Chapter 19
The SANE Approach to Real Collective Responsibility

Sara Rachel Chant

Abstract In this paper, I offer an argument for the existence of 'real collective responsibility' and the beginnings of an analysis of it. 'Real collective responsibility' refers to the responsibility that is borne by a group of individuals, but which is not reducible to the responsibility of each individual in the group. The approach I take is to draw an analogy between the uncontroversial way in which an individual's moral responsibility may be mitigated when her behavior is coerced, and the way in which group dynamics may exert pressure constraining the behavior of each member of a group. This sort of consideration suggests that real collective responsibility may occur when a group finds itself in a highly stable, accessible Nash equilibrium, which I refer to as the SANE condition for real collective responsibility.

1 Introduction

The claim that a group bears collective moral responsibility for its action has both a trivial and a non-trivial reading. On the trivial reading, the claim that a group is morally responsible for an action is merely shorthand for the claim that every individual in the group is morally responsible for her contribution to the action. For instance, if a set of bank robbers is morally responsible for a bank heist, we may mean only that each individual robber is morally responsible for her contribution to the robbery. This claim is widely assumed to be unproblematic.

The non-trivial reading is that the group may bear moral responsibility above and beyond the responsibility borne by the individual members. On this reading, the group of bank robbers is morally responsible in some sense that does not reduce to the responsibility of each robber. In contrast to the first sense of 'collective moral

S.R. Chant (✉)
Department of Philosophy, University of Missouri-Columbia, Columbia, MO, USA
e-mail: chants@missouri.edu

A. Konzelmann Ziv and H.B. Schmid (eds.), *Institutions, Emotions, and Group Agents*, Studies in the Philosophy of Sociality 2, DOI 10.1007/978-94-007-6934-2_19,

responsibility', the latter reading is extremely contentious in at least two ways. First, there is no consensus as to whether there are such cases at all; second, even those authors who have argued for it have not agreed on how this sense of 'collective moral responsibility' is to be analyzed.[1]

I shall refer to this former sense of 'collective moral responsibility' as 'distributive' because the responsibility of the group is distributed to the members, with nothing 'left over', so to speak. I shall refer to the latter sense as 'real collective responsibility'. In this paper, I argue for the existence of real collective responsibility, and I offer the beginnings of an analysis of it. The strategy I shall take is to draw an analogy between cases of real collective responsibility and ordinary cases in which an individual's moral responsibility is mitigated or excused by the fact that she was coerced into performing the action. Although coercion is not typically brought into discussions of real collective responsibility, I shall argue that coercion is relevant. This is because cases of real collective responsibility are ones in which (at least part of) each individual's moral responsibility has been mitigated by some situational feature, despite the fact that this mitigating feature fails to mitigate the responsibility of the group as a whole. This mitigating feature may be understood in analogy to coercion. According to the argument I shall offer below, real collective responsibility is to be understood entirely in terms of such mitigating features.

I shall begin by rehearsing an uncontroversial set of cases in which individuals are excused of (at least part of) their moral responsibility for their actions. These are cases in which the person is coerced into action on the force of a credible threat. Thus, I shall begin in the first section with Harry Frankfurt's famous arguments concerning (what he calls) the 'doctrine that coercion excludes responsibility' in his paper, 'Alternate Possibilities and Moral Responsibility' (Frankfurt 1969). Although there is a lively debate about the status of the broader so-called 'principle of alternate possibilities' which is the target of Frankfurt's analysis, it will not be necessary for me to become involved in those larger issues. Instead, I shall focus entirely on the uncontroversial cases which Frankfurt uses to motivate his main argument. In the second section, I examine an argument due to Joel Feinberg (Feinberg 1991), in which he argues that there is such a thing as (what I call) real collective responsibility, as well as some theoretical results in judgment aggregation due to Christian List and Philip Pettit (List and Pettit 2004), which have been interpreted to show that at least some facts about collectives are not reducible to facts about its members (as in Pettit 2004; Copp 2006). Although I shall conclude that the arguments from Feinberg and the impossibility results of List and Pettit cannot establish the existence of real collective responsibility, I will argue in the third section that they do point the way to a general account. That section contains an equilibrium account of real collective responsibility, which I argue provides the basis of a satisfactory account. I expand upon this account in the fourth section, where I offer the beginning of an analysis of the concept of real collective responsibility.

[1]There is a wide literature on the subject of collective responsibility, including Copp (2006), Feinberg (1991), Lewis (1991), May (1990), May and Hoffman (1991), Mellema (1988), Miller (2001), Narveson (2002), Sverdlik (1987), Pettit (2007).

2 Coercion Excludes Responsibility

My main argument in favor of the existence of real collective responsibility is that the dynamics that give rise to collective action are sometimes relevantly similar to cases in which individuals are coerced into performing actions that they would normally be unwilling to perform. Because cases of individual coercion are ones in which the individual's moral responsibility is mitigated or eliminated entirely, it is possible for group dynamics to have the same mitigating effect on the responsibility of the members of the group. In this way, any moral responsibility for the group's collective action must attach to the group *qua* group, without being distributed to the group's members.

Cases of individual coercion are often not discussed in detail in the literature on moral responsibility or collective action. Instead, the focus is typically on more difficult questions concerning moral responsibility and related concepts. However, individual coercion was one focus of 'Alternate Possibilities and Moral Responsibility' (Frankfurt 1969). In it, Frankfurt attacks the so-called 'principle of alternate possibilities' (hereafter, PAP), according to which an individual does not bear moral responsibility for her action if she couldn't have done otherwise. Frankfurt's argument against PAP takes the form of a single counterexample, the notorious Jones-4 case. In it, Jones-4 performs an immoral action of his own volition, but could not have done otherwise because another agent, Black, would have forced him to decide to perform the action if Jones-4 had wavered in his decision. Thus, Jones-4 bears responsibility for his action (presumably because the action was a result of his own motivations and was brought about 'in the right way') despite the fact that he couldn't have done otherwise (because Black would have ensured that he decide to perform the action in any event).

Given the tremendous intuitive plausibility of PAP, Frankfurt bolsters the Jones-4 counterexample by arguing that PAP gains its plausibility from a previously unquestioned assumption that it is relevantly similar to another principle, which Frankfurt takes to be unproblematic and uncontroversial. This is the principle that a person is not morally responsible for her action if she has been coerced into performing that action. According to Frankfurt, the so-called 'principle that coercion excludes moral responsibility' is sound, but PAP illicitly gains its plausibility from the assumption that PAP underwrites it. However, on closer examination, Frankfurt concludes that the relationship between coercion and moral responsibility does not depend on PAP after all.

For our purposes here, it is useful to rehearse Frankfurt's discussion of the relationship between coercion and moral responsibility. This discussion highlights a few uncontroversial features of coercion, which I will rely upon to motivate the SANE approach to real collective responsibility.

In his discussion, Frankfurt leads the reader down a garden path of cases, in which various agents are subjected to forces that are intended to constrain their choice of action. These cases lead Frankfurt to note that, although people may be under the same threat, some may be genuine cases of coercion, while others are

not. For example, Jones-1 has decided to perform some action, and is subsequently issued a threat from someone intending to coerce him into performing that very same action. But Jones-1, due to his peculiar psychology, is completely unmoved by the threat—he would have performed the action in any event, and the threat is completely irrelevant to him. So when he performs the action, he is fully responsible for having performed it, and so the threat does not count as constituting coercion at all.

On the other end of the spectrum, Jones-2 is 'stampeded' by the threat. Once the threat has been issued, no previous decision or intention of his is relevant at all. Thus, when Jones-2 carries out the action, it is entirely because of the threat. Thus, in contrast to Jones-1, we would say that Jones-2 is *not* responsible for the action. His responsibility is mitigated (or even excused entirely) because he is coerced.

Of course, when we say that coercion excludes moral responsibility, we are not typically thinking of cases that are as extreme as those of Jones-1 and Jones-2. Because Frankfurt is most interested in understanding why we treat coercion this way in ordinary cases, he goes on to consider one last case of coercion before examining PAP. In this penultimate case, Jones-3 has similarly decided before the threat is issued that he will perform a particular action. Subsequent to this decision, he is given the same threat that Jones-1 and Jones-2 were given. However, Jones-3 is a reasonable person who recognizes that the threat is credible, and so the consequences of failing to carry out the action are figured into his decision. That is, he is not (like Jones-1) impervious to threats, nor is he stampeded by them (like Jones-2). Rather, he rationally figures the consequences of inaction into his deliberations. Recognizing that the threat is credible, and that the consequences are severe, Jones-3 performs the action.

At this point, Frankfurt admits that the case might not provide a clear counterexample to the principle that coercion excludes moral responsibility. This is because Jones-3 had already decided to perform the action prior to being given the threat, and so it is difficult to attribute his action to either the threat or to his prior decision. The Jones-4 case, in which Black has the power to directly cause Jones-4 to decide on a particular course of action, is structured the way it is because it eliminates the ambiguity in attributing the origin of the action to either the prior decision of the agent or the influence of Black. But for the purposes of the present paper, we need not take a stand on whether Jones-3 has performed the action because of the threat or because of his prior intention. For the present discussion, we need only consider cases that are clearer; in particular, we can restrict our attention to cases in which an individual has not previously formed an intention to perform the action, but her decision is swayed by a sufficiently severe and highly credible threat. If we hold—as I think we should—that a rational person with an ordinary level of willpower and self-determination could be swayed by a credible threat, then we can have cases in which the person's moral responsibility is at least mitigated, and perhaps eliminated entirely. In short, the principle that is easily motivated by an example like Jones-3 is that:

> An ordinary person, with an ordinary level of autonomy or self-determination may have her moral responsibility at least mitigated by a credible threat of sufficiently serious harm.

Later, I will argue that this uncontroversial principle is relevantly similar to another principle that underwrites the existence of real collective responsibility. In short, my argument will be that there are cases in which every individual in a group is under a coercive threat which emanates from the entire group as a whole; this distinctive type of coercion mitigates the moral responsibility of the individuals in the group, but does not excuse the group itself of its moral responsibility. But before I can give this argument in detail, we must examine some other principles that have been taken to imply the existence of real collective responsibility.

3 Ought Implies Can and the Discursive Dilemma

Attempts to argue for the existence of real collective responsibility fall broadly into two categories. The first category consists of attempts to deploy the 'ought implies can' principle to show that real collective responsibility exists. The second sort of attempt uses structural features of the group to show that the group—but not the individuals—bears moral responsibility. These two strategies are not entirely unrelated. But although both are contentious, and neither of them yield clear examples of real collective responsibility, together they point the way toward a better strategy. In this section, I will briefly discuss each strategy in turn.

The first strategy, that of deploying the 'ought implies can' principle, has a simple logical form. We begin by arguing that there is collective moral responsibility in some case, while leaving open the question as to whether it is merely distributive moral responsibility or whether it is 'real collective responsibility'. We then argue that there is some aspect of the case for which there is moral responsibility, but which no individual could have prevented. If so, then the principle that 'ought implies can' will entail, by a simple *modus tollens* argument, that no individual bears responsibility for that aspect of the example. Therefore, because there is moral responsibility that is borne by no individual in the group, we conclude that the group *qua* group must be the bearer. So such a case must be one of 'real collective responsibility'.

Examples of this argumentative strategy go back at least to Joel Feinberg's paper, 'Collective Responsibility' (Feinberg 1991). This paper centers around a particular example that Feinberg takes to be a clear instance of real collective responsibility. In the example, a train is robbed by Jesse James. The robber is well-armed, and the passengers aboard the train are not. Although Jesse James is clearly morally responsible for the robbery, Feinberg specifies the example so that the passengers also bear responsibility for having been robbed. According to how the case is stipulated, no individual passenger is capable of preventing the robbery (since Jesse James is armed and the passengers are not); any attempt of any single passenger to fend off the robber will be unsuccessful. However, Feinberg stipulates that if the passengers were to rise up together, they could collectively fend off Jesse James at no risk to themselves.

Feinberg argues that if the passengers fail to do so, then they are collectively responsible for being robbed (at least, they bear a portion of the responsibility). If we accept the principle that 'ought implies can', then according to Feinberg, we have to excuse each individual passenger—after all, the example stipulates that no individual could have done anything to prevent the robbery. Thus, if there is responsibility, it must attach to the entire group of passengers as a whole, for only the group *qua* group could have done anything to prevent the robbery from occurring.

Although I shall argue below that the Jesse James case turns out to be quite close to the kind of case that is required, Feinberg's diagnosis of it is not compelling. The major difficulty with the case is that Feinberg explains it by appealing to the 'ought implies can' principle, and his argument depends upon the pair of assertions that the group could have repelled the robbery, while no individual could. There is at least a tension in holding both that:

1. No individual passenger could have risen up to stop the robbery, and
2. the group as a whole could have risen up to stop the robbery.

After all, if (2) is true, then its truth entails that all of the individuals could have stopped the robbery, since the group is simply composed of the individuals. But if so, then the truth of (1) entails that (2) is false. To put the point another way, because the group's behavior supervenes on the behavior of the individuals, it is difficult to so cleanly separate the possible behavior of the group from the possible behavior of the individuals; if it is possible for the group to act in a particular way, then it must be possible for the individuals to act correspondingly. We may put the argument more generally, in the following way. First, we assume that the behavior of a group supervenes on the behavior of the individuals who compose it, in the sense that if the individuals all behave in a particular way, then this fully determines the collective behavior of the group. Thus, if it is possible for the group to perform an action, then this entails that it is possible for the individuals to perform the component actions upon which the group action would supervene. So if the group is judged to have the power to perform a particular action, then we must say of the individuals that they each have the power to perform the corresponding individual actions. Therefore, examples of the form specified by Feinberg cannot be stipulated.[2]

A more subtle approach to the question of real collective responsibility has been used, which depends upon a set of impossibility results originating with Kenneth Arrow's (1950) seminal work on rational preferences, and extended in important recent work by Christian List and Philip Pettit (List and Pettit 2004). These results show that there are situations in which the judgments of a group have logical properties that are in a sense 'disconnected' from the judgments of the individuals. In particular, they show that no method of aggregating the judgments of individuals can guarantee logical consistency, even if the judgments of every individual in the group are logically consistent.

[2]I am grateful to Kirk Ludwig for pressing me on this point in an earlier draft of this paper.

Table 19.1 Alice, Bob, and Carol's beliefs about global warming

	(1)	(2)	(3)
Alice	True	True	True
Bob	False	True	False
Carol	True	False	False
Total	True	True	False

To take a simple example, suppose that a committee of three people, Alice, Bob, and Carol, are charged with writing a report on global warming. They are to decide on the truth or falsity of each of three propositions:

1. Human carbon dioxide emissions have reached a particular threshold.
2. If human carbon dioxide emissions were to reach that threshold, then this would cause global warming.
3. There is global warming.

Let us assume that each of these propositions is open to reasonable disagreement among rational, well-informed people. However, there is one set of beliefs that would be irrational; namely, if someone were to believe (1) and (2), then it would be irrational not to believe (3), because it is simply the logical consequence of the first two propositions. Any other combination is rational (or so we shall stipulate). Now suppose that Alice, Bob, and Carol have the following opinions about (1)–(3):

- Alice believes that all three propositions are true.
- Bob believes that (1) is false. However, he does believe that (2) is true. Thus, because he does not accept the antecedent of (2), he is not rationally required to accept (3). And indeed, Bob does not believe that (3) is true.
- Carol also believes that (3) is false. However, she has different reasons. She believes that (1) is true (thus, disagreeing with Bob), but she does not believe that this level of carbon dioxide emission is sufficient to cause global warming (2).

We may represent their beliefs in the chart in Table 19.1. The relevance to collective responsibility comes into the picture when we consider how their judgments would be combined. To make the example more vivid, suppose that the committee has been charged with writing a report on global warming, divided into three sections corresponding to the three questions above. It may appear to be a reasonable plan for the committee members to vote on the conclusions to be asserted in each section. Suppose that they agree to do so, thereby deploying what List and Pettit refer to as a 'premise-centered' approach to judgment aggregation. When they vote on the first question, there is a majority in support of the claim that human emissions of carbon dioxide have reached the critical threshold, with only Bob dissenting. Accordingly, their report will assert that the first question has a positive answer. The same holds for the second question (with only Carol dissenting), and so they would write in their report that this level of emissions would be sufficient to cause global warming. But when they turn to the conclusion of the report, they vote that global warming does not exist, despite the fact that they have collectively agreed to a set of conditions that logically entails that there is global

warming. Thus, they collectively believe an inconsistent set of propositions, despite the fact that no member of the group does.

It has been suggested that in a case of collective judgment that is structured like this one, moral responsibility may attach to the group without any of the individuals bearing responsibility (Copp 2006). For if irrationality is a morally culpable fault, then it attaches to the group as a whole, but not to any individual member of the group.

More generally, such phenomena in judgment aggregation, preference aggregation, and the so-called 'discursive dilemma' show that the behavior and judgment of individuals may come apart from the behavior and judgment of the group. To put the point another way, although the collective judgment of the group supervenes on the judgments of the group's members, the rationality of the group judgment is not entailed by the rationality of the individuals' judgments. An argument for real collective responsibility says that when the moral responsibility of the group is tied to the rationality of the group, then so too, the moral responsibility of the group can come apart from the moral responsibility of the members. This sort of case is a much more substantive argument for the existence of real collective responsibility because it allows us to deny any version of the premise that caused problems for Feinberg's case. That is, in Feinberg's case, the fact that group actions supervene on individual actions makes it difficult to conclude that we have a genuine case of real collective responsibility. But the formal results due to Arrow and to List and Pettit show that the rationality of a group does not supervene on the rationality of the individuals. And this fact opens up the possibility that group rationality and individual rationality may come apart, thereby creating the possibility of real collective responsibility.

The argument does face a different challenge, however. This challenge is to point out that in order for no individual to bear moral responsibility, the example must be specified in such a way that there is no individual upon whom we can blame the failure of the judgment aggregation procedure. For example, if a committee were in danger of falling into the sort of situation faced by Alice, Bob, and Carol, then one might reasonably argue that it would be the responsibility of each of them to seek a way to avoid that outcome. Perhaps, for instance, each could have proposed a new method of aggregating their judgments—one that would not have yielded such paradoxical results. Or suppose, for instance, that the committee was charged with their task by the president of their university, who also required that they aggregate their judgments by voting on each question separately. Then it would be reasonable to lay the moral fault at the feet of the president, since it was at least foreseeable that the group would fall into this logical trap.

But suppose that there was, in fact, no way for any of the individuals to have foreseen or taken steps to avoid the situation that Alice, Bob, and Carol found themselves in. If so, then it seems that the group necessarily found itself in this judgment aggregation paradox. But if that is the case, then it is difficult to see how the group could have failed in any meaningful way to have met its responsibilities. For the principle that ought implies can would seem to imply that, because there was no way for the group to avoid its collective irrationality, then that failure cannot be a failure to meet any moral obligation.

4 Equilibria and Real Collective Responsibility

The failures of the previous cases to provide a clear-cut example of real collective responsibility do not show that the task is hopeless. On the contrary, I think that we learn a few important lessons that help us see how to construct a genuine example. In this section, I shall draw out those lessons and argue for the existence of real collective responsibility. The class of examples I shall develop will point the way toward a general account of this phenomenon.

Recall that in the Jones-2 case, we excuse Jones-2 because he faces a credible threat of serious harm. In examining the first three of the Jones cases, we are led to the common-sense conclusion that a reasonable human being may be excused of moral responsibility for at least some (otherwise) immoral acts if she faces such a threat. Furthermore, it is not required that the person be 'stampeded' into performing the action. Rather, a person can be in control of her rational faculties and simply perform the required action because she properly understands the risk to herself if she refuses. Of course, a person could still refuse on general principle. But we would typically say of such a person that she acted heroically, and that her refusal to perform the action was supererogatory, not morally required. For instance, if a person is ordered to rob a bank under threat of being killed by a bomb, and the threat is credible, then only a hero would refuse the request. Any reasonable person would comply, and we would ordinarily excuse the person of any moral responsibility for robbing the bank, despite the possible existence of such heroic individuals.

Of course, such cases do not concern *collective* moral responsibility—they are merely cases in which individual responsibility is mitigated. However, as I have mentioned above, there is a substantive link between mitigation of moral responsibility and real collective responsibility. This link is that real collective responsibility occurs if there is moral responsibility, but every individual's responsibility has been mitigated to a sufficient degree by a credible threat of serious harm.

The second lesson is from both Feinberg's 'ought implies can' case, as well as from the judgment aggregation case of List and Pettit. What both of these purported examples have in common is that they rely on some structural feature of the group to support the claim that there is real collective responsibility. That is, real collective responsibility is to be explained by the fact that the individuals' actions are somehow constrained by the dynamics the group finds itself in. In the Feinberg case, the relevant dynamic is that in order to successfully stop the robbery, everyone on the train would have to act together, and this coordination is difficult to achieve. In the judgment aggregation case, the dynamic is that the group must aggregate its judgment according to a specific procedure that is vulnerable to the judgment aggregation problem that arises. If either of these dynamics were changed, it would be much more difficult to maintain that there is real collective responsibility. For example, if it were easy for the passengers on the train signal to each other that they should all rush the robber at the same time, we would probably be far less likely to excuse the individuals of their moral responsibility. Similarly, if each member of

the committee had the power to suspend their voting procedure and call for a new procedure to be developed, then we would also be far less likely to say that this is a case of real collective responsibility.

Combining these two lessons, we are led to consider whether there are cases in which each individual in the group faces a credible threat constraining her individual action, but in which that threat is due to a structural feature of the group that no individual can evade.

But this set of conditions is quite familiar. It describes cases in which a group has an equilibrium behavior, where that equilibrium is highly suboptimal, but in which deviation from that equilibrium will be severely penalized. In the following section, I will explain this condition in more detail, and argue that it both motivates the existence of real collective responsibility, while also explaining some of the most plausible features of that form of responsibility.

5 The SANE Approach

The equilibrium concept I shall use here is due to John Nash, from his seminal paper, 'The Bargaining Problem' (Nash 1950). To understand the Nash equilibrium concept, we consider a set of agents, each of whom faces a choice between two or more actions. In order to be non-trivial, the situation should be a *strategic game*, meaning that each individual will be rewarded or penalized based not only on her own choice of action, but also upon the choice of action of the others. We say that a set of players is in a Nash equilibrium if each player has no incentive to switch strategies, provided that nobody else switches. Put in a slightly different way, each player is getting as high a payoff as possible, given the strategies of the other players.

Perhaps the most widely discussed such game is the so-called 'Prisoners Dilemma', which also happens to be relevant to the present discussion. In this game, each individual has a choice between a cooperative and a non-cooperative action, typically labeled *C* (for cooperation) and *D* (for defection). The game is characterized by the fact that the sum of their payoffs is highest when both players cooperate, despite the fact that each player is better off by defecting, no matter what the other player does. It is therefore a dilemma in the sense that the agents jointly prefer to both cooperate, but each individual has an incentive not to cooperate. A payoff matrix for the game is given in Fig. 19.1.

The Prisoners Dilemma is a particularly clear case for illustrating the Nash equilibrium concept because it is so easy to see that (*D*,*D*) is the unique equilibrium. For suppose that a player is playing *C*. No matter what the other player does, it would be best to switch from *C* to *D*, for either her payoff would improve from 4 to 5, or from 0 to 2. Because both players face exactly the same choice, they must both play *D*, and so that is the unique Nash equilibrium. This is notwithstanding the fact that their combined payoff would have been better if they had both played *C* (with a combined payoff of 8 if they both cooperate, but merely 4 if they do not).

Fig. 19.1 The Prisoners Dilemma. The row player's payoff is first in each pair, and the column player's payoff is second

		Player 1	
		C	**D**
Player 2	**C**	4,4	0,5
	D	5,0	2,2

Of course, I am not arguing that any simple formal condition such as an equilibrium will fully characterize real collective responsibility. However, with appropriate additions, the Nash equilibrium concept goes a long way toward understanding it.

Recall the Jesse James train robbery case advanced by Feinberg. With some appropriate specifications, it is reasonable to interpret the situation as a sort of Prisoners Dilemma. Suppose that it would not be necessary for every single passenger on the train to collectively rise up and prevent the robbery, but that it would take many of them to do so. If sufficiently many people cooperate to stop the robbery from happening, there would be only a small risk to any of them, but there is no risk whatsoever to those passengers who just sit passively and allow themselves to be robbed. In this situation, every passenger prefers that a sufficiently large number of them work together to stop the robbery; but each passenger prefers to sit passively and let others take the risk of doing so. Thus, the passengers will be in a Nash equilibrium if they all sit passively allowing themselves to be robbed. For if we assume that everyone else is sitting passively, there is no incentive (indeed, there is a powerful disincentive) to try to stop the robbery from occurring.

With the example respecified in this way, consider what we would say about the moral responsibility of the passengers. Let us suppose, with Feinberg, that there is moral responsibility borne by the group for failing to prevent the robbery. If we agree that a person's moral responsibility can be mitigated by a credible threat of serious harm, then each passenger's moral responsibility is mitigated here as well; for the example stipulates that even if a sufficient number of passengers cooperates to stop the robbery, each individual who does so is still taking a serious risk. Thus, if there is moral responsibility in such a case, there must be real collective responsibility.

Note that this argument does not depend upon the principle that ought implies can. In fact, we have stipulated that the individuals can rise up, individually and collectively, to stop the robbery. But they can do so only in the sense that a heroic person could do so. And the group of passengers is collectively capable of doing so only insofar as it is possible for the train to contain a large number of heroic individuals. What mitigates each individual's moral responsibility is not that they *cannot* stop the robbery, but that it is very *risky* for them to stop the robbery.

Another case will help clarify important differences between Feinberg's argument and the argument I am advancing here. Let us suppose that global warming will occur unless a sufficiently large number of people recycle their trash. However, no individual's recycling their trash will have any positive effect at all. Now suppose—realistically enough—that given these facts, nobody recycles their trash and global warming occurs. What are we to say of each individual's responsibility for global warming?

On Feinberg's account, we must say that we are collectively, but not individually, responsible for global warming. This is simply because it is not possible for any individual to make any positive impact by recycling her own trash. Thus, if Feinberg were right to deploy the 'ought implies can' principle, then we are each excused of any moral responsibility for global warming.

I think this is not a reasonable conclusion to draw about our moral responsibility. Rather, we each surely bear some, perhaps small, responsibility for our contribution to global warming. If anything is a clear example of distributive moral responsibility, it is when each person knowingly contributes to the production of a serious harm, despite the fact that each person's contribution is fully voluntary, and there is no risk for refusing to contribute to the harm.

5.1 Relevance of Equilibria to Real Collective Responsibility

The Nash equilibrium concept is useful for understanding real collective responsibility because it concisely expresses an important way in which the behavior of rational individuals is constrained by the structure of the group and the decision problem the group faces. It shows that a group may find itself trapped in an undesirable pattern of behavior because each individual is constrained by the collective behavior of everyone else in the group, despite the fact that the individuals may find this collective behavior to be quite undesirable.

Here, it is useful to consider coercion once more, and how coercion excuses or at least mitigates an individual's moral responsibility. Suppose I put a gun to your head and order you to perform an action that is morally repugnant. If you justifiably believe that I will fire the gun if you refuse, then you are surely excused for performing the action. And this is true even if the situation is less than life-threatening. If I 'merely' threaten to shoot you in the leg if you refuse, then only a hero would be expected to stand her ground. You are surely not blameworthy for failing to behave heroically. In this way, the blame is properly refocused on me, because I coerced you into performing the action.

In some situations, an entire group of individuals may face an equally unattractive choice—it may be the case that every member of a group faces a similar threat of serious harm if they fail to behave in a particular way. Clearly, these considerations apply without serious modification if someone is literally holding a gun to the head

of every member of a group. But what the Nash equilibrium concept shows us is that the threat may be directed at each member of the group *by every other member of the group*.

To see this, consider the following example. Suppose that in a particular country, it is difficult for individuals to survive economically unless they invest in a dubious financial scheme. Worse yet, as resources are diverted into it, there is less left over for those who refuse to take part, making it more difficult for people to exit the scheme. If the threat of serious harm is great enough, it is only reasonable to judge that the individuals who are involved in the financial scheme are 'trapped', and the consequences are dire enough that we would largely excuse them for failing to take a principled stand against it. Suppose that as a result, there is an economic crisis and serious harm ensues. If we take seriously the analogy between individual-level coercion (as in the Jones cases) and group-level coercion, then we should conclude that the moral fault lies with the source of the coercive force. In this case, the coercion came from the group of investors as a whole, and the coercion was directed at the very individuals who compose the group. Accordingly, the group bears moral responsibility, while the individuals in the group do not. In other words, this is a case of real collective responsibility.

5.2 Beyond Formal Considerations: Stability and Accessibility

Of course, no substantive moral claim will be fully understood in terms of a purely formal concept such as that of a Nash equilibrium. For example, suppose that a terrible consequence can be averted if a large enough number of people contribute one penny to a charity. If nobody contributes to the charity, and the terrible consequence ensues, then the individuals are in a Nash equilibrium. To see this, note that no individual has an incentive to contribute a penny to the charity, on the assumption that nobody else does; for by stipulation, no single individual's contribution will avert the disaster. So each person faces a choice between contributing a penny and not contributing a penny, with the disaster occurring regardless of their choice (assuming that nobody else is contributing). Each person therefore prefers to keep their penny, so this is a Nash equilibrium.

Obviously, it is too much to simply excuse everyone of their moral responsibility if the disaster could have been averted with only a trivial contribution from enough people. In other words, the threat of losing a penny is not sufficient to constitute a credible threat of serious harm of the severity that normally excuses an individual in cases of coercion.

In game-theoretic terms, the severity of the threat—or more generally, the strength of the penalty for deviating from the equilibrium—is often characterized as determining the 'stability' of the equilibrium. The motivation for referring to this as the equilibrium's 'stability' is that although deviations from equilibria are virtually always possible, deviations are less likely to occur if the penalty for deviating is

very great. For this reason, the equilibrium will be more likely to persist in such a case. On the other hand, if the penalty for deviation is very small—as in the case where it only costs a penny to behave out of equilibrium—we would be much more likely to see the group move away from the equilibrium behavior.

The fact that the stability of an equilibrium comes in degrees fits well with our pre-theoretic attributions of real collective responsibility, as well as with our intuitive judgments about coercion. After all, we are less likely to excuse a person of moral responsibility if they were 'threatened' with some trivial harm than we would be if they were threatened with a more serious harm. Clearly, mitigation of moral responsibility comes in degrees, regardless of whether we are considering the actions of an isolated individual, or the actions of a number of individuals in a group. Because the degree of moral responsibility is at least partially determined by the severity of the threat, the equilibrium account of real collective responsibility coheres well with our pre-theoretic notions.

It is also worth considering a second way in which the equilibrium concept highlights an important feature of real collective responsibility. We may think of stability as a measure of how likely or unlikely it is that a group will leave an equilibrium state once that equilibrium has been reached. But there is also a question of how likely it is that a group will enter an equilibrium state in the first place. This is not answered merely by citing the stability of the equilibrium, since an equilibrium may be highly stable and yet difficult to reach, owing to a variety of possible features of the situation. To take a timely example, it may be the case that there is a set of financial and economic reforms that would be very efficient and therefore highly stable if those reforms were ever enacted. But because of various political obstacles, it may be very unlikely that such reforms would ever be enacted, despite the fact that there is widespread agreement that the reforms are quite desirable. Let us refer to this property of an equilibrium as its 'accessibility'.

Accessibility comes in degrees, just as stability does. On one end of the spectrum, the equilibrium may simply be the *status quo*, the state the group immediately finds itself in. This is the case in Feinberg's description of the Jesse James robbery case. Presumably, when the robbery begins, the passengers are not particularly organized, and there is no mechanism in place for them to coordinate their actions with each other. This feature, despite the fact that it is only implicit in Feinberg's description of the case, lends plausibility to the assertion that the passengers, in some sense, could not have coordinated a response to the robbery. If, for some reason, the passengers had been in a highly organized state—perhaps if they were all members of the same organization and they frequently are called-upon to coordinate their actions—we would be more likely to find them morally blameworthy.

On the other end of the spectrum, it may be the case that a group goes through a lengthy and deliberate process to place itself in an equilibrium state. For example, once the bank robbery has begun, it may be true that each robber has a strong disincentive to stop performing their part of the robbery—for instance, it may increase their chance of getting caught if any of them stopped doing their part. In this way, the robbers may be in a very strongly stable equilibrium. But we do not normally excuse the individuals of their moral responsibility in such a case. This

is because, despite the fact that they are in an equilibrium, that equilibrium state was not very accessible insofar as it took a significant and deliberate effort to place themselves in that state to begin with.

The proposal, therefore, is that real collective responsibility exists in a situation to the extent that the group acts as a result of finding itself in a Stable, Accessible, Nash Equilibrium. In a slogan, this is the SANE approach to real collective responsibility.

6 Conclusion

Real collective responsibility is a particularly difficult concept because there are intuitively compelling arguments both for and against its existence. For instance, we frequently speak as if there is real collective responsibility—as when we say of a corporation *qua* corporation that it is guilty of a crime. Thus, common usage seems to militate in favor of the existence of real collective responsibility. However, it is equally intuitively compelling that if a group is morally blameworthy, then there must have been some failure at the individual level which accounts for it. After all, we are used to thinking of the actions of groups as determined by the actions of the individuals, and this habit makes it reasonable to think of real collective responsibility along the same lines.

For this reason, it is extremely difficult to motivate the existence of real collective responsibility by giving intuitively compelling examples, or by relying on simple moral principles such as 'ought implies can'. But conversely, there is no obvious reductive argument establishing that real collective responsibility must supervene on individual responsibility, especially given the judgment aggregation results showing that other group-level predicates such as rationality are not necessarily a function of the corresponding individual-level predicates.

Thus, I have attempted to take a different tack in this paper. Rather than directly examining purported cases of real collective responsibility, I have argued that such cases are specific instances of a more general class of cases. These are ones in which an individual (or group of individuals) has their moral responsibility mitigated or excused entirely. In such cases, the moral responsibility is transferred, as it were, to the agent who is the source of the coercion. Once we have adopted this perspective, it is clear that there are cases in which every member of a group is coerced, but this coercive force comes from the group itself. Such cases are all too common. We may understand a large variety of collective action problems, moral hazards, economic inefficiencies, and instances of systemic social corruption in these terms.

Despite the fact that real collective responsibility is not fully explicable in mathematical or other formalisms, there is a family of theories that may be quite valuable for describing these examples. As we might hope, the theory of Nash equilibria, which has proven so useful for studying group behavior, can also be pressed into service here. Cases in which the group exerts a coercive force on its own members turn out to be a subclass of Nash equilibria. The concept of a Nash equilibrium proves its worth in this context in much the same way as in economic

contexts. That is, despite the fact that it is not the 'end of the story' for explaining either real collective responsibility or group behavior in general, it does provide an extremely useful framework, which highlights the most important ways in which the analysis must be expanded.

In the case of real collective responsibility, additional questions are relevant. For example, we want to know about the relative strength of the coercive force that the group exerts on its members, and how likely it was the group would find itself in that equilibrium. It is an important virtue of the present approach that these additional factors do not need to be added in any *ad hoc* manner. Rather, these factors are naturally described in ways that cohere well with each other within the equilibrium framework. This has led us to the SANE approach—that real collective responsibility exists when the group is in a stable, accessible, Nash equilibrium. No doubt there are other features that will turn out to be relevant to a fuller understanding of real collective responsibility. The present account will be judged to be successful to the extent that these other features cohere equally well within the equilibrium framework.

References

Arrow, K. 1950. A difficulty in the concept of social welfare. *Journal of Political Economy* 58(4): 328–346.

Copp, D. 2006. On the agency of certain collective entities: An argument from 'normative autonomy'. *Midwest Studies in Philosophy* 30: 194–221.

Feinberg, J. 1991. Collective responsibility. In *Collective responsibility: Five decades of debate in theoretical and applied ethics*, ed. L. May and S. Hoffman, 53–76. New York: Rowman & Littlefield Publishers.

Frankfurt, H. 1969. Alternate possibilities and moral responsibility. *The Journal of Philosophy* 66(23): 829–839.

Lewis, H. 1991. Collective responsibility. In *Collective responsibility: Five decades of debate in theoretical and applied ethics*, ed. L. May and S. Hoffman, 17–33. New York: Rowman & Littlefield Publishers.

List, C., and P. Pettit. 2004. Aggregating sets of judgments: Two impossibility results compared. *Synthese* 140: 207–235.

May, L. 1990. Collective inaction and shared responsibility. *Nous* 24: 269–278.

May, L., and S. Hoffman. 1991. *Collective responsibility: Five decades of debate in theoretical and applied ethics*. New York: Rowman & Littlefield Publishers.

Mellema, G. 1988. Causation, foresight, and collective responsibility. *Analysis* 48(1): 44–50.

Miller, S. 2001. Collective responsibility. *Public Affairs Quarterly* 15(1): 65–82.

Narveson, J. 2002. Collective responsibility. *The Journal of Ethics* 6(2): 179–198.

Nash, J. 1950. The bargaining problem. *Econometrica* 18: 155–162.

Pettit, P. 2004. Groups with minds of their own. In *Socializing metaphysics*, ed. F. Schmitt, 167–193. New York: Rowman & Littlefield Publishers.

Pettit, P. 2007. Responsibility incorporated. *Ethics* 117: 171–201.

Sverdlik, S. 1987. Collective responsibility. *Philosophical Studies* 51(1): 61–76.