latter assertion is particularly attractive, in light of the vast scale of harm that global warming is expected to cause.²⁰

This intuition about what *we* are morally obligated to do seems, at first, to be inconsistent with the view that *each one of us*, taken as an individual, can permissibly engage in activities that emit greenhouse gases (GHGs)—e.g., going on a Sunday drive just for fun. But this inconsistency can be resolved by the notion of group agency. It is open to us to say that (a) we constitute a group agent, and (b) this group agent is obligated to take (collective) action to reduce its GHG emissions, *even though* (c) any given individual can permissibly engage in GHG-emitting activities. By asserting (b), we preserve the intuition that there is something deeply morally wrong with the status quo; by asserting (c), we preserve the view that GHG-emitting activities, such as going on a Sunday drive, are *individually* harmless, and therefore permissible.

Suppose that those who believe that it is permissible *for an individual* to go on a Sunday drive just for fun take our advice, and endorse (a)–(c). So, they agree that members of industrial society constitute a group agent, and this group agent is morally obligated to

¹⁸ We do not claim that Sinnott-Armstrong endorses this argument, although he might. In any case, if he did endorse this argument, he would have much more to say than we can say here about why the conditional “If going on a Sunday drive is harmless, then it is morally permissible” is true.

¹⁹ For some interesting, although questionable, objections to Sinnott-Armstrong’s view, see Hiller (2011).

²⁰ It has been estimated that global warming will lead to the suffering and death of four billion people. See Nolt (2011).

reduce its GHG emissions. How would that group agent go about *fulfilling* its obligation? It might use governmental means. For instance, we might create laws that restrict the sale or purchase of gas-guzzling cars; or we might hike gasoline taxes; or we might fund research into alternative fuel sources—etc. These types of actions, when taken by democratic governments, might reasonably be regarded as actions performed by a group agent constituted by democratic citizens. If this is right, then the view we are recommending here is not far from the position that Sinnott-Armstrong ultimately defends. Sinnott-Armstrong maintains that, although each *individual* can permissibly drive a gas-guzzling car just for fun, it may be that the *government* ought to restrict individuals' ability to do so.

Thus the position that groups of agents can constitute moral agents, whose obligations differ from the obligations of their members, is not just *theoretically* interesting; it is potentially applicable to at least one major ethical problem—the problem of anthropogenic climate change. There may also be applications to the ethics of voting. In a large election, an individual voter can be sure that her vote will not influence the outcome, and this may lead her to conclude that she is not obligated to vote. Such a view is not unreasonable. Nevertheless, electoral outcomes are often morally consequential. If Gandhi were running against Hitler in the next presidential election, it may seem quite obvious that we would be morally obligated to elect Gandhi rather than Hitler—even *if* one additional vote for Gandhi could not measurably improve the odds that Gandhi will be elected. To resolve this problem, we might say that (a) citizens constitute a group agent, and (b) this group agent is morally obligated to elect Gandhi rather than Hitler, even though (c) no *individual member* of this group agent is obligated to vote for Gandhi, or indeed to vote at all.

These points suggest that there are ways in which the views regarding group agency that we have defended can have practical significance. But there are further issues to be addressed before the notion of group agency can be applied to practical issues. We have only scratched the surface here.

Acknowledgments We are immensely grateful to Jeff Behrends, Harry Brighouse, Justin Horn, Matt Kopec, Jonathan Lang, Pete Nichols, and Russ Shafer-Landau for their feedback, advice, and conversation regarding previous drafts of this paper.

References

- Block N (1980) Troubles with functionalism? In: Block N (ed) Readings in philosophy of psychology vol. 1. Methuen, London, pp 268–305
- Gilbert M (2002) Collective guilt and collective guilt feelings. *J Ethics* 6:115–143
- Hiller A (2011) Morally significant effects of ordinary individual actions. *Ethics Policy Environ* 14(1):19–21
- Jackson F (1987) Group morality. In: Pettit P, Sylvan R, Norman J (eds) *Metaphysics and morality: essays in honour of J. J. C. Smart*. Basil Blackwell, Oxford, pp 91–110
- Kagan S (2011) Do I make a difference? *Philos Public Aff* 39(2):105–141
- Nolt J (2011) How harmful are the average American's greenhouse gas emissions? *Ethics Policy Environ* 14(1):3–10
- Pettit P, List C (2011) *Group agency: the possibility, design, and status of corporate agents*. Oxford, New York
- Regan D (1980) *Utilitarianism and cooperation*. Oxford, New York
- Shafer-Landau R (2005) *Moral realism: a defence*. Oxford, New York
- Sinnott-Armstrong W (2005) It's not *my* fault: global warming and individual moral obligations. In: Sinnott-Armstrong W, Howarth R (eds) *Advances in the economics of environmental resources vol. 5: perspectives on climate change*. Elsevier, Oxford, pp 285–307
- Vranas P (2007) I ought, therefore I can. *Philos Stud* 136(2):167–216
- Zimmerman M (1996) *The concept of moral obligation*. Cambridge University Press, Cambridge