COLLECTIVE ACTION AS INDIVIDUAL CHOICE

SARA RACHEL CHANT AND ZACHARY ERNST

1. INTRODUCTION

Collective action raises both conceptual and descriptive questions. The conceptual question concerns when we should count a set of actions performed by individuals as a unified 'collective action'. For example, when a group of passengers aboard a broken bus spontaneously exit the bus and push it off the railroad tracks, we would call this event an instance of 'collective action' – we would say that, in addition to the actions of individual passengers, the group performed the collective action of pushing the bus off the tracks. At least the surface level of our description treats the group as a single agent, and that single agent is the author of an action. The problem of determining when (if ever) such attributions are appropriate belongs to philosophical action theory, and has been undertaken by numerous authors.[1]

On the other hand, we can assume that the concept of collective action is intuitively clear enough, and ask the descriptive question, 'Under what circumstances will groups of people coordinate their individual actions and perform a collective action?' For example, we might ask what feature of our example makes it so salient to the passengers that they should coordinate their efforts. This is a descriptive problem, which may be addressed using the tools of game theory.

There is an important point of contact between these two projects. Epistemic conditions are recognized as playing an important role in both the conceptual and descriptive problems. For an important mark of paradigmatic examples of collective action is that the agents who take part in the action have certain beliefs about the beliefs, desires, and motivations of the other agents. Returning to our example, it is plausible to suppose that part of the reason why the agents each do their part of pushing the bus is that each believes that the others desire that the bus be moved, attribute such desires to the other people on the bus, and so on. Indeed, it is likely that in such cases, the existence of such supporting beliefs is a necessary condition for the performance of the collective action. Accordingly, several action theorists

---

[1]See [5, 6, 8, 16, 21–23, 26–28].

have embedded epistemic conditions into their analysis of the concept of collective action.

Furthermore, it is a familiar point that epistemic conditions about the beliefs of the other agents are frequently assumed in game-theoretic accounts of collective action. For example, equilibrium analyses frequently assume that the payoff structure of the game is common knowledge between the players, and rationality assumptions frequently entail that everyone knows that everyone else is fully rational.[2]

In this paper, we argue that the conceptual problem of characterizing collective action should be informed by game-theoretic analyses of collective action that place epistemic conditions at their center. Specifically, we propose to use modified versions of Ariel Rubinstein's so-called 'Electronic Mail Game' [17] as a model of collective action. By examining this model, we will see that some conceptual analyses of collective action have focused on features of collective action that should be regarded as peripheral, and have failed to consider those features that are most central.

2. Approaches to Collective Action

The concept of individual human action is central to philosophical puzzles regarding agency, free will, rationality, and moral responsibility. For example, it has commonly been assumed that a person can be morally responsible only for her actions and their consequences. Similarly, problems of free will only arise for human action and not for non-actional events; a person is an agent only in virtue of having the ability to perform actions; and a person's actions make her evaluable for rationality.

Because many of these problems can be reiterated at the level of the group, problems of collective action are immediately raised. For example, the concept of moral responsibility is reiterated when we ask under what conditions a group (e.g. a business, government, or angry mob) can be help morally responsible for a collective action [11]. We may also ask under what conditions a group of agents behaves rationally, or under what conditions we may say that a group has an appropriate form of collective agency.

The fact that such an important cluster of problems straightforwardly reemerges at the level of the group suggests that whatever analysis we endorse for individual action might also be reiterated at the level of the group. And although there are several competing analyses of the concept of individual action, they all give central

---

[2]For example, see [1, 2, 17].

place to the role played by individual intentions. Without spending the time on a review of this literature, we may fairly say that there is a consensus that an individual action is some behavior of the person which is brought about in some appropriate way by an intention of that person. To take a famous example from Donald Davidson [10], we may say that 'sinking the Tirpitz' refers to an action of the submarine commander because the Tirpitz's sinking is a causal consequence of something the submarine commander did – namely, pushing the button that launches the torpedo – and that the submarine commander pushed that button in order to carry out his intention. Of course, we are led to different accounts of individual action when we give different accounts of individual intention, as well as different accounts of how the intention 'issues forth' in the behavior.

Thus, a natural approach to a conceptual analysis of collective action is to give a central place to the appropriate kind of intention that issues forth in collective action. This kind of intention is commonly called a 'collective', 'joint', or 'group' intention, and accounts of this form of intention have been offered by Michael Bratman [5], Margaret Gilbert [12–14], Raimo Tuomela [21–26, 28], John Searle [18, 19], Seamus Miller [16], J. David Velleman [30], and others.[3] Those who follow the strategy of analyzing collective action by giving an analysis of collective intention[4] typically assume that once the analysis of collective intention is in place, the analysis of collective action will follow immediately.[5]

One challenge facing an account of collective intention is that there is an extremely wide variety of collective intentions, which may have various necessary and sufficient conditions, depending upon the specific circumstances. For example, we may correctly say that 'Russell and Whitehead had the intention to write the *Principia Mathematica*' and that 'the angry mob had the intention to storm the Bastille'. But the level of coordination, planning, and epistemic conditions required for the formation of the first intention is very different from the conditions required for the second. Furthermore, Russell and Whitehead may have had similar motivations underlying their intention, whereas the various members of the angry mob may have different motivations for storming the Bastille.

---

[3]Elsewhere, we have offered a game-theoretic characterization of the concept of collective intention [9].

[4]In what follows, we shall use the term 'collective intention' neutrally to refer to the entire range of intentions attributable to groups. Such intentions have been called 'group intentions', 'joint intentions', and 'we-intentions'.

[5]The strategy of analyzing collective action by giving an account of collective intention has been criticized in [8].

Because of these important differences among various collective actions and collective intentions, the concept of collective action has proven to be difficult to analyze. In particular, writers on collective action have tended to exhibit a taxonomy of collective actions and collective intentions, offering a slightly different set of necessary and sufficient conditions for each. For example, Raimo Tuomela counts as a kind of collective action cases in which several agents simply have different token intentions of the same type, as when several people simultaneously form the intention to open their umbrellas when it starts to rain [27]. Other types of collective action, according to Tuomela, require highly formalized institutions, as when Congress votes to enact a bill into law.

In this paper, our contention is that we should not be satisfied with a taxonomy of collective intentions, and therefore of collective actions. But we also cannot deny that collective intentions (and hence, collective actions) may vary in the way that other authors have indicated. Rather, we should aim for a unified account which allows us to recover and explain the diversity that is exemplified by various collective intentions and collective actions. In other words, we need to ask the question, 'What explains the diversity among collective intentions and collective actions?' The answer to this question will serve as the foundation of a unified account of collective action.

3. The Electronic Mail Game

If the the approach we advocate is correct, then the task at hand is to come up with a model of collective action having a small number of parameters. This model should have the property that when those parameters are allowed to vary over a reasonable range, the model should allow us to recover the range of conditions which have been taken to be necessary and sufficient for the formation of a collective intention and the performance of a collective action.

Elsewhere, we have offered informal arguments that epistemic conditions are the crucial features of an analysis of collective intention [9], and that a game-theoretic model of those epistemic conditions is appropriate. Here, we consider a well-known model – namely, Ariel Rubinstein's 'electronic mail game' [17] – in which epistemic conditions *prevent* the formation of a collective intention and thereby the performance of a collective action. By diagnosing the features of the example which prevent the performance of the collective action, we shall be led indirectly to a better understanding of the concept of collective action.

$\mathbf{G_a}$	A	B
A	M,M	$0,-L$
B	$-L,0$	$0,0$

$\mathbf{G_b}$	A	B
A	$0,0$	$0,-L$
B	$-L,0$	M,M

TABLE 1. Payoff matrix for the Electronic Mail Game. Nature chooses whether the state of the world is a or b, where $P(G_b) = \rho < \frac{1}{2}$. If the state of the world is a, then the players must play G_a. If the state is b, then they play G_b.

3.1. **Common Knowledge and 'Almost Common Knowledge'.** The literature on interactive epistemology now follows David Lewis's definition of 'common knowledge'. According to Lewis, a proposition p is common knowledge to a group of agents just in case:

$$(1) \qquad \qquad \text{(everyone knows that)}^n p$$

for all n [15]. Intuitively, of course, we would not expect to find a significant difference in people's behavior when they have very high degrees of knowledge on the one hand, and full-blown common knowledge on the other. For example, if two agents know that p, and satisfy Lewis's schema for all $n < 10$, we would not expect their behavior to change if they could satisfy the schema for all $n < 11$. In such cases, we would say that the agents have 'almost common knowledge'.

Ariel Rubinstein, however, shows that there can be situations in which agents will coordinate their actions if they have full-blown common knowledge, but will be unable to coordinate if their knowledge falls short of that extreme. Rubinstein's example is the so-called 'electronic mail game'. In it, we suppose that there are two agents – call them Alice and Bob – who face a particular decision problem. Each of them has a choice between playing two actions, which we shall label A and B. If the state of the world is a, then they will benefit by both performing A. If the state of the world is b, then they will benefit by mutually performing B. But if either player plays B alone in state b, then that player incurs a large penalty. The payoff matrices for the game are given in Table (1).

Only Alice is able to observe the actual state of the world. She has the opportunity to communicate a message to Bob through a faulty communication channel (an email system) which has a small probability $(1 - \psi)$ of failing to deliver any particular message. When either player's computer receives any message, a confirmation is automatically sent back to the sender, where that confirmation message also has a probability $(1 - \psi)$ of failing to reach its destination. Eventually, the

computers display a number indicating how many messages including confirmations have been received.

Rubinstein shows that no matter what number is displayed on the players' computers – even if it is a very large number – then the players will not be able to successfully coordinate their actions on B in state b. Informally, the proof goes like this. Suppose that Alice's computer displays the number 1 after all messages have been received. Then she knows that she has received a confirmation message from Bob, but she does not know what happened after that. In particular, she knows that that there are two distinct series of events which may have happened after she received her confirmation:

(1) Her next confirmation was sent to Bob, but it failed to reach him. This may have happened with probability $1 - \psi$.

(2) Her next confirmation was sent to Bob, he received it, but his confirmation failed to reach her. The probability of this is $\psi(1 - \psi) = \psi - \psi^2$.

Alice will reason that option (1) is more probable than option (2), since $\psi - \psi^2 < 1 - \psi$. So Alice believes that Bob most likely did not receive the confirmation for Alice's message.

Intuitively, one might suspect that this should not matter, since Alice does know that Bob received her first message. However, Alice now imagines herself as Bob. In particular, she considers what Bob will conclude after he received Alice's first message, but does not receive a confirmation from her. In such circumstances, Bob will realize that there are two possible explanations for why he did not receive a confirmation:

(1) Perhaps his confirmation of Alice's first message failed to reach her. The probability of this occurring is $1 - \psi$.

(2) Or perhaps his confirmation did reach her, but her reconfirmation did not. As above, the probability of this option is $\psi - \psi^2$.

So now we have the curious conclusion that Alice believes that Bob believes that she did not receive the message. In other words, although Alice's computer has the number 1 on it (signifying that she did receive a confirmation from Bob), Alice believes that Bob believes that her computer shows the number 0.

The argument may be iterated to show that no matter what number is displayed on either person's computer at the end of the email exchanges, each person will act as if the number displayed is zero. Hence, they will be unable to coordinate their actions on option B in state b.

Clearly, this conclusion relies upon several assumptions of the email game which may be unrealistic. For example, it attributes an extremely high degree of rationality to each player, so that they will reiterate the above argument as many times as is necessary.[6] It also relies on the fact that each confirmation message is automatically sent, with neither player having any choice. If either of these assumptions is weakened, then the conclusion may not follow.

However, this is not an objection to using the email game as a model of coordination and collective action. Rather, by determining which features of the situation make coordination impossible, we may better characterize those features which may aid coordination in the real world. This, in turn, may lead us to a better analysis of the concept of collective action. In fact, the email game is a very appropriate case, because it places epistemic conditions at the center of the problem.

3.2. **Formal Preliminaries.** In analyzing the email game, we follow Binmore and Samuelson's [4] game-theoretic presentation of the problem. The first move is made by Nature, in which either G_a or G_b is chosen, where the probability that Nature chooses G_b is $\rho < \frac{1}{2}$. Player 1 receives a message from Nature which tells her which state of the world is actual. If the state is G_b, then Player 1 automatically sends a message to Player 2 informing her of this fact. The message is sent by a communication channel that is not completely reliable, having a probability $\psi < 1$ of reaching its destination. Upon reaching its destination, a confirming message is automatically sent back along the same communication channel. The process of sending confirming messages is repeated until one of the messages fails to reach its destination.

3.3. **Electronic Mail in an Evolutionary Setting.** Binmore and Samuelson propose to examine the electronic mail game within an evolutionary game-theoretic framework. That is, they determine which strategy for each player will be evolutionarily stable in the sense of Maynard Smith. This is closely related to Rubinstein's analysis, which proceeds by way of computing Nash equilibria, since it is a necessary condition for evolutionary stability that the strategies compose a Nash equilibrium. But since the converse does not hold – Nash equilibria are not always evolutionarily stable[7] – the approach of Binmore and Samuelson holds out the possibility of a more fine-grained analysis in which some unlikely equilibria are excluded.

---

[6]Although recent research by Colin Camerer on the Centipede Game suggests that this may not be as improbable as one might suspect [7].

[7]For proof, see Jörgen Weibull [31].

But another argument for examining this problem within an evolutionary game-theoretic framework is that this approach avoids inappropriately strong rationality assumptions. For one of the most implausible features of the electronic mail game is that it attributes an implausible level of rationality to each player. In contrast, evolutionary game theory makes no such assumptions. It simply models a population of agents and assumes only that the most successful strategies will gradually come to predominate in the population. This process by which the successful strategies spread throughout the group may be one driven by rationality – as when players deliberately modify their behavior by a process of imitation. But it may not. It may simply be that the strategies spread by a process of differential reproduction, as in a biological context.

The avoidance of rationality assumptions is particularly appropriate here, since we are examining a situation in which it is supposedly difficult to elevate a proposition to the status of 'common knowledge'. But in Rubinstein's formulation of the game, the players do have some common knowledge – in particular, they have common knowledge of each others' rationality, as well as common knowledge of the payoffs and probabilities required by their calculations. Although it is certainly not impossible that the players have common knowledge of those elements of the game, but not have common knowledge that their various messages have been received, the evolutionary game-theoretic analysis avoids this complication by eliminating the need for strong rationality assumptions. We therefore begin by considering two versions of the electronic mail game that have been given an analysis by Binmore and Samuelson.

3.4. **Electronic Mail with Voluntary Communication.** The first and most obvious modification of the electronic email game is to consider the case where communication is voluntary. So in this form of the game, which Binmore and Samuelson call 'Electronic Mail with Voluntary Communication', each player may choose whether to send a confirmation back to the other player.

Theorem 1. *Let a strategy in the electronic mail game with voluntary communication be a number $n \in \mathbb{N} \cup \{\infty\}$, where the player plays B just in case she has received at least n messages. Then the strategy $n = 1$ is evolutionarily stable.*

Proof. First, we show that a population playing strategy $n = 1$ cannot be invaded by a small group of mutants playing $n = \infty$.

Let us denote the expected payoff of a player $i \in \{1, 2\}$ who plays strategy n in a population α as $\text{EXP}_\alpha(i, n)$. Let α be a monomorphic population of players with strategy $n = 1$. Then player 1 plays strategy B just in case she receives a message

from Nature that the state of the world is G_b. Note that if she does not receive that message, then she will not send a message to player 2; therefore player 2 will play strategy A in state G_a. Thus, we have the following expected payoff:

$$(2) \qquad \text{EXP}_\alpha(1,1) = (1-\rho)M + \rho[\psi M + (1-\psi)(-L)]$$

Player 2 will play strategy B only if the state of the world is G_b and Player 1's message gets through:

$$(3) \qquad \text{EXP}_\alpha(2,1) \quad = \quad (1-\rho)M + \rho M$$

On the other hand, a group of mutants who always play strategy A will receive a payoff of M in G_a and 0 in G_b. So the expected payoff of their strategies are:

$$(4) \qquad \text{EXP}_\alpha(1,\infty) = \text{EXP}_\alpha(2,\infty) = (1-\rho)M$$

Evidently, $(1-\rho)M + \rho M > (1-\rho)M$. So the strategy to play B after receiving a message yields a higher payoff than the strategy of always playing A. Thus, the population α cannot be invaded. $\qquad \square$

At this point, it is worth pausing to reflect on how we have reached the heartening conclusion that coordination may be possible in the electronic mail game after all.

Considerations of rationality and efficiency may easily come apart. That is, it is often the case that a rational strategy is not necessarily the one that will lead to the highest expected payoff among a population of agents. The most common example of this phenomenon is the Prisoners Dilemma. There, simple dominance arguments show that defection is the only rational strategy. But it is still the case that a population may be structured in such a way that cooperation comes to predominate. But as Binmore forcefully argues [3], such cases do not show that cooperation is rational; they merely show that there are environments in which players are better off if they behave irrationally.

The same phenomenon has happened in this simple evolutionary game-theoretic setting. We have here an evolutionary dynamic and set of available strategies in which coordination on B in state b come to predominate by virtue of the fact that such behavior yields high payoffs. But because the evolutionary setting causes coordination to predominate in the population solely in virtue of its higher expected payoff, the model does not reveal which other features may play a role in encouraging coordination to evolve in an arbitrary population.

3.5. **Electronic Mail with Costly Communication.** This observation suggests that we complicate the model in order to examine other features that may affect the evolutionary dynamic. An obvious next step in that direction is by imposing a cost on messages. In Binmore and Samuelson's presentation of this variation, a cost c is paid for the ability to listen to each message. So each agent's strategy is now a pair $\langle n, m \rangle$, where n is the number of messages to which the agent may pay attention, and m is the number of messages which will trigger a play of strategy B. So the cost paid by the agent for the ability to listen to messages is cn. Furthermore, it is important to note that this cost is paid by the player, even if less than n messages are received.

It is easy to show that the strategy $\langle 0, \infty \rangle$, corresponding to the evolutionarily stable strategy of the original electronic mail game, is evolutionarily stable in this new game. We shall follow Binmore and Samuelson in calling this the 'tacit' equilibrium. We now consider whether the tacit equilibrium can be invaded by a strategy in which the players pay an up-front cost to listen to a single message.

Theorem 2. *There exists a cost c such that the strategy $\langle 1, 1 \rangle$ cannot be invaded by the tacit strategy $\langle 0, \infty \rangle$.*

Proof. As before, we consider a population α in which all players play $\langle 1, 1 \rangle$. We consider whether the expected value of $\langle 0, \infty \rangle$ is greater than the expected value of $\langle 1, 1 \rangle$ in α.

For convenience, we denote the tacit strategy t and the strategy $\langle 1, 1 \rangle$ as u (for the 'utilitarian' strategy). The expected value of u for player 1 in the population α is given by:

$$\text{EXP}_\alpha(1, u) = (1 - \rho)M + \rho[\psi M - (1 - \psi)L] - c$$

Similarly, the expected value for player 2 is:

$$\text{EXP}_\alpha(2, u) = (1 - \rho)M + \rho \psi M - c$$

The expected payoffs for mutant t players will be:

$$\text{EXP}_\alpha(1, t) = (1 - \rho)M - (1 - \psi)\rho L$$
$$\text{EXP}_\alpha(2, t) = (1 - \rho)M$$

Player 2's who are tacit can invade only if:

$$(1 - \rho)M > (1 - \rho)M + \rho \psi M - c$$

which reduces immediately to:

(5) $c > \rho\psi M$

In other words, tacit player 2's can invade only if the cost of paying attention is sufficiently large – specifically, only if $c > \rho\psi M$. A similar calculation shows that the same inequality is necessary if tacit player 1's are to invade.   □

3.6. **Lessons from the proofs.** A peculiar feature of Theorems $(1)-(2)$ is that the values of the payoff matrix G_a drop out of the proofs. For upon inspection of those proofs, we find that the only payoff value from G_a is M, and only when it occurs inside the term $(1-\rho)M$.

Intuitively, this fact may strike us as surprising. After all, one might suppose that agents who are attempting to coordinate their actions will care a great deal about all of the potential payoffs in both G_a and G_b. So we have the question of why the structure of G_a appears to be irrelevant.

The reason for this is another assumption that plays a large role in each version of the electronic mail game. In each of these proofs, we assume that Player 1 comes to believe that the state of the world is G_b only if the state of the world really is G_b. Accordingly, when Player 2 receives a message from Player 1, she knows with absolute certainty that the state of the world is G_b. It is not possible for either player to believe that the state is G_b when it is actually G_a. Thus, whenever any messages are received by both players, they know that the values in the matrix G_a are irrelevant. This fact shows itself in the above proofs when the term $(1-\rho)M$ drops out of the relevant calculations.

3.7. **The Electronic Mail Game with False Positives.** But it is reasonable to wonder what happens when the message that Player 1 receives from Nature may be wrong. In other words, we may consider the case in which there is a positive probability $\tau > 0$ that Player 1 receives a message from Nature, even though the state of the world is G_a. We shall call this new version of the game the *Electronic Mail Game with False Positives*. As we shall show, this new versio of the electronic mail game reveals considerably more about the strategic situation than do the versions of the game considered by Binmore and Samuelson.

Since we shall be concerned with whether the values of G_a play a role in the new game, we shall mark them as in Table (2). We now consider the introduction of costs into this augmented game.

$\mathbf{G_a}$	A	B
A	α,α	γ,β
B	β,γ	δ,δ

$\mathbf{G_b}$	A	B
A	$0,0$	$0,-L$
B	$-L,0$	M,M

TABLE 2. For generality, we now consider arbitrary values in a symmetric G_a.

As before, we consider whether a population α consisting entirely of $\langle 1,1\rangle = u$ players can be invaded by a small group of mutants playing strategy $\langle 0,\infty\rangle = t$. First, we calculate the expected payoff of the native u players in the population:

$$\mathrm{EXP}_\alpha(1,u) \;=\; (1-\rho)\Big[\tau\big(\overbrace{\psi\delta}^{M+}+\overbrace{(1-\psi)\beta}^{M-}\big)+\underbrace{(1-\tau)\alpha}_{\text{No error}}\Big]+\rho\Big[\overbrace{\psi M}^{M+}+\overbrace{(1-\psi)(-L)}^{M-}\Big]-c$$

$$\mathrm{EXP}_\alpha(2,u) \;=\; (1-\rho)\Big[\tau\big(\overbrace{\psi\delta}^{M+}+\overbrace{(1-\psi)\gamma}^{M-}\big)+\underbrace{(1-\tau)\alpha}_{\text{No error}}\Big]+\rho\Big[\overbrace{\psi M}^{M+}+\overbrace{(1-\psi)0}^{M-}\Big]-c$$

Next, we have the expected payoff of the mutant t players:

$$\mathrm{EXP}_\alpha(1,t) \;=\; \overbrace{(1-\rho)\big[\tau(\underbrace{\psi\gamma+(1-\psi)\alpha}_{\text{Erroneous message}}+\underbrace{(1-\tau)\alpha}_{\text{No error}}\big]}^{\text{State is }G_a}+\overbrace{0}^{\text{State is }G_b}$$

$$\mathrm{EXP}_\alpha(2,t) \;=\; \overbrace{(1-\rho)\big[\underbrace{\tau\gamma}_{\text{Erroneous message}}+\underbrace{(1-\tau)\alpha}_{\text{No error}}\big]}^{\text{State is }G_a}+\overbrace{0}^{\text{State is }G_b}$$

Mutant Player 1's can invade the population just in case:

$$(1-\rho)[\tau\beta+(1-\tau)\alpha] \;>\; (1-\rho)[\tau(\psi\delta+(1-\psi)\beta)+(1-\tau)\alpha]+$$
$$\rho[\psi M+(1-\psi)(-L)]-c$$

and mutant Player 2's can invade if:

$$(6) \qquad\qquad c > (M\rho+(\gamma-\delta)(\rho-1)\tau)\psi$$

Condition (6) reveals some relevant facts about the role of the game G_a in determining whether the players will coordinate their actions. In particular, it shows that the relative values of γ and δ determine whether the possibility of a false positive message from Nature helps or hinders coordination.

Fact 1. *If $\gamma > \delta$ [$\gamma < \delta$], then increases in the value of τ make it less [more] likely that tacit Player 2's can invade a population of u players.*[8]

We note that Condition (6) contains information about the game G_a whereas Condition (5) does not. This implies a fact which is potentially important for characterizing collective action – namely that the structure of G_a has a stronger effect upon the players' actions when there is some degree of uncertainty about the state of the world.

4. Toward a Unified Account

One goal of a philosophical account of collective action should be to subsume as much of the diversity of collective action under set of concepts that is as small as possible. The problem of providing an analysis of the concept of collective action is an especially challenging one, because collective actions may take a wide variety of different forms under a wide range of circumstances. It is therefore an especially worthwhile test for the tools of game theory, and particularly for formal epistemology within a game-theoretic framework.

When an agent recognizes that there is an opportunity to take part in a collective action, she faces a simple choice – she may either participate in the collective action or strike out on her own.[9] Described as such a choice faced by two or more individuals, collective actions are the result of individuals' strategic decisions. Therefore, a conceptual analysis of collective action should have something to say about the calculations and circumstances that give rise to the decision to participate. To put the picture another way, we may hold out hope for an etiological account of collective action, according to which a group of actions is properly called 'collective' just in case it arose as a result of the individuals' having made a particular kind of strategic decision.

In spite of the fact that the electronic mail game is (at best) a very coarse-grained model of collective action, the considerations we have made so far lead to a number of lessons about the nature of collective action. And although we feel that

---

[8]Furthermore, in the special case where $\gamma = \delta$, Condition (6) reduces to Condition (5) in the analysis of the electronic mail game with listening costs (without listening costs).

[9]Put this way, the problem of collective action is similar to Brian Skyrms's justification for focusing on the Stag Hunt game in his study of the evolution of the social contract [20].

the game-theoretic analysis places this study on a solid formal foundation, those lessons can be motivated informally.

4.1. **The Importance of the *Status Quo*.** If collective actions arise as the result of a strategic decision to participate, then the individuals who are potentially involved in the collective action must compare the likely results of participating with the likely results of not participating. It is therefore important for an analysis of collective action to give an important place to what we might call the '*status quo*' – that is, the likely payoffs that the agent will obtain without taking part in the collective action.

Any mention of the *status quo* is conspicuously absent from existing accounts of collective action. Tuomela's proposed conditions for (what he calls) 'we-intending' to take part in a collective action is representative:

> A member A_i of a collective G we-intends to do X if and only if:
> (1) A_i intends to do X (or his part of X), given that every member of G will do X (or his part of X),
> (2) A_i believes that every member of G will do X (or his part of X),
> (3) there is a mutual belief in G to the effect that (1) and (2).

If an account of collective action must explain why individuals put aside their individual actions in favor of performing 'their part' of the collective action, then we will not be able to explain why 'A_i intends to do X' without having relevant information about A_i's alternative to doing X.

To approach this issue from another direction, we may note that if the analysis of collective action is to have the explanatory function we have claimed it must have, then it must meet necessary conditions for comprising an adequate explanation. It is now a familiar point that explanations serve a *contrastive* role – an explanation of why some proposition p is true must contrast p's truth with the falsity of other propositions in the explanation's contrast class [29]. In the case of explaining why A_i took part in a collective action X, the obvious contrast class consists of a proposition asserting that A_i performed a merely individual action.

If we use the electronic mail game as a model of collective action, then the relevant information about the contrast class is given by the structure of G_a – the game faced by the players in the *status quo*. Furthermore, the electronic mail game shows us that the importance of the *status quo* rises when the agents are faced with a degree of uncertainty about the actual state of the world. It also shows us that the importance of the *status quo* is increased when there is an incentive for at least one

agent to *mis*coordinate in the game G_a; this is a fact we learn from Condition (6), for the relationship between γ and δ determines whether it is better for the player to coordinate or miscoordinate in G_a. Interestingly, the relative value of coordinating on option *A* in the *status quo* game G_a does not show up in Condition (6). Whether or not this is a general feature of collective action problems is a question we leave open here.

4.2. **Degrees of Interactive Knowledge.** Existing accounts of collective action uniformly recognize that many paradigmatic cases of collective action require the agents to have knowledge of each others' state of knowledge. For example, it is a commonplace observation that I might be willing to participate in a collective action only on the condition that I believe that you will participate. Furthermore, if I recognize that you have the same reservations about participating, then I have some incentive to make sure that you know that I know that the right opportunity for collective action exists. So questions concerning the appropriate level of interactive knowledge are clearly important.

However, existing philosophical accounts of collective action have tended to be ambiguous about the level of interactive required, or to be implausibly precise by requiring a particular level of interactive knowledge. In order to make this discussion precise, we shall follow the some distinctions made in the interactive epistemology literature. We say that a proposition *p* is 'mutual knowledge' just in case everyone knows *p*. If Lewis's definition of common knowledge is not true for all $n \in \mathbb{N}$, but is true for all $n \leq m$, then we shall say that *p* is known to degree *m*. And we shall continue to use Lewis's definition of common knowledge as before.

What degree of interactive knowledge is required in order to guarantee collective action is not clear from existing accounts of collective action. Michael Bratman's account of (what he calls) 'shared intentions' is that:

> We intend to *J* if and only if:
> (1) (a) I intend that we *J* and (b) you intend that we *J*.
> (2) I intend that we *J* in accordance with and because of 1*a*, 1*b*, and meshing subplans of 1*a* and 1*b*; you intend that we *J* in accordance with and because of 1*a*, 1*b*, and meshing subplans of 1*a* and 1*b*.
> (3) 1 and 2 are common knowledge between us. [6, p. 106]

Bratman is careful to point out that he is not using the term 'common knowledge' in the technical sense. According to Bratman, we may leave that term unspecified,

presumably because the concept of collective action is to be explained largely in terms of the 'meshing subplans' held by the people involved.

In contrast, our discussion of the electronic mail game is meant to motivate the view that Bratman's explanatory strategy should be inverted. If what we have argued above is correct, then it is the interactive knowledge held by the various agents which explains why they have the plans they do. To return to the variants of the electronic mail game, if Alice has the plan to play strategy B in G_b, then this is because she and Bob have acquired the requisite level of interactive knowledge. In contrast, if we were to follow Bratman's strategy by temporarily setting aside the epistemological issues, then the agents' plans would appear in the explanation of their behavior as an unexplained assumption.

4.3. **Interactive Knowledge as Prior to Social Organization.** Other accounts of collective action (for example, see Tuomela's proposals in [21]) begin by noting that different social structures often correspond to different kinds of collective actions. For example, a large organization like a business or a government might have institutional practices in place to spread information throughout the organization. Or conversely, a hierarchical organization may be structured so that collective actions can immediately be performed as a result of one person's (an executive decision-maker) having formed an intention to initiate that collective action.

But when we examine collective action through the lens of interactive epistemology, we note that the level of interactive knowledge required is a function of the potential costs and benefits the agent must incur if she is to take part in the collective action. The various conditions we have been led to in our analysis of the electronic mail game predict that different levels of interactive knowledge may be required, depending upon the *status quo*, the cost of communication, and the reliability of the information possessed by the players about the actual state of the world. In cases where there is relative certainty about the state of the world, small risk for attempting to take part in the collective action, and high costs for communication, the model predicts that only mutual knowledge will be necessary. By allowing the risks and uncertainty to increase, and by lowering the cost of communication, higher levels of interactive knowledge may be required. The original form of the electronic mail game is therefore a limiting case of this phenomenon in which the costs of communication are zero, the information about the state of the world is perfectly reliable, and the risk for attempting to coordinate is very high. Accordingly, only full-blown common knowledge is sufficient to justify the collective action in that case.

These observations suggest that the structure of the group engaged in the collective action may be relatively unimportant for determining the nature of collective action. Rather, the group structure may simply be a side-effect of the strategic situation in which the agents find themselves. For one of the important functions of social structures is to affect the transfer of information. For example, in an environment in which the risks associated with collective action are relatively high, there is some incentive to organize the group structure so that interactive knowledge is gained cheaply. Conversely, if only mutual knowledge is required for collective action, then there may be no incentive to organize the group in any particular way. And similarly, a group may be organized hierarchically so that the need for interactive knowledge is eliminated – in such a case, many members of the group are obligated to take orders from the top of the hierarchy without having any more complex information about the beliefs, knowledge, or desires of those who are giving the orders.

5. Conclusion: What Is 'Collective Action'?

Formal models of interactive knowledge are valuable tools for understanding the nature of collective action, predicting when it is likely to occur, and identifying the strategic features that either encourage or inhibit it. Furthermore, the game-theoretic models that are used to study the effects of interactive knowledge reveal facts that are important for a philosophical theory of collective action.

When a theory tells us that a phenomenon's apparent diversity is the result of a small number of variables, then the theory thereby reveals which features should figure into a conceptual analysis of that type of phenomenon. For if a large number of that phenomenon's characteristics are observed to vary over a wide range as the result of changes in a small number of its properties, then those latter properties should be taken as central to a philosophical account. In contrast, those characteristics which display a great deal of variation should be understood as more peripheral.

Conceptual analyses of *individual* action assert that the difference between actions and non-actional events is to be characterized by considering the explanation of the behavior in question. For example, we might ask whether Bob's raising his arm is an action of his, or merely some non-actional event that happened to him. The received view on this question is that Bob performed the action of raising his arm only if the correct explanation of why his arm rose involves – in some appropriate way – Bob's intentions.

When we turn to collective action, the relevant contrast is not between collective actions and non-actional events, but between collective actions and sets of individual actions. Successful accounts of individual action suggest that we should identify collective action with a particular kind of *explanans*. We should say that a set of actions is a collective action just in case that set's occurrence admits of a particular kind of explanation.

The majority of this paper has been extended argument that collective actions are explained in large part by interactive knowledge. That is, a set of actions composes a collective action just in case the level of interactive knowledge obtained by the agents explains why they were moved from the best available action in the *status quo*.

REFERENCES

[1] Robert J. Aumann. Interactive epistemology I: Knowledge. *International Journal of Game Theory*, 28:263–300, 1999.

[2] Robert J. Aumann. Interactive epistemology II: Probability. *International Journal of Game Theory*, 28:301–314, 1999.

[3] Ken Binmore. *Playing Fair*. MIT Press, Cambridge, Massachusetts, 1994.

[4] Ken Binmore and Larry Samuelson. Coordinated action in the electronic mail game. *Games and Economic Behavior*, 35:6–30, 2001.

[5] Michael E. Bratman. Shared cooperative activity. *The Philosophical Review*, 101(2):327–341, 1992.

[6] Michael E. Bratman. Shared intention. *Ethics*, 104:97–113, 1993.

[7] Colin Camerer. *Behavioral Game Theory: Experiments in Strategic Interaction*. Princeton University Press, 2003.

[8] Sara Rachel Chant. The special composition question in action. *Pacific Philosophical Quarterly*, forthcoming.

[9] Sara Rachel Chant and Zachary Ernst. Group intentions as equilibria. *Philosophical Studies*, forthcoming.

[10] Donald Davidson. *Essays on Actions and Events*. Clarendon Press, Oxford, 1980.

[11] Joel Feinberg. Collective responsibility. In Larry May and Stacey Hoffman, editors, *Collective Responsibility: Five Decades of Debate in Theoretical and Applied Ethics*, pages 53–76. Rowman and Littlefield Publishers, Inc., 1991.

[12] Margaret Gilbert. *On Social Facts*. Routledge, London, 1989.

[13] Margaret Gilbert. Walking together: a paradigmatic social phenomenon. *Midwest Studies in Philosophy*, 15:1–14, 1990.

[14] Margaret Gilbert. *Living Together: Rationality, Sociality, and Obligation*. Rowman & Littlefield, Lanham, Maryland, 1996.

[15] David Lewis. *Convention: A Philosophical Study*. Harvard University Press, Cambridge, 1969.

[16] Seumas Miller. *Social Action: A Teleological Account*. Cambridge University Press, Cambridge, 2001.

[17] Ariel Rubinstein. The electronic mail game: Strategic behavior under "Almost common knowledge". *American Economic Review*, 79:385–391, 1989.

[18] John Searle. Collective intentions and actions. In Philip R. Cohen and Jerry Morgan, editors, *Intentions in Communication*, pages 401–416. MIT Press, 1990.

[19] John Searle. *The Construction of Social Reality*. The Free Press, New York, 1995.

[20] Brian Skyrms. *The Stag Hunt and the Evolution of Social Structure*. Cambridge University Press, Cambridge, 2004.

[21] Raimo Tuomela. *A Theory of Social Action*. D. Reidel Publishing Company, Dordrecht, 1984.

[22] Raimo Tuomela. Actions by collectives. *Philosophical Perspectives*, 3:471–496, 1989.

[23] Raimo Tuomela. We will do it: an analysis of group-intentions. *Philosophy and Phenomenological Research*, 51(2):249–277, 1991.

[24] Raimo Tuomela. *The Importance of Us: A Philosophical Study of Basic Social Notions*. Stanford University Press, Stanford, California, 1995.

[25] Raimo Tuomela. *The Philosophy of Social Practices: A Collective Acceptance View*. Cambridge University Press, Cambridge, 2002.

[26] Raimo Tuomela. Joint action. *Workshop on Holistic Epistemology and Theory of Action*, 2004.

[27] Raimo Tuomela. We-intentions revisited. *Philosophical Studies*, 125:327–369, 2005.

[28] Raimo Tuomela and Kaarlo Miller. We-intentions. *Philosophical Studies*, 53:367–390, 1988.

[29] Bas van Fraassen. The pragmatic theory of explanation. In Joseph Pitt, editor, *Theories of Explanation*, pages 136–155. Oxford University Press, 1988.

[30] J. David Velleman. How to share an intention. *Philosophy and Phenomenological Research*, 57(1):29–50, 1997.

[31] Jörgen W. Weibull. *Evolutionary Game Theory*. MIT Press, Cambridge, Massachusetts, 1995.