Author: Simon Rosenqvist

Supervisor: Wlodek Rabinowicz

Collective Consequentialism

ABSTRACT

Any plausible consequentialist theory should fulfill the following intuitions: (a) if an agent satisfies the theory she should produce the best possible consequences she can in that situation, (b) if all agents satisfy the theory they should produce the best possible consequences as a group, and (c) all agents should, in fact, be able to satisfy the theory. This thesis shows that while act-, rule- and cooperative consequentialism each fails to fulfill at least one of these intuitions, generalized act-consequentialism is successful by directing itself to both collective and individual agents. It is further argued that these collectives act only in a weak sense, according to which they need not have the ability to act intentionally, or to entertain beliefs or desires. This opens up the possibility for a new theory, collective consequentialism, which directs itself only to the collective agent constituted by all other agents.

CONTENTS

1. INTRODUCTION

When citizens fail to rise against a dictator, we blame their individual self-concern. Had they all acted in the common good, we believe they would have deposed the tyrant and reinstated a rule of the people. But sadly, this need not be true. Even if each individual does what is best, the group as a whole may fail to bring about the best possible consequences. It may be true of each individual that had she tried to protest, she would only have provoked a harsh crackdown without weakening the regime. Because of this, each agent would do her best by remaining passive. Paradoxically, their joint wrongdoing would make each of their acts right.

This problem have long troubled those consequentialists who search for a theory of rightness, or for some criterion of which acts are right. It is particularly worrisome for act-consequentialism, which holds that each agent ought to produce the best consequences she can, but which cannot also guarantee that our acts are coordinated in the right ways. Two other theories have sometimes been offered as alternatives: rule-consequentialism, which calls for each agent to follow an ideal set of rules, and cooperative consequentialism, which requests that the agent cooperate with others. In this thesis, I show how all of these theories are defective in some regard. Merely telling individuals what to do is a non-starter.

As of today, the only successful solution to this problem is provided by so called generalized act-consequentialism, which directs itself to both individual and collective agents. Incorporating collective agents into a consequentialist theory is controversial, but as I shall argue, a weaker view of collective agency can be made plausible enough. The best consequentialist theory is therefore a collectivist one. When we fail to do our best as a group, it is the collective which has done wrong.

The disposition of the thesis goes as follows. Section two suggests three intuitions which any consequentialist theory of rightness should fulfill. Section three shows how act- and rule-consequentialism fail to fulfill our first or second intuitions as a result of certain coordination problems, and section four demonstrates how cooperative consequentialism is successful only when it fails to fulfill our third intuition. Section five suggests that these results are likely to hold for all theories directing themselves to only individual agents, and section six moves on to discuss generalized act-consequentialism and its inclusion of collective agents in morality. Finally, section seven suggests a new theory: collective consequentialism.

2. THREE INTUITIONS

Consequentialist theories combines some axiological account of which consequences are good with a theory of rightness. Regarding which consequences are good, the literature abounds with suggestions. For example, utilitarians claim that the best outcome is a possible state of affairs which contains the maximum total utility, but they differ on how to interpret this claim: while hedonist utilitarians claim that maximum utility consists in the maximum amount of pleasure minus pain, preference utilitarians argue that it consists in the maximum amount of satisfied preferences minus frustrated preferences. Other consequentialists take a broader stance than the mere promotion of utility. For example, some include loyalty and friendship as directly contributing to the best outcome. In what follows, I say little of such questions of which consequences are good, and I expect most views of goodness to be compatible with the theories I am to discuss. Also, note that while many philosophers cited in the text speak of utilitarianism rather than consequentialism, these differences concern mostly which consequences are good, and not which theory of rightness is correct. Since I bracket the question of goodness in its entirety, I choose to speak consistently of consequentialism. This should not affect any of the conclusions which follows.

When discussing which theory of rightness is correct, my aim is to find an adequate consequentialist criterion of which acts are right, rather than to suggest a method of decision making, a theory which we should try to follow, or even one we should accept.[1] Furthermore, I am concerned with the objective consequences of our acts; that is, with the actual rather than the expected outcomes. In short, I am interested in which theory of rightness we ought to successfully follow, and I will adopt the common usage of saying that an agent *satisfies* a theory whenever the agent has achieved whatever requirements the theory directs her to fulfill. Whether a theory directs itself to an agent is indicated in the beginning of each statement of the theory.

To choose between different theories of rightness, we need some criteria which they should aspire to fulfill. In what follows, I proceed from three intuitions. The first and second of these have been suggested by Donald Regan:[2]

Intuition 1: For any theory T, if an agent satisfies T, then that agent should

---

1 For a discussion on the difference between consequentialism as a criterion of rightness and a method of decision making, see R. E. Bales, "Act-utilitarianism: Account of Right-making Characteristics or Decision-making Procedure?", *American Philosophical Quarterly* 8 (1971), pp. 257-65.

2 Regan's formulations are modified to fit the current text. Instead of intuitions, Regan speaks about two properties which different theories can have, which he calls PropCOP and PropAU. See his *Utilitarianism and Co-operation* (New York, 1980), pp. 3-5.

produce by her act the best consequences she can possibly produce in that
situation.

Intuition 2: For any theory T, if all agents satisfy T, then the class of all
agents should produce by their acts taken together the best consequences
that they can possibly produce by any pattern of behavior.

The first intuition states that each agent should relate to the best possible consequences by way of *maximizing* those consequences. The second intuition adds that the agents must be coordinated in this task, so that they also maximize the best consequences *as a group*.

Fulfilling these two intuitions is a necessary constraint, but not a sufficient one, on any consequentialist theory of rightness. In addition, we also want the agents to *to be able to satisfy* the theory. Otherwise we could produce a theory which fulfills our first and second intuitions simply by making it impossible for any agent to satisfy. All agents satisfying the theory would then, trivially, produce the best consequences possible both individually and as a group. But this is obviously not the kind of theory we are looking for. We must therefore add to our first and second intuitions a variant of the "ought-implies-can"-principle:

Intuition 3: For any theory T: all agents should be able to satisfy T.[3]

Our third intuition is simply a plausible extension from our second intuition; if we demand that the universal satisfaction of a theory guarantees that the agents produce the best possible consequences together, it is reasonable to want such universal satisfaction to be possible.

3. ACT- AND RULE-CONSEQUENTIALISM

In this and the following two sections, I assume only the existence of individual agents. More precisely, I take an individual agent to be an isolated time-slice of a person with a specified list of alternative acts open to her. This gives us what we may call an *atomic* conception of individual agency. On this view a person is made up of a number of individual agents during the course of her life. This is perhaps slightly counterintuitive, but upon closer reflection it may seem alright. Our

---

3   Note that this is a strong version of the "ought-implies-can"-principle. We are demanding that *all agents* are able to satisfy the theory, rather than that *any* agent can do so. To see the difference, we need only imagine a theory which any agent can satisfy, but which can only be satisfied by one agent at a time. Such a theory fulfills the weaker, but not the stronger, version of the "ought-implies-can"-principle. I believe that the stronger version makes more sense in the context of our second intuition, but perhaps this can be disputed.

present and future selves certainly differ from each other to some extent, both regarding our intentions and how we act. For example, yesterday-you may have intended to write a paper, but today-you may be more interested in taking a day off. A way to make sense of this discrepancy is to say that yesterday-you and today-you are two different agents. This account of individual agency also helps us to avoid intrapersonal coordination problems.[4] Without it, the problems which we are to discuss may simply reappear within the course of a person's life.

Assume now that each agent acts from among her list of available alternatives, and that she produces different consequences depending on which alternative is selected. Let us borrow a term from Regan and call theories which directs the agent only to pick one alternative over another for *exclusively act-oriented*.[5] Since, by stipulation, agents can always select one of the acts available to them, constructing exclusively act-oriented theories makes for an expedient way of fulfilling our third intuition. But as Regan and others have argued, such theories cannot at the same time fulfill both our first and second intuitions.

Let us look closer at this claim. Consider first act-consequentialism (abbreviated AC, from this point onward):

> AC: For each individual agent: The agent ought to do the act which has at
> least as good consequences under the circumstances as any other act open to
> the agent.

As Regan notes, AC clearly fulfills our first intuition.[6] If an agent satisfies AC she does the act which has at least as good consequences as any other act; that is, she produces the best consequences she can in that situation.

Furthermore, AC is an exclusively act-oriented theory, since it requires only that the agent does one of the acts which are open to her. Since this is something which the agent always can do, AC fulfills our third intuition as well.

However, AC does not fulfill the second intuition. Satisfaction by all agents of AC does not guarantee that the agents as a group produce the best consequences they can possibly produce by any pattern of behavior. This is shown by the following example:[7]

---

4   Cf. Regan, *Utilitarianism and Co-operation*, p. 150.

5   Regan, *Utilitarianism and Co-operation*, pp. 105-23.

6   Regan, *Utilitarianism and Co-operation*, pp. 17-18.

7   Several philosophers have pressed this challenge against AC. See, for example, Allan F. Gibbard, "Rule-utilitarianism: Merely an Illusory Alternative", *Australasian Journal of Philosophy* 43 (1965), pp. 211-20; Gerald Barnes, "Utilitarianisms", *Ethics* 82 (1971), pp. 59-62; Regan, *Utilitarianism and Co-operation,* pp. 12-53; Fred

Imagine two persons, Sarah and Joshua, who are working in a field hospital, in which they attempt to help a number of wounded patients. Assume that each of them because of the high level of noise and the crowded room cannot communicate with or otherwise influence the actions of the other. Also, suppose that both Sarah and Joshua can help some moderately wounded patient on their own, thereby producing a modest outcome (with a value of 6). On the floor lies a seriously wounded patient on a stretcher. If both Sarah and Joshua walk over to the patient and lift the stretcher, they carry the patient to a doctor, which results in the best possible outcome (valued 10). However, if only one of them walks over, he or she cannot lift the stretcher, while at the same time eschewing the opportunity to help another patient (valued 3). This tragic outcome is the worst possible. The problem for AC is this: if Sarah and Joshua both decide not to walk over and lift the stretcher, it is true for each agent that, according to AC, he or she did the right thing. Had Sarah walked over, the worst outcome would have resulted. The same goes for Joshua. Strangely enough, their joint wrongdoing makes each of their acts right.

The above case is an example of a *pure coordination game*, played by two consequentialist agents with the same evaluation of each pattern of action:

	Sarah lifts	Sarah does not lift
Joshua lifts	10	3
Joshua does not lift	3	6

Figure 1: Example of pure coordination game.

In this situation, there are two possible situations in which no agent can single-handedly improve the outcome by acting otherwise (so called Nash equilibria). The problem is that one of these outcomes is sub-optimal. As a result, even if all agents satisfy AC, there is no guarantee that they end up in the optimal situation, thereby producing the best consequences together.[8] Because of this, AC fulfills our first and third, but not our second, intuition.

Could we solve the problem, not by telling each individual to do what is best, but by providing each agent with a set of rules to follow? Consider rule-consequentialism (RC):

Feldman, *Doing the Best We Can* (Dordrecht, 1986)*,* pp. 154-6.

8   It is worth noting a parallel discussion on whether act-consequentialism is even *consistent* with the production of the best consequences possible. Regan has argued for such consistency, though Wlodek Rabinowicz has noted that this is true only if the theory gives the agents common aims, which it may fail to do under some circumstances. In what follows I presuppose the agents being given common aims as a constraint on any consequentialist theory. For more on this, see Regan, *Utilitarianism and Co-operation*, pp. 54-65; Rabinowicz, "Act-utilitarian Prisoner's Dilemmas", *Theoria* 55 (1989), pp. 1-44.

RC: For each individual agent: An act is right if and only if it conforms with an ideal set of rules; an ideal set of rules is any set of rules such that if everyone always did, from among the things he could do, what conformed with that set of rules, then at least as much good would be produced as by everyone's always conforming with any other set of rules.[9]

Suppose a set of rules prescribes that both Sarah and Joshua lift the stretcher. This set of rules is ideal, since if both Sarah and Joshua follow these rules, at least as much good would be produced as if they conformed with any other set of rules. In this case, when both Sarah and Joshua satisfy RC they produce the best outcome together. So it may seem as if RC fulfills our second intuition.

RC clearly fulfills our third intuition. Satisfying RC requires nothing but that the agent selects an act available to her; though this act is now recommended by an ideal set of rules rather than, as in the case of AC, on the basis of whether it produces the best consequences. RC is therefore an exclusively act-oriented theory.

Regretfully, however, there are serious defects with RC. One objection goes as follows. Let us assume that one agent declines to follow the ideal set of rules. This raises the question of what, in such a case of less than universal satisfaction, RC tells everyone else to do. Return to our first example. If Joshua decides not to lift the stretcher, what should Sarah do? Satisfying RC requires her to conform with an ideal set of rules, but if these rules tells her simply to lift the stretcher, then conforming to them does not lead her to produce the best consequences she can in that situation. In this case, given Joshua's intransigence, it is clear that Sarah ought instead to help a moderately wounded patient.

As Regan has argued, the most natural way to solve this problem is to introduce conditional rules which prescribe different acts to an agent depending on what the other agents do.[10] Suppose that we list all the different circumstances which can arise for an agent, and by the side of each item write a prescription for action for that agent. The idea is that once a set of rules take into account all the possible ways in which Joshua might behave, Sarah's compliance with this set should make her produce the best possible consequences she can in that situation.

But according to Regan, this does not work. To see this, let us follow him in considering the

---

9   This formulation is fetched from Barnes, "Utilitarianism", p. 57, and is discussed by Regan in *Utilitarianism and Co-operation*, pp. 90-3. Derek Parfit has recently argued for an almost identical theory, stating that "[e]veryone ought to follow the principles of which it is true that, if they were *universally followed*, things would go best", in his *On What Matters (*Oxford, 2011)*, pp. 404-6. Regrettably, Parfit suggests no solution for the problems facing RC.

10 Regan, *Utilitarianism and Co-operation,* pp. 86-7.

following two sets of conditional rules:[11]

Sarah	Joshua
If Joshua lifts, lift.	If Sarah lifts, lift.
If Joshua does not lift, do not lift.	If Sarah does not lift, lift anyway.

Figure 2: Ideal conditional rules, set A.

Sarah	Joshua
If Joshua lifts, lift.	If Sarah lifts, lift.
If Joshua does not lift, lift anyway.	If Sarah does not lift, do not lift.

Figure 3: Ideal conditional rules, set B.

Both of these sets are ideal, since the universal satisfaction of any set would cause the agents to produce the best possible consequences together. But suppose that both Sarah and Joshua do not lift the stretcher. Both of them satisfy RC, since each of them follows an ideal set of rules: Sarah follows set A, and Joshua follows set B. Each agent also produces the best consequences he or she can in that situation. But they do not produce the best possible consequences as a group. Just as AC, RC fails to fulfill our second intuition.

Another problem concerns the possibility of there existing multiple ideal sets of rules.[12] This point is best made by considering a case with two equally good outcomes. Take a different example with Amanda and John, where the benefit of both agents helping a moderately wounded patient on their own is four units higher. The outcome is illustrated by the following matrix:

	Amanda lifts	Amanda does not lift
John lifts	10	5
John does not lift	5	10

Figure 4: Example of game with multiple best outcomes.

Suppose Amanda lifts the stretcher, while John does not lift. Both of them satisfy RC, since each conform to an ideal sets of rules; i.e. had both either lifted or not lifted the stretcher, the best consequences would have been brought about. But since they conform to different sets of rules, neither of them produces the best possible consequences on their own. Nor do both of them produce

---

11 Regan, *Utilitarianism and Co-operation,* pp. 90-3.

12 This has been argued by Jordan Howard Sobel, "Rule-utilitarianism", *Australasian Journal of Philosophy* 46 (1968), pp. 154-55, and Regan, *Utilitarianism and Co-operation,* pp. 83-93.

the best consequences together. In fact, RC would fulfill neither our first or second intuition.[13]

We may think that this problem can be solved by referring to a salient feature or focal point in the formulation of our theory, thereby somehow guiding the agents to pick the same set of ideal rules. For example, Brad Hooker has suggested that we follow the ideal set of rules which lies closest to the set of conventionally accepted rules.[14] But Hooker's suggestion suffers from similar problems. First, he does not explain what to do when two ideal sets of rules lies equally close to the conventionally accepted set of rules. And second, there could be multiple sets of conventionally accepted rules as well.

How serious is this problem for RC? Somehow RC needs to recommend a specific ideal set of rules, but it is not clear how it may accomplish this. Perhaps RC is a plausible method of decision making, or a theory we should accept. But as a theory of rightness it is just as defective as AC.

4. COOPERATIVE CONSEQUENTIALISM

If coordination problems arise because each agent is acting on her own, perhaps we should require that the agents *cooperate*? Let us label such theories under the fitting heading of cooperative consequentialism (CC). Can CC help us to bypass the problems facing AC and RC?

Regan has suggested a variant of CC, according to which each agent ought to cooperate with whoever else is cooperating, thereby producing the best possible consequences given the behavior of non-cooperators. This theory takes the form of a certain procedure which the agents must go through, of which a rough sketch goes as follows:

> CC1: For each individual agent:
> (a) the agent should hold herself ready to cooperate with other cooperators,
> (b) the agent should determine which other agents:
>> (i) hold themselves ready to cooperate,
>> (ii) understand the basic situation, and
>> (iii) correctly identify other cooperators,

---

13 Note that the second example is not a problem for AC. If Amanda chooses to lift and John chooses not to lift, neither of them have satisfied AC. For Amanda, not lifting the stretcher would be the better alternative, while for John, lifting the stretcher would be better.

14 For Hooker's discussion on multiple ideal sets of rules, see his *Ideal Code, Real World* (Oxford, 2002)*, pp. 2-3, 114-17. His formulation of RC differs significantly from those considered here. According to him, those acts are wrong which are forbidden by the set of rules whose internalization by the overwhelming majority of everyone everywhere in each new generation has maximum expected value in terms of well-being (with some priority for the worst off). Also, note that Hooker's theory considers the expected rather than the actual consequences of our acts.

(c) the agent should determine which agents are *not* available to be cooperated with and how these agents are disposed to behave,

(d) the agent should determine the best possible pattern of behavior for the group of cooperators, given the behavior of non-cooperators,

(e) the agent should do her part in the best possible pattern of behavior for the class of cooperators.[15]

CC1 is according to Regan part of a possibly larger group of similar solutions. Also, CC1 is only a rough sketch of Regan's final theory, because of a number of logical difficulties. I will not go into any detailed exposition of these problems. Suffice to say that Regan introduces a more complicated procedure which while elaborate solves all or at least most of the logical difficulties.[16] For my purposes it is sufficient to consider the more intuitive procedure CC1.

Also, note that according to CC1, to be a cooperator is simply to be an agent who satisfies CC1, a usage which may or may not coincide with out pre-philosophical intuitions about what it means to cooperate. This is not of any importance here; we are merely interested in whether CC1 fulfills our three intuitions.

Finally, Regan claims that having the agents move through this procedure should not be interpreted in a temporal fashion.[17] Obviously, conceiving of going through these steps without simultaneously assuming the passage of time is somewhat difficult. It becomes slightly easier if we remember that Regan's procedure is not primarily meant as a method of decision making, but is rather a way of ensuring that the agents have done whatever is required of them. At any rate, let us grant Regan that this can be done.

How can CC1 solve our previous problems? First, it appears to fulfill our first intuition.[18] Consider, for example, a case where Joshua satisfies CC1 and Sarah does not. Since Joshua satisfies

---

15 Regan, *Utilitarianism and Co-operation,* pp. 124-45. Note that, strictly speaking, CC1 does not just give an account of what we ought to do or which acts are right, but also requires us to move through a certain procedure and to hold certain beliefs. It is not entirely correct to characterize this as a criterion of which acts are right. Nevertheless, since theories based upon cooperation have been suggested as a solution to the coordination problems discussed earlier, it will be worthwhile to see why these solutions fail.

16 The interested reader is encouraged to consult Regan's extended argument in *Utilitarianism and Co-operation,* pp. 146-63, 212-28.

17 Regan, *Utilitarianism and Cooperation*, pp. 212-14.

18 CC1 is designed by Regan to be a theory which fulfills both our first and second intuition. Regan actually tries to prove that CC1 has an even stronger property, which he calls *adaptability.* Since adaptability is stronger than, and entails, our first and second intuitions (PropAU and PropCOP, in Regan's terminology), this does not affect my discussion of Regan's view.

CC1, Joshua holds himself ready to cooperate with all other cooperators. Consequently, he determines that he himself is available to be cooperated with. He then determines that Sarah is not available to be cooperated with, and determines how she is disposed to behave. If Sarah lifts the stretcher, the best plan for Joshua is to lift the stretcher as well. If Sarah does not lift the stretcher, the best plan for Joshua is to not lift the stretcher. In both cases Joshua produces the best possible consequences he can.

CC1 also seems to fulfill our second intuition. If both Joshua and Sarah satisfy CC1, they hold themselves ready to cooperate with each other. They then correctly determine that both of them are cooperators. Upon doing so, they find no non-cooperators. They subsequently determine their best possible pattern of behavior for the group, which is to lift the stretcher. Finally, they do their part in this optimal pattern. So if both Joshua and Sarah satisfy CC1, they both lift the stretcher and produce the best possible consequences together.

Wary of the challenges facing RC, we may wonder if CC1 avoids the problem with multiple best plans. Recall the case of John and Amanda. In their case, both of them either lifting or not lifting the stretcher produces the best possible consequences. But to achieve either of these outcomes, the agents must embark on the same best plan (for the class of cooperators). As in the case of RC, it may be possible for John and Amanda to satisfy CC1 even as they pick different best plans, thereby failing to produce the best possible consequences together.

Regan solves this problem by suggesting that we, in his more detailed variant of CC1, can add to the procedure that the agents choose between the plans at random, and wait with carrying out their acts until they all have selected the same best plan.[19] This is possible because CC1 in its more complicated version involves several looping stages where the agents return to a previous step unless they fulfill certain conditions. Assuming that the number of plans are finite, Regan plausibly suggests that the agents eventually end up with the same plan. So the detailed version of CC1 can accomplish what RC cannot, because Regan demands that the agents go through a particular procedure.[20]

At this point the reader may suspect that requiring the agents to have certain beliefs, or moving through a particular procedure, is itself a considerable source of complication. The reader would be correct. In fact, three problems have been discussed in the literature, two of which are discussed by Regan himself.

First, we may agree with Earl Conee that it is highly implausible that no act is morally correct

---

19 Regan, *Utilitarianism and Co-operation*, p. 162, 261.

20 But this need not imply that the agents should use a particular method for decision making. That agents have followed the procedure is simply a way of deciding whether the agents have done what is right.

unless selected by a particular procedure.[21] In fact, Conee notes that it is in such a case likely that no act has ever been right according to CC1 (!), since it is unlikely that any agent has ever been able to ascertain the behavior of all non-cooperators. And if some agents cannot satisfy CC1, the theory does not fulfill our third intuition.

Second, the theory also violates our third intuition in a more direct way.[22] CC1 tells us that we ought to go through a certain procedure, or that we should acquire certain beliefs. But there is no guarantee that we have the ability to perform any of these tasks.

Third, there are Regan's so-called "Mad Telepath"-examples.[23] We imagine a third party, a "Mad Telepath", which in the event of us carrying through with a particular procedure or successfully acquiring certain attitudes or beliefs, will do some destructive act, such as bombing a hospital, which makes the total consequences produced worse. Also, note that theories evaluating only the consequences of individual acts have no problems with mad telepaths, since the consequences of the telepath bombing the hospital may then be included in the evaluation of the act. The possibility of these side effects constitute a specific problem for CC1, because CC1 requires something more than simply having the agents perform certain acts. In fact, this suggests a more general argument against CC1: whenever we go beyond the mere consequences of acts, we can always imagine a third party hindering the best consequences being brought about. In these cases, the best action would be to *not* follow CC1. Or, in other words, we can always construct a scenario under which doing some act other than that endorsed by CC1 will result in better consequences. Because of this, CC1 fails to fulfill, not only our third intuition, but our first and second as well.

Regan admits that this is a real problem, and agrees that CC1 is "not perfect".[24] He goes on to suggest that no theory can be perfect. Perhaps this is true for theories considering only individual agents, but as will be shown later this does not hold for those theories including collective agents as well. At any rate, CC1 is clearly defective. Is there a more plausible alternative?

Michael Zimmerman has discussed two other variants of CC.[25] Zimmerman is mainly interested

---

21 Earl Conee, "Utilitarianism and Co-operation, by Donald Regan. Review by: Earl Conee", *The Journal of Philosophy* 80 (1983), pp. 420-2.

22 Conee, "Utilitarianism and Co-operation, by Donald Regan. Review by: Earl Conee", p. 421; Regan, *Utilitarianism and Co-operation,* pp. 173-6.

23 Regan, *Utilitarianism and Co-operation,* pp. 176-81.

24 Regan, *Utilitarianism and Co-operation,* pp. 181-5. See also Sobel, "Utilitarianism and Cooperation", *Dialogue* 24 (1985), pp. 150-2.

25 Zimmerman, *The Concept of Moral Obligation* (Cambridge, 1996)*,* pp. 254-76. For an earlier and somewhat different presentation, see his "Cooperation and Doing the Best One Can", *Philosophical Studies* 65 (1992), pp. 295-301. Also, see Feldman, *Doing the Best We Can*, pp. 173-8.

in the concept of moral obligation, and his solution is broader in scope than the more narrow search for a consequentialist theory of rightness. The context of this discussion therefore differs slightly from Zimmerman's. This is not a problem, since my aim is not to argue against Zimmerman's view. I am content with showing that his view cannot aid the consequentialist who tries to appease our three intuitions.

According to a rough sketch of Zimmerman's first theory, each individual agent ought to both (a) do the best she can in that situation, and (b) enable the other members of the group, of which she is a member, to produce the best possible consequences available to the group.

It is easy to see how this theory supplies a natural way of dealing with our first and second intuitions. While condition (a) ensures that our theory fulfills the first intuition, condition (b) helps us to fulfill the second. Let us once again consider the case with Joshua and Sarah. We can show that whenever both Sarah and Joshua fulfill these two conditions, they both lift the stretcher, since for one agent to lift the stretcher is the only way for him or her to enable the other agent to lift as well. So when both Joshua and Sarah satisfy this theory, they both lift the stretcher and produce the best possible consequences they can, both as individuals and as a group.

Let us consider a more formal statement of Zimmerman's theory. To see how (b) is constructed, we first have to introduce Zimmerman's notion of *openness*.[26] Assume that S is an individual agent distinct from G, and that G is a group of individual agents, possibly with just one member. Zimmerman now defines openness as follows: S is open to G concerning p if and only if (a) *{S, G}* [the group of S and G] can act so that p is true; and (b) G can act so that p is true.[27]

To make this more clear, let us illustrate with our first stretcher case. Ask yourself what it takes for Sarah to be open to Joshua. According to the definition of openness, Sarah is open to Joshua

---

26 Openness should, according to Zimmerman, be understood as the adoption of an attitude, though as Brian Kierland notices, this "attitude" is plausibly understood not as based on the agent having certain beliefs, but rather as saying something about the state of the agent's environment. See Kierland's "Cooperation, 'Ought Morally', and Principles of Moral Harmony", *Philosophical Studies* 128 (2006), p. 390.

27 Zimmerman's account of openness is found in his *The Concept of Moral Obligation,* pp. 265-7. In the following discussion I leave out Zimmerman's references to different time frames, both in his definition of openness and in the more formal statements of his theories. His full definition of openness goes as follows: "S is open at T to G concerning p at T' if and only if (a) *{S, G}* [the group of S and G] can at T so act at T' that p is true; and (b) G can at T so act at T' that p is true". There are two reasons for excluding this extra complexity. First, while distinguishing between time frames could help the agents to coordinate their actions, this is true only if they can communicate with each other or otherwise influence the other agent, which is something the agents in our examples, by stipulation, cannot do. Second, on my account of individual agency, referring to different time frames has no obvious advantages. There can, for example, never be the case that an individual agent ought at T do something at a later or earlier time T', because in this case we are considering two different agents.

concerning lifting the stretcher if and only if (a) Sarah and Joshua can together act so that the stretcher is lifted, and (b) Joshua can act so that the stretcher is lifted. In the present case, (a) is clearly true. But under what circumstances is (b) true? Since Joshua cannot lift the stretcher on his own, the only way for him to to lift the stretcher is for Sarah to lift as well. The only way for Sarah to be open to Joshua is therefore for her to lift the stretcher.

Keeping this definition of openness in mind, consider now the first of Zimmerman's theories:

> CC2: For any individual agent S, S ought to:
>
> (a) do the best that she can; and
>
> (b) for any group of agents G and proposition p, be open to G concerning p
>
> if and only if $\{S, G\}$ so acting that p is true is the best that $\{S, G\}$ can do.[28]

CC2 seems to fulfill our first intuition. When Joshua satisfies CC2 he does the best he can on his own. And *vice versa* for Sarah. In both cases, the agent satisfying CC2 produces the best possible consequences he or she can in that situation.

Moreover, CC2 appears to fulfill our second intuition. Since the only way for Sarah and Joshua to satisfy CC2 is to be open to each other, and since the only way for each them to be open to the other is to lift the stretcher, their joint satisfaction of CC2 makes them produce the best consequences as a group.

But whatever the success regarding our first and second intuitions, CC2 fails to fulfill our third. Zimmerman discusses the following problem.[29] Consider a case where Joshua is intransigent and declines to lift the stretcher. What does CC2 tell Sarah to do? According to clause (a), Sarah should help a moderately wounded patient, since this would produce the best possible consequences in that situation. But clause (b) tells Sarah to remain open to the group of Sarah and Joshua, which she can do only if she lifts the stretcher. So regardless of whether Sarah lifts the stretcher or not, she cannot satisfy CC2.

There are other problems with CC2 as well. Zimmerman suggests that whenever the agent is part of multiple groups, it may be impossible for the agent to simultaneously be open to all groups. And there is a related difficulty concerning the possibility of there existing multiple best plans, which Zimmerman does not discuss, but which is made apparent by the case of Amanda and John. In their case, both either lifting or not lifting the stretcher produces the optimal outcome. So openness requires Amanda to both lift and not lift the stretcher, which is impossible. So even when the agent is *not* a member of multiple groups, it may be impossible for the agent to satisfy CC2.

---

28 Zimmerman, *The Concept of Moral Obligation*, pp. 265-6.

29 Zimmerman, *The Concept of Moral Obligation*, pp. 266-7.

Consider now another theory discussed by Zimmerman, which is meant to solve some of the previous difficulties:

CC3: For any individual agent *S*, *S* ought to:

(a) do the best she can; and

(b) for any group of agents *G* and proposition *p*, be open to *G* concerning *p*

if and only if

(b1) {*S, G*} so acting that *p* is true is the best that {*S, G*} can do, and

(b2) for any act *A'*, if *A'* is the best that *S* can do, then *S* can both be open to *G* concerning *p* and do *A'*.[30]

This theory is simply CC2 with the added clause (b2), which states that the agent ought only to be open to the group if and only if this is compatible with her doing the best she can. Can this theory accomplish what the others cannot?

According to CC3, Sarah no longer faces the impossible task of both lifting and not lifting the stretcher when Joshua does not lift, because openness is only required when this coincides with her doing the best she can. So in one sense, progress has been made. However, the agents may still confront the impossible task of having to be open to two different best outcomes at the same time. Zimmerman's modification of CC2 is therefore not sufficient for our purposes. Something more is needed for the consequentialist.

But even if we grant that the problem with multiple best outcomes can be solved by introducing a more complicated variant of CC3, adding (b2) wrecks havoc with our second intuition. If all agents satisfy CC2, are they guaranteed to produce the best consequences as a group? They are not. Consider a case in which both Joshua and Sarah do not lift the stretcher. In such a case, each agent picks the best alternative he or she can, which is the alternative of not lifting the stretcher. Since openness requires that the agent lifts the stretcher, and since it is not possible both to lift and not lift the stretcher, (b2) now rules out openness for both agents. In this situation, Joshua and Sarah both satisfy CC2. But their satisfaction of CC2 does not guarantee that they produce the best possible consequences as a group.

Zimmerman accepts this tragic result and believes that, in these cases, the individuals do no wrong.[31] But this is an unacceptable conclusion for those who want to maintain our second intuition.[32]

---

30 Zimmerman, *The Concept of Moral Obligation*, p. 268.

31 Zimmerman, *The Concept of Moral Obligation*, pp. 266-7.

32 Kierland suggests that while we could revert to the ordinary conception of attitudes as based on beliefs, this would

It appears then that CC1, CC2 and CC3, who first promised to free us from the problems which AC and RC had us trapped within, can fulfill our first and second intuitions only by breaking our third. What has gone wrong?

5. THE FAILURE OF INDIVIDUALIST THEORIES

The theories which we have considered so far have directed themselves only to individual agents. To gain some insight into why success for these theories proves so elusive, it will be worthwhile to return to our discussion about AC. Michael Neumann has suggested that our initial argument against AC is faulty, and that when we have worked out what it means to consult the different principles in these situations the problem disappears.[33]

Neumann's argument goes as follows. He first distinguishes between general choice situations, in which all facts about the world are not given, and particular choice situations, in which all such facts are given.[34] Particular choice situations can be generated from general ones by determining facts about the world. To illustrate with our popular stretcher case, Joshua and Sarah are initially, when the example is presented, situated in the general choice situation in which the outcomes which are possible and the value of each outcome are given; and in which the alternatives the agents will choose are not. By imposing additional constraints on this situation, by for example supposing that the agents do not in fact lift the stretcher, we transform the general situation into a particular one, and thereby place ourselves in a position to evaluate whether the agents satisfy AC. Satisfaction of theories such as AC can happen only in particular choice situations, since it is only in these situations that the acts produce determinate consequences. If we do not know which act Joshua carries out, Sarah's act produces no determinate consequences.

Neumann now suggests two different ways of understanding the phrase "the best consequences possible".[35] Either we understand it as (a) the best consequences that are *logically possible,* or as (b) the best consequences that are *possible within the confines of the given choice situation*. Consider first sense (a). While it is logically possible that Joshua and Sarah could have reached a better outcome through being in a better particular choice situation, this is not surprising. Of course,

---

just reintroduce the threat from mad telepaths. In fact, CC seems predestined to be thrown between violating the "ought-implies-can"-principle and facing third party threats. Kierland's own solution is to evaluate agents rather than acts. Such a solution, while interesting, clearly does not help our current project, which is concerned with finding some theory of rightness, not of the evaluation of character. For more on this, see Kierland, "Cooperation, 'Ought Morally', and Principles of Moral Harmony", pp. 399-401.

33 Michael Neumann, "Co-ordination and Utility", *The Philosophical Quarterly* 39 (1989), pp. 66-74.

34 Neumann, "Co-ordination and Utility", pp. 68-70.

35 Neumann, "Co-ordination and Utility", p. 70.

Neumann argues, they could have achieved better results if they were in a better choice situation.[36] This is simply to say that, had things been different, a better outcome could have been brought about. It is possible that either Joshua or Sarah had been in the different choice situation where the other agent lifted the stretcher, but it is also possible that they had been in a situation in which not lifting the stretcher produced a value of twelve rather than ten. Neumann suggests that because of this "no plausible moral theories will [fulfill our second intuition] in sense (a) – in the sense guaranteeing the best logically possible outcome".[37]

If we, on the other hand, understand "the best possible consequences" in sense (b), the problem disappears.[38] In such a particular choice situation it is stipulated that the other agent does or does not lift, because the fact of what the other agent will do is given. If Sarah is certain that Joshua will not lift the stretcher, then she should not lift either: by not lifting, she makes sure that the agents produce the best consequences together, as are possible in the confines of this particular choice situation. Because of this, the example of Sarah and Joshua fails, and theories such as AC fulfill our second intuition.

Now, I believe that Neumann contributes some important insights, but that he nevertheless misses the gist of the difficulty. It seems that we are looking not for a sense (a) or (b), but rather a third sense (c) of "best consequences possible", namely the best consequences that are possible in cases where all facts are given *except the behavior of all agents*.

It is not difficult to explain why (c) is interesting for a moral theory: we may believe that moral theory is concerned with us navigating a world populated by a multitude of agents, all of which are equipped with some weaker or stronger sense of free will, but where everything else runs according to the laws of nature. The problem is that in these situations, there remains for any agent an undetermined factor: the behavior of the other agents is not given.[39] As we have seen, since the behavior of the other agents is not given, we may sometimes fail to coordinate our actions with them in the appropriate ways, or fail to embark on the same best plan.

To ensure that the agents execute the same optimal pattern, the theories considered so far would need to direct themselves to all agents at once. But in their present form, these theories direct themselves only to one individual agent at a time. As long as there exists such unpleasant

---

36 Neumann, "Co-ordination and Utility", p. 70, 72.

37 Neumann, "Co-ordination and Utility", p. 72.

38 Neumann, "Co-ordination and Utility", p. 70.

39 We can now see another difficulty with CC1. CC1 asks of us to ascertain the behavior of non-cooperators, but this is something we must now assume is impossible. In a situation (c) we do not know the behavior of non-cooperators, anymore than we know the behavior of cooperators. It is quite strange, as Regan does, to suppose that we can know the behavior of non-cooperators, but not the behavior of cooperators.

possibilities as those facing Joshua, Sarah, Amanda and John, we should therefore expect the coordination problems to persist. The alternative is to fulfill our first and second intuitions only by violating our third.

At this point we may suspect that all theories directing themselves only to individual agents are bound to fail. How can we modify our theory so that we direct ourselves to several agents at once? One possibility, which is briefly considered by Neumann, is to introduce collective agents into morality.[40]

6. GENERALIZED ACT-CONSEQUENTIALISM

How does the introduction of collective agents solve our problem? Recall our three intuitions:

> Intuition 1: For any theory T, if an agent satisfies T, then that agent should produce by her act the best consequences she can possibly produce in that situation.

> Intuition 2: For any theory T, if all agents satisfy T, then the class of all agents should produce by their acts taken together the best consequences that they can possibly produce by any pattern of behavior.

> Intuition 3: For any theory T: all agents should be able to satisfy T.

Let us assume that, in addition to the individual agents considered so far, each collection of two or more agents constitutes an additional collective agent.[41] This ontological commitment needs to be argued for. But suppose for now, by hypothesis, that these collectives are agents who are capable of acting. How could this assumption aid the consequentialist?

Some philosophers have suggested a theory of generalized act-consequentialism (GAC):[42]

---

40 Neumann, "Co-operation and Utility", p. 74.

41 Two comments are in order. First, an alternative is to define all agents in terms of collective agents, by stating that a collective agent is constituted by one or more agents. However, I believe that ordinary language is better accommodated if we consider as a group or collective an entity constituted by two or more agents. Second, we could admit only individuals, but not collectives, as members of collectives. I can see no dangers with such a sparse ontology, at least not for the present purposes, but no apparent advantages either. I shall therefore take the more generous view, though I must stress that any of the two views will be consistent with my conclusions.

42 This theory was first suggested by Lars Bergström in "Vad är nyttomoral?" ["What is utilitarianism?"], *Filosofi och samhälle*, ed. Gunnar Andrén *et al.* (Bodafors, 1978), pp. 32-4; and B. C. Postow in "Generalized Act Utilitarianism", *Analysis* 37 (1977), pp. 49-52. It has been defended by Torbjörn Tännsjö in "The Myth of

> GAC: For any individual or collective agent: The agent ought to do the act which has at least as good consequences under the circumstances as any other act open to the agent.

According to GAC, when Sarah and Joshua fail to lift the stretcher, there is a third agent which fails to do the act which would produce the best possible consequences, namely the collective of Sarah and Joshua.

We can show that GAC fulfills all of our three intuitions. First, as in the case of AC, if an individual or collective agent satisfies GAC, the agent picks the alternative which produces the best possible consequences in that situation. GAC therefore fulfills our first intuition.

Second, if all agents satisfy GAC, this means that the collective agent constituted by all other agents produces the best possible consequences in that situation. Since this collective agent produces the best consequences possible, this means that there is no alternative pattern of behavior for all agents which would produce better consequences. All agents therefore produce the best consequences as a group. GAC therefore fulfills our second intuition as well.

Third, GAC requires nothing from each agent but that she picks the best alternative available to her. As in the case of all other theories exclusively concerned with directing each agent to do some act, GAC fulfills our third intuition as well.

Let us now turn to the question of whether collectives can act. In what sense could this be true?

The easiest way to introduce collective agents into our ontology would be on a pragmatic basis. Since we need collective agents to explain what we ought to do, we may think ourselves entitled to assume their existence on pragmatic grounds.

However, perhaps we want some additional constraints on which entities may be classified as agents. One reasonable requirement is for the agent to have been able to do otherwise. Our collective agents clearly fulfill this criteria. Had one of its individual members done otherwise, the collective would have acted otherwise as well.

Another requirement would be to demand some connection between acting and acting intentionally. But what could such a relationship look like? Requiring that all acts are intentionally carried out may seem enticing, but such a claim is too strong. For instance, when I intend to drink a cup of coffee but mistakenly drink a cup of water, the latter event is an act, even if it is not an

---

Innocence: On Collective Responsibility and Collective Punishment", *Philosophical Papers* 36 (2007), pp. 295-314; "The Morality of Collective Actions", *The Philosophical Quarterly* 39 (1989), pp. 221-8; and *Hedonistic Utilitarianism* (Edinburgh, 1998)*,* pp. 54-62. Also, see Parfit, *Reasons and Persons* (Oxford, 1984), pp. 67-86; Feldman, *Doing the Best We Can*, pp. 160-2; Jan Österberg, *Self and Others: A Study of Ethical Egoism* (Dordrecht, 1988), pp. 177-80.

intentional one.

In a now classic paper, Donald Davidson claims that these acts which are not themselves intentional are instead intentional under some other description.[43] For example, "I drink a cup of water" can be redescribed as "I intentionally drink what I believe to be coffee". According to Davidson then, someone is the agent of an act if what she does can be redescribed as intentional.[44] We could describe this as letting the non-intentional acts inherit their status as acts from the intentional ones.

Perhaps the collectivist could try a similar maneuver as well.[45] Just as events such as "I drink a cup of water" and "I intentionally drink what I believe to be coffee" are acts, the same would be true for "the collective agent lifts the stretcher". Let me suggest two ways in which this could be carried out.

A weaker view could hold that for a collective to be the agent of an act, it is sufficient that the *individuals* act intentionally. "The collective agent lifts the stretcher" would thus be an act in virtue of it being redescribable as "Sarah intentionally lifts the stretcher, Joshua intentionally lifts the stretcher". On this weaker view, while the collective can act, it need not have the capacity to intend to act in certain ways. Nor need it be able to take decisions, cherish beliefs etc. In a sense, we might say that the collective act inherits its status as an act from the intentional acts of its individual members.

Conversely, a stronger view holds that collectives as such have specifically *collective* intentions. Torbjörn Tännsjö seems to hold such a view.[46] According to Tännsjö, the collective intention is made up by a complex conjunction of all the individual intentions. As a result, the collective intention has a different content from that of any of the individual intentions. To illustrate, the collective of Sarah and Joshua can intend that "Sarah lifts the stretcher, Joshua lifts the stretcher", even though neither Sarah nor Joshua has this complex intention. Also, Tännsjö holds that the collective has certain collective beliefs and desires, and that we may think of these as "represented by a matrix where the relevant beliefs and desires of each individual making up the collectivity are represented".[47]

---

43 Davidson, *Essays on Actions and Events* (Oxford, 1980)*,* pp. 43-61.

44 Davidson, *Essays on Actions and Events,* pp. 44-7.

45 For two attempts to use Davidson's account of agency to defend a collectivist position, see Tännsjö, "The Morality of Collective Actions", pp. 226-7, and French, "The Corporation as a Moral Person", *American Philosophical Quarterly* 16 (1979), pp. 211-12*.*

46 Tännsjö, "The Morality of Collective Actions", p. 227; *Hedonistic Utilitarianism,* pp. 57-9; "The Myth of Innocence: On Collective Responsibility and Collective Punishment", pp. 300-4.

47 Tännsjö, "The Myth of Innocence: On Collective Responsibility and Collective Punishment", pp. 302-3.

There are two worries with this stronger account of collective agency. First, there may exist no straightforward way to derive the complex collective properties solely on the basis of individual ones. For example, Björn Petersson has suggested that so called *discursive dilemmas* show that "complete information about the member's individual beliefs and desires is not sufficient to explain the action of a collectivity".[48]

Consider for instance the following example of a discursive dilemma. The board of an international oil company is about take a majority vote on three separate issues, all of them concerning whether to start drilling in a foreign country. There are three board members, and three separate questions to be decided. The corporate desire, whether to start drilling or not, is itself based on the desires of the individual members. The results are as follows:

	(a) Should we set up the fields?	(b) Should we negotiate with the foreign government?	(c) Should we hire foreign employees?	(a)&(b)&(c) Should we start the drilling operation?
Mary	No	Yes	Yes	No
Carl	Yes	No	Yes	No
Elizabeth	Yes	Yes	No	No
The Board	Yes	Yes	Yes	Yes / No

Figure 5: Example of a discursive dilemma.

There are two procedures available to the Board. On a premise-based procedure, the Board's decision is the outcome of the votes on each separate question (the result is "yes"), and on a conclusion-based procedure, the Board's decision is the outcome of each individual's vote on the final decision ("no"). The problem is that, even if we know everything about the desires of the individuals members, we still cannot determine the desire of the collective. This is possible only if we know whether it has used a premise- or conclusion-based decision procedure. And similar examples could be set up to show that full information of the individual intentions and beliefs are not sufficient to explain the intentions and beliefs of the collective.

Does the existence of discursive dilemmas show that Tännsjö's account is defective? I think not. Take, for example, the premise-based procedure in the above example. It may seem as if the collective desires to start the drilling operation. But Tännsjö could reply that the collective desire is not the Board's desire to start drilling. Rather, the collective desire is the conjunction of Carl's desire to set up the fields, Mary's desire to negotiate with the foreign government, etc. And this

---

48 Björn Petersson, "Collective Omissions and Responsibility", *Philosophical Papers* 37 (2008), pp. 252-4. Discursive dilemmas are discussed by Philip Pettit in his "Responsibility Incorporated", *Ethics* 117 (2007), pp. 181-2.

conjunctive desire, held by the collective, is clearly reducible to the desires of the collective's individual members.[49]

The second problem concerns the metaphysical status of collective intentions, as well as the corresponding collective beliefs and desires. It is natural to think that human beings with their capacity for centralized decision making have the ability to intentionally carry out certain acts. But it seems preposterous to think that collections of spatiotemporally dispersed individuals, many of which will never meet or even know about each other's existence, can nevertheless intend to act as a collective.[50]

Now, Tännsjö suggests that for the collective to entertain the collective intention, we need not suppose that the collective is *conscious.* He compares the collective to a computer program, and suggests that we can attribute an intention to a computer program without at the same time supposing it is conscious.[51] But even if we accept that acting intentionally does not require consciousness, we still need to establish that a computer program can intend to act in some non-conscious way. What could give a non-conscious program such a capacity for acting intentionally? Well, one plausible suggestion is that such programs have some centralized locus of control, or are sufficiently organized in other ways. It seems plausible that it is these features which gives computer programs the ability to behave in a coordinated manner and to exhibit behavior which is consistent with acting intentionally. But the collectives we are discussing are not even remotely organized in such ways. Consider for instance a number of independently working computer programs. Even though each of these programs are plausibly described as acting intentionally, surely the collective of all computer programs is not?

But even if the stronger view is unacceptable, I suggest that Tännsjö could accept the weaker view without losing anything important. He could hold that while there exist no such things as collective intentions, the collective nevertheless acts in a weaker sense. If the stronger view is controversial, and if it contributes nothing which the weaker view has not already accomplished, I suggest we go with the weaker one. Let us then briefly consider what can be said in favor of the

---

49 If "the Board" is not a collective in the sense discussed here, this raises additional questions of the metaphysical status of this entity. One possibility is to claim that corporations, and similar entities who can have desires without these being reducible to the desires of a set of individuals, are in fact distinct non-human individuals on their own. If we consider corporations as individual agents in their own right, the collective agent under consideration here would consist of four individuals: Mary, Carl, Elizabeth and the Board (or "The Corporation"). I am not convinced that this concept of corporate agency is metaphysically sound, but at least it is compatible with the additional view of collective agency discussed here.

50 This has been argued by Conee, "Hedonistic Utilitarianism", *The Philosophical Review* 110 (2001), pp. 429-30.

51 Tännsjö, *Hedonistic Utilitarianism,* pp. 58-9.

weaker sense of collective agency.

First, as has already been noted, there are significant advantages for us to assume that collectives can act, regardless of whether they do so intentionally or not. As we have seen, none of the consequentialist theories which directs themselves exclusively to individual agents have so far fulfilled our three intuitions. So if we are looking to accept a consequentialist theory, assuming that collectives can act gives us perhaps the only way to formulate an adequate theory of rightness.

Second, there are no apparent disadvantages with the weaker collectivist position. Perhaps we could argue that a theory of rightness should concern itself only with those agents who are capable of choice and decision making. But it is difficult to explain why this should be the case. Since we have already given up the supposition that our theory should be used a method for decision making, what remains is merely the somewhat vague idea that it should guide or inform the decisions of agents. But this is likely true of only a few individual agents. For many individual agents, our theory will lend itself poorly to guide deliberation, and some are perhaps not even able to use our theory at all. And if this is the case, we should not be worried about collective agents lacking this ability as well.[52]

Third, and finally, the weaker view of collective agency provides a plausible way of understanding some cases of conjunctive acts. Examples of conjunctive acts are such as "Sarah managed to get through college", "John wrote a book", and "Amanda went and bought an ice cream". All of these acts are carried out over large swaths of time, and involve countless other acts. On the account of individual agency considered here, these acts are carried out by series of individual agents. Now, it may be true that in these cases no single individual agent intended the conjunctive act. To give an example, Amanda may first have intended to buy an ice cream, then intended to walk down the street, and finally intended to enter a store. But while no single part of her ever had the complex intention of "walking down the street, entering the store, and buying the ice cream", it still seems reasonable to say that, in some sense, this was precisely what she did. Our account of collective agency promises a straightforward way of understanding such conjunctive acts: namely as the non-intentional acts of intrapersonal collectives. Accepting that collectives can act is therefore not only a reasonable way of viewing interpersonal undertakings, but may help us to better understand intrapersonal ones as well.

7. COLLECTIVE CONSEQUENTIALISM

Assume now that the previous conclusions have been accepted: that no theory considering only

---

52 But that the agents cannot use the theory to guide deliberation does not, of course, mean that the agents cannot satisfy the theory when put forth as a theory of rightness: that is, as a criterion of which acts are right.

individual agents fulfills our three intuitions, and that a weak sense of collective agency can be introduced to solve this problem. If you are unconvinced by the arguments presented this far, I ask that you at least consider them hypothetically correct for the remaining few pages.

While showing that GAC fulfilled our second intuition, I briefly considered the collective agent which was constituted by all other agents. Let us call this agent the *universal agent*. Is should be obvious at this point that it is the universal agent which makes it possible for GAC to fulfill our second intuition. If we exclude the universal agent, new coordination problems may reappear between the other collectives.

In fact, it seems that the universal agent does its job almost too well. Upon reflection we find ourselves in a position where there is nothing left for the individuals to do, and where all of the work is carried out by the universal agent. Consider first the case where the universal agent does what is right, and where as a result the best outcome is brought about. In this case, no agent has at her disposal an alternative act which would bring about a better outcome. So when the universal agent does what is right, all other agents do so as well. And similarly, in all cases where an agent fails to produce the best possible outcome on her own, there is a corresponding failure for the universal agent; there is in such cases an action available to the universal agent which would produce better consequences.

If the individualist position had prevailed, we would have had no reason to introduce any collective agents. But why is this so? A reasonable explanation is that theories which manage to achieve equally good results with fewer moral agents are, everything else being equal, at a significant advantage. And it is, I believe, the assumption that collectives are fully reducible to their individual members which lies behind the individualists idea that we should not consider collective agents in moral theorizing. Why consider collective agents, we may ask, when all the moral wrongdoing of the collective is reducible to the wrongdoing of the individual agents? However, as the previous sections have shown, there are strong reasons for consequentialists to accept a weaker sense of collective agency. And so the reductionist argument points in the opposite direction: why consider individual agents, when all the moral wrongdoing of the individuals is reducible to the wrongdoing of the universal agent?

Consider then collective consequentialism:

> COC: For the universal agent: The agent ought to do the act which has at least as good consequences under the circumstances as any other act open to the agent.

How does COC handle the case of Sarah and Joshua? While GAC directs itself to three agents: Sarah, Joshua and the collective of Sarah and Joshua, COC directs itself only to the collective of Sarah and Joshua. According to COC, there is no fact of the matter as to whether Sarah and Joshua as individual agents have done what is right or wrong. Just as an advocate of AC can admit the existence of collective agents while holding that all the wrongdoing is attributable to individual agents (perhaps she mistakenly believes that if all individuals do what is best on their own, the best outcome is brought about), the proponent of COC can admit the existence of individual agents while holding that all the wrongdoing is attributable to the universal agent. Or, in other words, COC holds that Joshua and Sarah may do whatever they like, as long as the universal agent does what is right.

What about our three intuitions? COC fulfills our three intuitions just as well as GAC. COC directs itself to only one agent, the universal agent, and whenever this agent produces by its act the best possible consequences it can, the best consequences are brought about. And, as in the case of all other exclusively act-oriented theories, this act is available to the universal agent.

We may object that COC is a worse theory than GAC because it gives no direction to individual agents, which at first sight may seem a very strange view. GAC did not provide a method of decision making, but at least it let some lucky individuals use it for guidance. But it is difficult to explain why COC is worse off than GAC when it comes to guide individual decision making. Instead of asking what we ought to do to realize the best possible outcome, we could ask what we ought to do as parts of the universal agent. And further, if GAC is really a better theory for guiding decision making, then having individuals take their decisions according to GAC (or AC, RC, CC or any other theory) is fully compatible with accepting COC as the correct criterion of rightness.

Another objection goes as follows: if we accept COC, we are unable to blame and punish individuals for their behavior. But it seems plausible that consequentialism should be concerned merely with producing the best possible consequences, and that *holding* agents responsible is what matters, irrespective of whether they are actually responsible or not.

There are also clear benefits in accepting COC instead of GAC. COC is a less complex theory, because where GAC directs itself to both individual and collective agents, COC accomplishes the same result by directing itself to only the universal agent. And as an added bonus, COC guarantees that all moral agents are given common aims, since there is now only one such agent.

Does COC have any practical implications? Not directly, since it is not designed to guide behavior or to function as a method of decision making. However, it may still benefit the consequentialist theorist in other ways. Let me suggest at least two such advantages. First, COC may resolve some of the confusions which currently plague consequentialist views. For instance,

consider some of the different formulations of utilitarianism. As of today, there are extreme act-utilitarians who hold that individuals should directly try to calculate the consequences of their acts, two-level utilitarians who call for working on two different levels, oscillating between a practical level of simple principles and a critical level for designing them, and government house-utilitarians who are skeptical to the whole idea of adopting utilitarianism as a personal morality, suggesting it is more fitting as a tool for designing public policy.[53] Now, as practical strategies for bringing about the best possible outcome, all of these ideas may very well turn out useful. When and which mixture of theories the utilitarian must employ is itself an interesting empirical question, not to be resolved from one's armchair. But sometimes utilitarians move from such practical matters to the similar but very different question of "which acts are really right". They may then try to find some criterion of rightness, some determinate point at which they can say "when I follow this theory, I have accomplished what utilitarianism requires of me". And as a consequence, they may end up arguing which of their specific strategies also provides the real formulation of utilitarianism.

What I suggest is that on this most abstract level, there is no need for disagreement. All of these theories would do better to accept COC as a common umbrella for their views, holding that after all is said and done, our ultimate obligation is to produce the most utility together. Since this overarching theory says nothing about individual behavior, nor about methods of decision making or of which beliefs we should acquire, it makes it abundantly clear that all strategies for individual conduct are contingent on empirical facts and to be evaluated as we go, and that whether the individuals are successful depends on the effort of the group as a whole.

Second, by accepting COC consequentialists may avoid some purported counterexamples to their view. For example, utilitarians have often been accused of recommending individual agents to plug into experience machines, or to kill innocents in order to benefit the greater good. But COC says nothing about what we as individuals should do. It is therefore immune to many of these examples.

I leave it an open question whether this theory is at all plausible. Let me close by tentatively addressing another worry. As we have seen, collective consequentialism claims that the moral project cannot be realized simply by an individual doing some act, following some rules or doing

---

53 Peter Singer seems to me the best example of an extreme act-utilitarian, though he might dispute the label. For example, see his *Practical Ethics,* 2[Nd] edn, (Cambridge, 2008 [1993]). For a discussion on two-level utilitarianism, see Richard Hare's *Moral Thinking* (Oxford, 1981), pp. 1-86, and Feldman's "True and Useful: On the Structure of a Two Level Normative Theory", *Utilitas* 24 (2012), pp. 151-71. As for government house-utilitarianism, the classic is Henry Sidgwick's claim that the utilitarian's conclusions should perhaps, according to utilitarianism, be rejected by mankind in general. See his *The Methods of Ethics* (Indianapolis, 1982 [1907]), p. 490. For a more recent defense of this idea, see Robert E. Goodin, *Utilitarianism as a Public Philosophy* (Cambridge, 1995).

her part. It therefore rejects the very thought of placing individuals at the heart of the moral enterprise. And so we may object to this theory, not because of its theoretical features, but because it seems to us morally repugnant. I do not know what to say of this, except that rampant individualism appears to me just as dangerous. I can only reiterate that while on the most abstract level only those acts made by the universal agent are right or wrong, this does not imply that there is nothing for us as individuals to try to do, or theories for us to accept. Our importance as individuals is therefore, I believe, not diminished by our partaking in the universal agent; on the contrary, it is precisely as part of this collective that we are given the ability to do what is right.

REFERENCES

Barnes, Gerald. "Utilitarianisms". *Ethics* 82 (1971), pp. 56-64.

Bales, R. E. "Act-utilitarianism: Account of Right-making Characteristics or Decision-making Procedure?". *American Philosophical Quarterly* 8 (1971), pp. 257-65.

Bergström, Lars. "Vad är nyttomoral?" ["What is utilitarianism?"], *Filosofi och samhälle*, ed. Gunnar Andrén *et al.* (Bodafors, 1978), 29-42.

Conee, Earl. "Book Review: Hedonistic Utilitarianism". *The Philosophical Review* 110 (2001), pp. 428-30.

— "Utilitarianism and Co-operation, by Donald Regan. Review by: Earl Conee". *The Journal of Philosophy* 80 (1983), pp. 415-24.

Davidson, Donald. *Essays on Actions and Events* (Oxford, 1980).

Feldman, Fred. *Doing the Best We Can* (Dordrech, 1986).

— "True and Useful: On the Structure of a Two Level Normative Theory", *Utilitas* 24 (2012), pp. 151-71.

French, Peter A. "The Corporation as a Moral Person". *American Philosophical Quarterly* 16 (1979), pp. 207-15.

Gibbard, Allan F. "Rule-utilitarianism: Merely an Illusory Alternative". *Australasian Journal of Philosophy* 43 (1965), pp. 211-20.

Goodin, Robert E. *Utilitarianism as a Public Philosophy* (Cambridge, 1995).

Hare, Richard. *Moral Thinking* (Oxford, 1981).

Hooker, Brad. *Ideal Code, Real World (*Oxford, 2002).

Kierland, Brian. "Cooperation, 'Ought Morally', and Principles of Moral Harmony". *Philosophical Studies* 128 (2006), pp. 381-407.

Neumann, Michael. "Co-ordination and Utility". *The Philosophical Quarterly* 39 (1989), pp. 66-74.

Parfit, Derek. *On What Matters* (Oxford, 2011).

— *Reasons and Persons.* Oxford: Oxford University Press, 1984.

Petersson, Björn. "Collective Omissions and Responsibility". *Philosophical Papers* 37 (2008), pp. 243-61.

Pettit, Philip. "Responsibility Incorporated". *Ethics* 117 (2007), pp. 117-201.

Postow, B. C. "Generalized Act Utilitarianism". *Analysis* 37 (1977), pp. 49-52.

Rabinowicz, Wlodek. "Act-utilitarian Prisoner's Dilemmas". *Theoria* 55 (1989), pp. 1-44.

Regan, Donald. *Utilitarianism and Co-operation* (New York, 1980).

Sidgwick, Henry. *The Methods of Ethics (*Indianapolis, 1982 [1907]).

Singer, Peter. *Practical Ethics*. 2ⁿᵈ edn. (Cambridge, 2008 [1993]).

Sobel, Jordan Howard. "Rule-utilitarianism". *Australasian Journal of Philosophy* 46 (1968), pp. 146-65.

— "Utilitarianism and Cooperation". *Dialogue* 24 (1985), pp. 137-52.

Tännsjö, Torbjörn. *Hedonistic Utilitarianism* (Edinburgh, 1998).

— "The Morality of Collective Actions". *The Philosophical Quarterly* 39 (1989), pp. 221-28.

— "The Myth of Innocence: On Collective Responsibility and Collective Punishment". *Philosophical Papers* 36 (2007), pp. 295-314.

Zimmerman, Michael J. "Cooperation and Doing the Best One Can". *Philosophical Studies* 65 (1992), pp. 283-304.

— *The Concept of Moral Obligation* (Cambridge, 1996).

Österberg, Jan. *Self and Others: A Study of Ethical Egoism* (Dordrecht, 1988).