

Editorial Manager(tm) for The Review of Philosophy and Psychology  
Manuscript Draft

Manuscript Number: ROPP42

Title: Stag Hunts and Committee Work: Cooperation and the Mutualistic Paradigm

Article Type: Joint Action: What is Shared?

Keywords: altruism; cooperation; ethics; shared intention

Corresponding Author: Dr. Jay Elliott,

Corresponding Author's Institution: Yale University

First Author: Jay Elliott

Order of Authors: Jay Elliott

Abstract: Contemporary philosophers and psychologists seek the roots of ethically sound behavior - including altruism and a sense of fairness - in the basic structure of cooperative action. I argue that recent work on cooperation in both disciplines has been hampered by what I call "the mutualistic paradigm." The mutualistic paradigm treats one kind of cooperative situation - what I call a "mutualistic situation" - as paradigmatic of cooperation in general. In mutualistic situations, such as the primeval stag hunt described by psychologist Brian Skyrms, every partner in a cooperative action has to do his part in order for the action as a whole to succeed. But many familiar cooperative situations - for example, serving on an academic committee - do not have this structure. Thinking about cooperation can shed light on how and why ethically sound behavior happens in human beings. But the deep connections between ethics and cooperation only come into view once we have a richer conception of our capacities for cooperation than the mutualistic paradigm provides.

Suggested Reviewers: Michael Tomasello

tomas@eva.mpg.de

Suggested reviewer's work is discussed in submission.

Brian Skyrms

bskyrms@uci.edu

Suggested reviewer's work is discussed in submission.

Michael Bratman

bratman@stanford.edu

Suggested reviewer's work is discussed in submission.

Joan Silk

jsilk@anthro.ucla.edu

Suggested reviewer's work is discussed in submission.

Stag Hunts and Committee Work:  
Cooperation and the Mutualistic Paradigm

Jay R. Elliott

Whitney Humanities Center  
Yale University  
53 Wall Street  
New Haven, CT 06511  
U.S.A.

Phone: 203 988-5761  
Fax: 203 432-1087  
[jay.elliott@yale.edu](mailto:jay.elliott@yale.edu)

Acknowledgment: For advice and encouragement in connection with this essay, I am indebted to Anton Ford, Nat Hansen, Rafeeq Hassan, Erica Holberg, Ben Laurence, and Candace Vogler.

## Stag Hunts and Committee Work: Cooperation and the Mutualistic Paradigm

**ABSTRACT:** Contemporary philosophers and psychologists seek the roots of ethically sound behavior – including altruism and a sense of fairness – in the basic structure of cooperative action. I argue that recent work on cooperation in both disciplines has been hampered by what I call “the mutualistic paradigm.” The mutualistic paradigm treats one kind of cooperative situation – what I call a “mutualistic situation” – as paradigmatic of cooperation in general. In mutualistic situations, such as the primeval stag hunt described by psychologist Brian Skyrms, every partner in a cooperative action has to do his part in order for the action as a whole to succeed. But many familiar cooperative situations – for example, serving on an academic committee – do not have this structure. Thinking about cooperation can shed light on how and why ethically sound behavior happens in human beings. But the deep connections between ethics and cooperation only come into view once we have a richer conception of our capacities for cooperation than the mutualistic paradigm provides.

**KEYWORDS:** altruism; cooperation; ethics; shared intention

### Introduction

Contemporary philosophers and psychologists are interested in using the idea of cooperation as a starting point for understanding ethics. Recently, influential thinkers in both of these disciplines have been particularly attracted to the idea that our ability to cooperate can show us something about the nature of human practical rationality. The hope is that, equipped with an account of how cooperation is rational, we can get some traction on the vexed question of how ethics is rational.

I think there is a lot of promise in this line of inquiry. But I will argue that these discussions, in both philosophy and psychology, have been hampered by an inadequate conception of cooperation. That inadequate conception has in turn kept from view the deep ethical insight that thinking about cooperation *can* give us.

My argument has four main parts. In the first part, I describe a dominant mode in recent thinking about cooperation, a mode that I call the “mutualistic paradigm.” The mutualistic paradigm has been fundamental to recent work on cooperation in both social

psychology and philosophy. But in each discipline the paradigm is driven by different motivations, and so I discuss them in detail separately, taking psychology first and then philosophy. I then introduce, in the second part, an aspect of our capacity to cooperate that the mutualistic paradigm cannot account for, namely our capacity to cooperate in the face of uncooperative partners. That criticism points us, in the third part, to a deep feature of human cooperation, a capacity for what I call “participant rationality.” In the final section, I begin to draw out the significance of the idea of participant rationality for understanding the kind of rationality at work in ethics.

### **The Mutualistic Paradigm In Psychology**

At the heart of contemporary behavioral psychology is a kind of “how possible?” question about ethics. We see, or at least we hope, that it is possible for people to be kind, generous, fair, and so on. But it is a bit of a puzzle how anyone ever actually grows up to have these admirable traits. For reflection on human evolution might lead you to believe that we’ve got to be pretty selfish organisms in order to survive and pass on our genes. And reflection on human development might lead you to believe that we’ve got to be pretty selfish creatures in order to survive childhood and make it to maturity.

Contemporary psychologists try to make a start on answering this “how possible?” question by thinking about certain basic capacities that might underlie the qualities we admire. That is, psychologists suggest that we should not take a sophisticated achievement like full-blown adult generosity as our starting point. Rather, we should look for simpler capacities that in some way resemble the full-blown virtue, and then try to tell a story about how we build from the simpler capacity to the virtue.

So in the case of generosity, we could begin with something like a tendency to *help others*, which you might think includes many of the essential materials for generosity, for example the ability to understand that others have needs. And in fact recent studies have shown that even quite young children display a propensity to help others. When an adult seems to inadvertently drop something, such as a clothespin, children pick it up and hand it to them, without any direction from the adult.<sup>1</sup> The ambition behind these studies is that this seemingly innate behavior may be at the root of a natural human tendency toward altruism.

Though a lot of recent psychological work has been focused on altruism, more recently psychologists have become interested in the distinction between *altruism* and *cooperation*. Cooperation seems particularly important for understanding how we develop a sense of reciprocity and fairness, and thus how we acquire sophisticated notions such as the notion of a right or an obligation. As with altruism, psychologists have looked for simple capacities to cooperate that might underlie and help explain our sense of fairness. And here, too, studies have shown that from an early age children exhibit a striking capacity to act cooperatively.

In one study, three-year old children worked together to lift a pole up a “step-like apparatus”, one child on each end of the pole. Once the children get the pole all the way to the top, each child receives a reward.<sup>2</sup> The task is not difficult, but it requires the participation of both partners. Three-year olds do show a striking ability to cooperate in this way, and this ability shows that they can pursue goals that they can only achieve given certain actions by others. But the results of this study get even more interesting.

---

<sup>1</sup> See two papers by Felix Warneken and Michael Tomasello: “Altruistic Helping in Young Infants and Chimpanzees”; and “Helping and Cooperation at 14 months of age”.

<sup>2</sup> For a description of this (as yet unpublished) research, see Michael Tomasello, *Why We Cooperate* 65-6.

When the experimental design was altered so that one child could get his reward first, before the task was finished, this “fortunate” child most often continued with the activity until the task was complete and the other child received his reward as well. This second phase of the experiment suggests that these young children are not only able to frame goals that they can only achieve given some action by others. They are able to frame a *shared goal* – in this case, that the task should be completed and both should get their reward. This study shows that children have an ability to engage in cooperative activities that is not strictly dependent on their aiming at rewards for themselves. Children do not lose all interest in the task of lifting the pole once they have their own reward, but persist until the task is complete and their partner is rewarded as well. Furthermore, this effect seems to be distinct from altruism, since children were more willing to complete the joint task than they were to simply help another child get a reward in a similar context where there was no collaboration.

Another study shows just how attached to cooperative activities children can be. In this study, led by Felix Warneken at the Max Planck Institute, an adult and a child engaged in a cooperative activity. In one case, the activity had a goal – to find a bell hidden inside a tube. In the other case, the activity was simply an amusing game that involved dropping a small cube through a pair of inclined tubes. Once the activity was under way, the adult experimenter would withdraw and refuse to go on.<sup>3</sup> Pre-linguistic children responded by making movements and gestures designed to get the adult to reengage in the activity.

This study supports the idea that children thought of what they were doing as something they were doing *with* the adult. The adult did not figure, for them, merely as a

---

<sup>3</sup> F. Warneken, F. Chen, and M. Tomasello, “Cooperative Activities in Young Children and Chimpanzees”

kind of necessary background condition to their being able to engage in the activity. If that were the case, the children might have simply dropped the activity once they saw that the adult was no longer participating. Warneken and his co-authors tell us that this is in fact what chimpanzees do in similar situations. Human children, by contrast, seem to think of the adult partner as taking part in the activity in something like the same way they themselves do. When the adult seemed unwilling, the children retained their interest in the activity, and urged the adult to keep participating.

Studies like these have led social psychologists such as Michael Tomasello to offer the radical suggestion that a capacity to cooperate may be intrinsic to human practical rationality. As Tomasello puts it:

From a young age, children...possess a kind of social rationality... in shared cooperative activities, my individual rationality – I want to transport the table to the bedroom so I should do X – is transformed into a social rationality of interdependence: *we* want to transport the table to the bedroom, so I should do X and you should do Y.<sup>4</sup>

Part of the appeal of thinking about cooperative capacities as building material for ethics is that cooperation seems more tractable than altruism from an evolutionary point of view. Evolutionary psychologist Brian Skyrms has proposed that the origin of many distinctive human capacities – including capacities for communication and trust – may lie in an ancestral “stag hunt.”<sup>5</sup> Suppose our ancestors were strong and swift and clever enough that each of them could catch a rabbit, but not so strong, swift, or clever that any of them could catch a stag. Two or more of them working together, however, might be able to catch a stag. And even though the proceeds of a stag hunt are shared, each hunter gets more meat from his part of the stag than he would from a whole rabbit. If only our ancestors could work together to bring down the stag, they would each be better off.

---

<sup>4</sup> Michael Tomasello, *Why We Cooperate* (Cambridge, Mass: MIT Press, 2009) 40-41

<sup>5</sup> Brian Skyrms, *The Stag Hunt and the Evolution of Social Structure*

Under these conditions, being a cooperative sort of creature is evolutionarily advantageous, since those who hunt stags together will be better off than those who only hunt rabbits separately. Thus it seems plausible that the cooperative tendencies we find ourselves with today were nurtured and reinforced by this kind of primeval stag hunt.

Skyrms' stag hunt is a prime example of what I call the "mutualistic paradigm" in thinking about cooperation. The hallmark of the mutualistic paradigm is that each agent's participation in a cooperative activity is necessary in order for the activity to be performed. Thus each agent's contribution is necessary in order for the end – in Skyrms' case, a stag feast – to be achieved. In situations that fit this paradigm – what I will call "mutualistic situations" – each agent can reason from the end he aims at, namely a stag feast, to something *he* needs to do. And each agent will find that in order for him to enjoy the feast, he needs to contribute to the hunt.

The practical reasoning of a stag hunter is social in the sense that he needs to coordinate his efforts with others. If a cooperative effort is necessary in order for any of us to bring down the stag, a hunter won't do himself any good if he insists on showboating and trying to bring down the stag himself. He needs to know when to hold back and give others a shot, or when to take advantage of the distraction offered by another hunter in order to strike. In general we can say that a stag hunter needs to reason socially in the sense that he needs to be able to frame and carry through projects which he cannot achieve by himself, but which he can achieve given certain intentional actions on the part of others. The mutualistic paradigm does capture these aspects of social rationality.



Nonetheless the bulk of my argument here is devoted to arguing that the mutualistic paradigm can't capture the full depth of our capacity to cooperate. That failure means that we still have work to do in understanding how we could come to have the cooperative capacities we do. For I am going to argue that we have capacities to cooperate that could not derive from mutualistic situations like that of the stag hunt.

The failure of the mutualistic paradigm also means that psychologists in this tradition misconstrue the significance that our capacity to cooperate has for ethics. Psychologists such as Tomasello argue that we can find “the seeds” of “normative judgments of rights and responsibilities” in our capacity to engage in activities where “participants [were] mutually aware that they were dependent on one another for success.”<sup>6</sup> Tomasello’s thought is that, for example, we can move from the idea that each agent’s contribution is necessary for the success of the group’s enterprise to the idea that if one of us should fail, he has thereby “wronged” the group or its members. This line of thinking raises several problems. One is the problem of how to understand obligations to those outside of our group. It will be hard to see how a sense of fairness that is keyed to such small-scale collaborative activities could “scale up” to anything like a doctrine of human rights, or even the rights of citizens in a modern nation. Though this “outsider” problem is important, it is not, I think, the deepest problem we face in trying to get a grip on ethical life by reflecting on cooperative situations. The deeper problem, as I have already suggested, is that many cooperative situations simply don’t fit the mutualistic paradigm. In many cooperative situations, there is simply a gap between the thought of what *we* have to do to attain some shared end, and the thought of what *I* have to do in

---

<sup>6</sup> *Why We Cooperate* 98-99. The idea that rights and obligations are generated by mutualistic cooperative activities has been defended by Margaret Gilbert. See, e.g., “What Is It for *Us* to Intend?”

order for us to attain that end. This gap, in turn, makes it hard to find a source of respect for others, or for communities, in the exigencies of individual practical reasoning.

I'm going to come back to the ethical issues toward the end. At that point I will argue that thinking about human practical rationality as containing an intrinsically cooperative dimension *can* help us to understand ethics. But in order to bring out the right connection between cooperation and ethics, we need the right conception of cooperation. I think the mutualistic paradigm can't be the right view of cooperation in general.

The mutualistic paradigm is not only attractive to psychologists for reasons connected with their interest in explaining human evolution and development. The work I have been discussing in psychology is indebted to work in the philosophy of action that has its own distinct set of motivations for hewing to the mutualistic paradigm. It is to these motivations that I now turn.

## **The Mutualistic Paradigm in Philosophy**

In this section I want to explain how the mutualistic paradigm works in recent philosophy of action and to highlight some of its philosophical motivations. Since my aim is ultimately to cast doubt on the mutualistic paradigm, I begin by drawing out the philosophical assumptions that can easily make it seem as if mutualism is the only way to make sense of cooperation.

Many philosophers of action, including Michael Bratman, David Velleman, Margaret Gilbert, and Raimo Tuomela, have developed mutualistic approaches to cooperation. But here I am going to focus on the version developed by Bratman, since it is, I think, the most sophisticated and the most sensitive to questions of underlying

philosophical motivation. Now in Bratman's work cooperation is thought of as a *species* of a broader genus, which Bratman calls *shared intentional action*. In Bratman's sense, cooperation involves not merely acting together for a shared goal, but also things like mutual aid and an absence of coercion between the parties. But for my purposes, Bratman's distinction between cooperation and shared intentional action is unimportant. Bratman's commitment to the mutualistic paradigm comes out in his account of shared intentional action in general, and so we can confine our attention there.

Bratman's discussion of shared intentional action begins with a concern he shares with other contemporary philosophers of action. The concern is that we should not think of the actions of groups as the doings of a mysterious "superagent" (my term). Thus Bratman writes:

Shared intentions are the intentions of the group. But...what they consist in is a public, interlocking web of the intentions of the individuals.<sup>7</sup>

Bratman's aim is to make sense of talk about what *we* do, or what *we* intend, that is consistent with the plausible assumption that, ultimately, there is nothing more to what *we* do than what you and I do. With this assumption in mind, Bratman pursues the following strategy for understanding what it is for us to intend to A: begin with a case in which I intend to A, and you intend to A; then think about what further we have to *add* to our story in order for us to have a *shared* intention to A.

From Bratman's point of view, this procedure is analogous to what he takes to be the standard procedure for thinking about intentional action in general. To understand what it is for me to do A intentionally, the standard story goes, we begin with a picture of what it is for me to do A that is neutral between my doing A intentionally and not.

---

<sup>7</sup> "I Intend that We J" 143

So, for example, if we wanted to understand what it is for me to *scald* you intentionally, we begin with a description of the action that is neutral with respect to whether it was intentional. “Scalding” is plausibly such a description, since one can, after all, scald someone without intending to, as when a nurse bathing a patient reaches for the hot tap, mistaking it for the cold one. Then, the standard story continues, we ask what we have to add to our inchoate story of a scalding in order for me to scald you intentionally. Here the natural candidate is going to be something about how the scalding was *caused*: perhaps it was caused by a special mental event of intending, or perhaps by a suitable combination of belief and desire. Bratman proposes that we can follow a similar procedure when it comes to shared intention.

We begin with a description of what we’re up to that is neutral on the question of shared intention. Say its “I’m carrying a table and you’re carrying a table.” Note that here the relevant description needn’t be neutral with respect to whether the action on either of our parts is *intentional*. “Carrying a table” is not, of course, plausibly neutral on this question. The description need only be neutral with respect to whether we *share* an intention. And, again, we’re interested in shared intention here as being the essential ingredient in cooperation.

Now Bratman proposes that in following this procedure we start from a place that may seem rather surprising. He suggests that we begin with a description of each agent as intending what he calls a “joint activity”. For example, if I intend to dance a tango with you, I intend a joint activity, since, after all, it takes two. This move on Bratman’s part may seem surprising, since intending a joint action might seem to already build in the idea of a shared intention. But Bratman denies that this is so: a shared intention is one

that I can have only provided others have it as well. But I can intend a joint activity, say, of tango dancing, even if, alas, no one else is interested. So Bratman has refined our initial question: what is the difference between our intending to A separately and our intending to A together? Bratman's refined version of the question is: what is the difference between each of us intending a joint activity of Aing and our sharing an intention to A?

Bratman's proposal that we start with each of us intending a joint activity is appealing. If we don't start with each of us having a joint activity in mind, it doesn't look like we could ever build up to a shared intention, since of course a shared intention must be an intention to do something together. But the idea of an individual agent intending a joint activity may itself seem puzzling.

Here we encounter what we might call the "control" problem, which has been pressed against Bratman's account by David Velleman among others.<sup>8</sup> How, the problem goes, can I intend that *we* do something, when whether *we* do something is not under my control? It seems impossible, given the plausible assumption that I can only intend what is under my control. That assumption looks indispensable, given that it plays an essential part in distinguishing bona fide intentions from things like wishes and fantasies, which aren't constrained by what is under my control.

If you were somehow under my authority, then I could form intentions about what you'll do, since in that case I do have control. Thus a gangster, for example, can form intentions about whom his underlings will kill. Similarly, a dance teacher might be able to intend that a student tango with her, because she has a certain authority over him as the

---

<sup>8</sup> See Velleman's "How to Share an Intention" and Bratman's response in "I Intend that We *J*"

teacher. But in cases where neither of us is under the authority or influence of the other, it isn't clear how I can form an intention about what we'll do.

Bratman has an ingenious solution to this problem. He proposes that I can form an intention about what we'll do on the basis of a *prediction* about how you'll respond to my intention. Thus I might know that you'll dance with me if I ask you. In that case, I can intend that we dance together. Whether we dance together *is* in a sense under my control, not because you are under my authority or influence, but rather because I know that my willingness to dance can move you to be willing, too.

At this point we are in a position to see how Bratman solves our initial problem in its refined form. Recall that the refined version of the problem was: what is the difference between each of us intending a joint activity of Aing and our sharing an intention to A? The difference, on Bratman's account, is shown by the fact that I can intend for us to dance together before my intention is known to you. Before my intention is known to you, I can intend a joint activity with you, say, to dance a tango with you, but I can't share an intention with you to dance a tango. Once my intention is known to you, and you, as I predicted, concur, *then* we have a shared intention. When you concur, you form an intention of your own that we dance together. Without that, of course, my intention to dance with you comes to naught. Before you know my intention, I can intend a joint activity with you on account of my being able to accurately predict that *if* my intention is known to you, you'll form a corresponding intention.

That you and I share an intention, for Bratman, consists essentially in 1) that you and I each intend a joint activity involving the other; and 2) that each of us is moved to

act by our knowledge of one another's intentions. Here is Bratman's full account of the conditions that are jointly necessary and sufficient for shared intention:

- (1a) I intend that we dance
- (1b) You intend that we dance
- (2) My intention is known to you, and yours is known to me
- (3a) The persistence of (1a) depends on my continued knowledge of (1b): if I did not know that (1b) I would not intend that we dance.
- (3b) The persistence of (1b) depends on your continued knowledge of (1a): if you did not know that (1a) you would not intend that we dance.
- (4) We will dance if but only if (1a) and (1b).
- (5) (1) – (4) are common knowledge between us.<sup>9</sup>

In this situation, as Bratman says, each of us has control over whether we dance, but a control that only works by engaging the will of the other.

Bratman's picture can seem a little complicated, but its appeal is simple. It allows us to explain how we can share intentions in terms of more basic capacities, in particular a capacity to intend our own actions, combined with a capacity to predict the responses of others.

The mutualistic paradigm plays an essential role in Bratman's account. Cooperative actions that fit Bratman's account must be ones in which, as in the tango, each partner's contribution is necessary in order for the action to come off. That is why it makes sense to assume, as Bratman's account does, that each partner is willing to participate in the action only if the others are.

### **When the Feeling Isn't Mutual**

I have just been describing how the mutualistic paradigm informs current thinking about cooperation in both psychology and philosophy. In this part I am going to make trouble

---

<sup>9</sup> Based on "I intend that We *J*" 153, but modified to fit the example I have been using.

for this paradigm. But I don't think the trouble I am going to make for the mutualistic paradigm casts doubt on the radical suggestion that human rationality is intrinsically social. In fact, I think philosophers and psychologists beholden to the mutualistic paradigm have underestimated how radically social human practical rationality really is.

The mutualistic paradigm can be boiled down to two basic conditions. First, that our intentions are interdependent, so that I intend to perform a certain sort of action if you will, and vice versa. And second, that this interdependence of our intentions is common knowledge between us. One criticism of the mutualistic paradigm might point out that the second condition, the common knowledge condition, is implausible. It seems clear that there can be large-scale cases in which people cooperate who are not even acquainted with each other, as when the citizens of a large city collectively engage in recycling by putting our cans and bottles out on the curb. In cases such as this, I don't even know everyone I am cooperating with, so it is surely not true that I know their intentions. This criticism is right as far as it goes, but I think there is a deeper problem with the mutualistic paradigm.

The deeper problem is that it doesn't get right the kind of dependence on others that cooperation involves. The mutualistic paradigm thus ends up mischaracterizing the structure of intention at work in cooperative action. According to the mutualistic paradigm, if you and I are cooperating in doing A, then my willingness to do A depends on your willingness to do A, and vice versa. But we very often cooperate while knowing full well that some of our partners aren't willing to do their parts.



Consider every academic's favorite example: committee work.<sup>10</sup> Suppose our department wants to hire a philosopher of science, and a committee is formed to review the applications and make a recommendation to the department about whom to hire. The committee undertakes a *single* process of review, and makes, at the end of it, a *single* recommendation. The point of having a committee will be lost if each member of the committee undertakes his own review and makes his own recommendation. So here we have a classic case of cooperation: each member of the committee, insofar as he participates, will aim to avoid duplicating the labor of the others, and to arrive at a recommendation that the group can agree on.

Actions such as *reviewing the applications* and *making a recommendation* can be ascribed to the committee, and the members of the committee likewise can be said to do these things, where it is understood that the members do them not separately but collectively. This way of talking is perfectly familiar, but it has the following striking feature: all these things can be true of the committee even if some members of the committee contribute little or nothing to the effort.

Fortunately, most committees are so constituted that a full contribution by every member is not required in order for the committee to accomplish its work. Each member goes into the committee knowing that the others may slack off without substantially impairing the committee's chances of success. And by the same token, each member goes into the committee knowing that *he* could slack off without substantially impairing the committee's chances of success.

So insofar as the committee members are acting for the end of making a good recommendation, each of them has to recognize that their own contributions are very

---

<sup>10</sup> My use of this example is indebted to Joan Silk's reply to Tomasello in *Why We Cooperate* 111-122.

often not necessary in order for this end to be achieved. Of course they may have other reasons to do their part: perhaps their reputation is at stake, or perhaps they are keenly interested in the outcome of this particular search. But those who *do* participate for the sake of making a good recommendation do so knowing that their participation is not actually necessary in order to bring that off.

Now if all this is right, then there is something missing in the mutualistic paradigm. For if what we have just been saying is right, then it can't be right to say of cooperative situations in general that each party to the action is only willing to engage in the action provided the others are. We've just seen that in cases like that of a hiring committee, cooperators have to be willing to cooperate despite the fact that some of their partners may not be.

In both the philosophical and the psychological literature there is a tendency to focus on two-person cases as the basic cases of cooperation. And this focus on two-person cases may partly account for the limitations of the mutualistic paradigm. It might be true that in two-person cases, such as dancing a tango, one partner's unwillingness is often enough to doom the whole enterprise. But many cases are not like this, and any account of the phenomena of human cooperation has to allow for situations in which the collective work goes ahead despite the fact that many of us simply couldn't be bothered or never got around to it.

At this point, I want to consider a couple of objections to the line of argument I have just been setting out. First, the very idea of uncooperative partners can sound strange. I'm saying that they are part of the group that performs the action, for example,

the committee, but that they don't contribute anything to what the committee does. So in what sense are they part of the committee?

Well, in cases like that of the committee we of course have an independent criterion for membership in the group: I'm a member of the committee, say, because I have agreed to be on it, or because I have been assigned to it by someone higher up. That doesn't say anything about what contribution I will or won't make. Some cases may not be so clear-cut. But the fundamental point here is that uncooperative partners *must* be part of the group, because otherwise we could not speak of them as *not* doing their part. There must be a difference, in other words, between an *uncooperative partner* who fails to do his part, and a mere bystander. A mere bystander might be helped or harmed in some way by the group's action. He may even do something to help or hinder the group. But he doesn't have any part to play in it, and so isn't a member of the group.

A second objection runs like this: surely what the group does is a function of what its members do. If we deny that, then we risk falling into the idea of a mysterious "superagent" which, we saw above, philosophers invoke the mutualistic paradigm in order to avoid. The argument I have been making can seem to indulge in the mystery of the "superagent", since it allows that I can be a member of a group that does something, even if I contribute nothing. But I don't think anything I have said implies the need for a mysterious "superagent". It is surely true that a committee's doing something, for instance making a recommendation, depends on what its members do. If *no one* gets around to reading the applications, for example, the committee surely didn't read the applications either. It's just that the ascription of the action to the group does not depend on what *each* of its members does. It remains true that the committee does and intends

nothing except by way of what (at least some of) its members do and intend. Our capacity for cooperation isn't independent of individual agents' capacity to do and intend things. It is just that among the powers of individual agents are powers to do and intend cooperative actions.

Mutualism is attractive because it makes cooperation tractable from a certain theoretical point of view. That comes out in its appeal to both philosophers and psychologists. One response to the kinds of arguments I have been making is to say that we should think of mutualistic situations as basic and try to build up from there to cases that allow for uncooperative partners. For example, we might say that mutualistic situations foster cooperative "skills" and "habits" that, once developed there, can be applied to more complex cases. We don't, as far as I know, have a real sketch of how this would go. Clearly the central problem would be to give an account of how skills and habits that are keyed to mutualistic situations could apply outside of those contexts. It doesn't do anything to help with this problem simply to give those skills and habits the generic label "cooperative."

But perhaps this strategy could after all be made to work. What we have gotten clearer on, though, is just how much this kind of strategy would have to explain. The same basic point would go for Bratman's suggestion that his approach might be seen as explaining one *kind* of shared intention.<sup>11</sup> It is hard to know how to take such a suggestion in the absence of an independent account of what the various "kinds" of shared intention are and how they are related. We don't, as far as I can see, operate with any intuitive distinction among kinds of shared intention. But even if Bratman's case could be made, my argument would still show how few cases his model covers, and

---

<sup>11</sup> See "I Intend that We *J*" 144, 160

surely the most ethically and philosophically interesting cases of cooperation are those that allow for uncooperative partners.

## **Cooperation without Mutualism**

Earlier I introduced what I called the “radical suggestion” that human practical rationality involves an intrinsic social dimension. As I mentioned before, I think the mutualistic paradigm considerably underestimates just how social human practical rationality can be. The foregoing considerations about uncooperative partners help to bring into focus the deep grounding for the radical suggestion.

There may be some ways in which mutualistic situations require social rationality. For example, they may require an ability to shape one’s action so as to take account of others’ efforts. In a stag hunt, for example, I won’t help to bring down the stag if, though I stab the stag with my spear, I do so clumsily in a way that interferes with the spearthrows of others, allowing the stag to get away. But cooperation as we know it involves a more radically social point of view than this. In a stag hunt, I can still think of what I do as what *I* need to do in order for *me* to enjoy any stag products. It’s just that it turns out that what *I* need to do in order to enjoy stag products is to coordinate my actions in certain ways with those of others. In that sense, I can fully participate in a stag hunt while still acting from an individual point of view. But in committee work, to return to our earlier example, I won’t be much of a cooperator if I am concerned solely with what I have to do in order for me to enjoy, say, a good philosopher of science as a colleague.

Many familiar cooperative activities, such as committee work, are characterized by what I call *the paradox of participation*: someone has to participate in order for the whole enterprise to succeed, but not anyone in particular. In order to go on participating

in the face of the paradox, I have to be willing to participate, knowing that my own contribution is neither necessary nor sufficient in order for the end to be attained. That knowledge is hard to hold together with the fact that it is only my action's contribution to the attainment of the end that gives it its point. In order to participate in cooperative activities with *this* structure, we have to be able to take up what I call a "participant" point of view. From this point of view, I act as *some of us* need to act in order for *us* to attain our end. Acting in this way follows a distinctive pattern of *participant rationality*, in which I move from thoughts about what some of us need to do to conclusions about what I'll do, given that I'm one of us.

Now this may seem fairly far from the picture of social rationality that contemporary psychologists are interested in, as exemplified by the stag hunt and other mutualistic situations. But I think we can see some evidence in the psychological studies I mentioned earlier of an ability to carry on cooperating in the face of others' uncooperativeness. Think back to the study of interruption and reengagement in cooperative activities. Warneken and his coauthors took their results to show that young children have the ability to form shared intentions. They argued that the child's urging the adult to reengage in the activity meant that the child was engaged in a shared activity with the adult, to which the child now wanted the adult to recommit.

That is right as far as it goes. But the children's urging the partner to recommit suggests something further: that the children's willingness to engage in the activity did not depend on their partner's being willing. In the face of an uncooperative partner, the children took steps to get that partner to recommit to their activity. That implies that the children's willingness to engage in the activity held up in the face of an uncooperative

partner. The children seemed to think in terms of *a goal* and *an activity* that the two of them shared, *even when the partner was unwilling*.

Now it is important to note that the situations in Warneken's study are still mutualistic in the sense that if anyone is going to attain the end he aims at, then every partner has to participate. It would be interesting to see experimental protocols that were designed explicitly to test children's ability to participate in cooperative activities that can succeed without everyone's participation. It seems plausible to think that young children should be able to do this, but this is clearly an area where further work remains to be done. It would be unwise, in any case, to conclude that young children do have this ability merely by some kind of "extension" from their ability to participate in mutualistic situations. For if what I have been arguing is right, then our ability to engage in activities that allow for uncooperative partners involves a fundamentally different kind of social ability, the ability to take up what I have called a "participant" point of view.

I have been arguing that our capacity to cooperate includes an ability to sustain an attachment to a cooperative project where some of one's partners are unwilling to do their parts. I want to stress that even in this kind of cooperation, we have something importantly different from altruism. In this way, my critique of the mutualistic paradigm differs from one that has been raised by primatologist Joan Silk. Silk rightly points out that mutualism can't provide a general account of cooperation, since it doesn't allow for situations in which we cooperate despite uncooperative partners. Mutualism makes sense for cases like the imagined primeval stag hunt, in which, as Silk puts it, the preferences of the parties are "perfectly aligned."<sup>12</sup> But in humdrum cases like that of committee work, some of us are more willing to contribute than others, and this may be true despite the

---

<sup>12</sup> See Silk's reply in *Why We Cooperate* 113.

fact that each of us can see that he stands to gain from a successful committee. The reason is that the success of our committee doesn't stand or fall on any one member's contribution, and all of us know this. Mutualistic situations are theoretically appealing because in these situations what each individual gets out is a function of what he puts in. Mutualistic situations thus seem to give us a kind of leverage from the theoretically easy case of separate actions to the theoretically hard case of cooperative actions. The problem is that there seems to be a lot of human cooperation that the mutualistic paradigm doesn't model very well. Someone who was only willing to cooperate insofar as his contribution was *necessary* wouldn't make a very good committee member. He'll do only what *he* has to in order for the whole committee to succeed, which, as it turns out, may be little or nothing. Committees as we know them could hardly function if everyone thought and acted that way.

So how, Silk asks, do committees ever succeed? Her answer is that "we have altruistic social preferences that motivate us to *value* the benefits to the group."<sup>13</sup> Her suggestion is that we should think of dutiful committee members as concerned with the good of the group, as opposed to their own. Consider, she suggests, another situation familiar to academic types: the public radio pledge drive. My measly contribution surely does not make a difference with respect to whether I or anyone else can enjoy public radio. Yet people keep giving, enough to keep public radio going, at any rate. Why? It might be those irresistible "premiums": the tote bags, the clever mugs, and so on. But Silk concludes that it is more likely that most people give because they want to contribute to a public good. They want to do something good for their community. Thus Silk argues that the failure of the mutualistic paradigm lands us back in the altruism problem.

---

<sup>13</sup> Silk's reply to Tomasello in *Why We Cooperate* 120



Now my proposal differs from this. I think that despite the failure of mutualism there may still be something important and distinctive about cooperative situations. Think again about Silk's public radio example. Of course some people give to their station as a philanthropic act. They want to provide something good for others. People who give very large amounts naturally think in this way, as if they are giving in place of those who can't. These people fit well with Silk's altruism model.

But when the folks on public radio urge me to contribute by reminding me of how much I get out of the programming, they aren't just talking nonsense. They certainly aren't deceiving me into thinking that I will enjoy the programming any less if I don't donate. It may rather be that they are pointing to the fact that *the listeners* support the station, and so as *a listener* I have a reason to give. This reason has the character of what I earlier called "participant rationality." Thinking as a participant, I give to my station in proportion to what I get out of it, not because I believe that my donation is necessary in order for me to enjoy the programming, but rather because I think of myself as one among many people who all support the station together by each giving something. The difference between this approach and Silk's altruistic approach is that a participant in my sense needn't think of his donation as conferring a benefit on other people or on his community as opposed to himself. In this way, my proposal holds on to the appealing idea that there is an important difference between altruism and cooperation. Our ability to cooperate reflects not an ability to act *for* others, but an ability to act *with* others.

## **Back to Ethics**

Now I want to return to the question of what significance all of this has for ethics. As I said before, there is a common way of drawing out the significance of cooperation for

ethics that I want to reject. I think it would be a mistake to look for the origin of rights and obligations in the dependence that cooperators in mutualistic situations have on one another. But I do think that the different approach to cooperation that I have sketched above suggests how, in another way, a capacity to cooperate does underlie our capacity for at least some aspects of moral virtue.

We can see the important connection by reflecting on ways in which *practices* share some of the features of cooperative action that I have been discussing. Take the practice of making and keeping promises. It is evident, and a very fortunate thing, that this practice can function despite its being frequently abused. That is to say, it is not required, in order for the practice to succeed, that everyone do his or her part in it.

Yet at the same time it is clear that the practice would suffer if no one ever made a promise intending to keep it, or kept a promise simply because he had promised. Practices like this are subject to their own form of the paradox of participation: someone has to do his part in them in order for them to function, but no one in particular. What we may call “practice-based accounts” attempt to give an account of the obligation to keep promises that appeal in some way to the value of the practice. Though practice-based accounts are onto something important, they have been known to run into trouble on account of the paradox of participation.

Practice-based accounts are on firm ground when they point out that the obligation to keep one’s promises depends on the *good* that the practice serves. The obligation to keep a promise is not like a requirement in a game, say that one does not pass “Go” on the way to jail in Monopoly. Nor is a genuine obligation like the “requirements” of fashion, or local custom, or *taboo*. A needful practice such as that of

promising has a kind of authority or necessity that games and fashions do not have, and this authority or necessity arises from the way the practice makes the lives of people who live with them go better. Likewise, practice-based accounts are attractive in proposing that we understand certain virtues – in the present case, the kind of justice or honesty involved in keeping promises – by seeing the place it has in the practice. The thought, roughly, is that to possess the virtue is simply to be a full participant in the practice, to be one who has internalized the practice and made it his own. It is plausible to picture this virtuous individual as acting not only from the practice, in the sense of having internalized it, but also as acting *for the sake of* the practice. The virtuous individual himself recognizes the good that the practice secures and can tell the difference between the authority of the practice and the whims of fashion or *taboo*.

But practice-based accounts have been known to stumble when they have to negotiate the transition from talking about the practice and what it requires to talking about the individual agent and his obligations. It can't be said that the individual agent is obligated to keep his promises on the ground that if he breaks them he will be doing harm to a valuable practice. The reason for this is, of course, the aforementioned and very fortunate durability of the practice. It is not necessary for me to keep my promises in order for *me*, or indeed *anyone else*, to enjoy the blessings of the practice. That is just to say that this practice is marked by the paradox of participation.

Now the difficulty posed by the paradox can lead us to be suspicious of practice-based accounts of the obligation to keep promises. Thoughts about my individual obligations, we may conclude, have to be protected from the corrosive effects of thinking too hard about whether *my* contribution is really necessary for the success of the practice.

The practice may be valuable, and that may in some way be the basis of the obligation, but those thoughts can figure only at some other “level” and must not be allowed to influence my deliberations about whether to keep *this* promise here and now.<sup>14</sup>

This conclusion seems peremptory, and it loses what was most appealing in practice-based accounts: namely that the just or honest individual is moved by his understanding of the value of the practice. The virtuous participant appreciates how the requirements of a genuinely valuable practice differ from the tyranny of fashion or taboo. We can avoid the peremptory conclusion, and hold on to what is valuable in practice-based accounts, if we can allow for the kind of strongly social rationality I have been describing.

Our puzzle, recall, was to understand how willing participants could make sense of what they were up to in the face of the fact that their participation was not necessary for the success of their enterprise. But we can make sense of this if we think of one who keeps his promises honestly as reasoning in what I called a “participant” mode. He understands that *someone* has to participate in order for the practice to work, and he understands that, so to speak, *he is someone*, just as much as anyone else. These reflections suggest that we may need social rationality in order to understand at least some of the rationality that goes into virtuous action, for example, in the keeping of promises.

The kind of rationality at work here – what I have called “participant” rationality – is social in a special sense. It does not simply involve the idea of acting *on* others,

---

<sup>14</sup> The *locus classicus* of this response on behalf of practice-based accounts is, of course, John Rawls’ “Two Concepts of Rules”. Michael Thompson argues that Rawls’ attempt to isolate considerations “within the practice of promising” from those “outside” it threatens to assimilate the practice too much to a mere game. See *Life and Action* 174-179.

whose interests I have reason to take into account. Nor is it simply the idea, made familiar by Thomas Nagel's work on altruism, that in some ultimate sense what I have reason to do depends on what reasons others could also accept. Participant rationality is also social in a sense that goes beyond the kind of sociality acknowledged by the mutualistic paradigm: an ability to reason toward actions which I can only perform given certain intentional actions on the part of others. My imagined promise-keeper acts and reasons socially in the sense of adopting as his practical point of view the point of view of a characteristic participant in a practice. That requires the idea of acting from a certain socially defined position, just as, for instance, acting as a committee member or a public radio listener does. In each case, a participant makes sense of his action in virtue of its being one among many others, each of which flows from the shared action or practice.

Now it is important to keep in mind that acting as a participant in some practice or other does not entail acting virtuously. This same "participant" structure can be exhibited in perfectly atrocious ways of life. We may even be wrong about the value of a practice such as promising: perhaps the making and keeping of promises is after all a petty, jealous business, and it would be better if we all sought one another's good spontaneously, without being bound by our word. All I have tried to show is that *if*, as we usually suppose, there is a virtue in making and keeping promises, then it must exhibit the "participant" structure I have described.

### **Concluding Remark**

In conclusion, I want to draw attention to the attitude I have taken toward the investigation of ethics in the behavioral sciences, such as the work of Tomasello and Skyrms. Philosophers tend to take one of two attitudes toward this kind of natural-

scientific approach. They either think that it's already turning up a lot of significant results, or that it can't, in principle, show us anything of importance. I have tried to cultivate a third attitude in this essay: an attitude of looking at the details, and criticizing some of the underlying *philosophical* conceptions that shape these scientific investigations. I have no interest in making the work of explaining ethics in evolutionary terms any harder than it has to be. I just want it to be as hard as it has to be. What I have tried to bring out here is how much of the work of explaining the full depth of our capacity to cooperate remains to be done.

## References

- Bratman, Michael. (1999). I Intend that We *J.* In *Faces of Intention: Selected Essays on Intention and Agency* (pp. 142-161). Cambridge: Cambridge University Press.
- Gilbert, Margaret. (2000). What Is It for *Us* to Intend? In *Sociality and Responsibility: New Essays in Plural Subject Theory* (pp. 14-36) Lanham, Md.: Rowman and Littlefield.
- Rawls, John. (1999). Two Concepts of Rules. In S. Freeman, ed., *John Rawls: Collected Papers* (pp. 20-46). Cambridge, Mass.: Harvard University Press.
- Skyrms, Brian. (2004). *The Stag Hunt and the Evolution of Social Structure*. Cambridge: Cambridge University Press.
- Thompson, Michael. (2008) *Life and Action: Elementary Structures of Practice and Practical Thought*. Cambridge, Mass.: Harvard University Press.
- Tomasello, Michael, with Dweck, C. Silk, J., Skyrms, B., and Spelke, E. (2009). *Why We Cooperate*. Cambridge, Mass.: MIT Press.
- Velleman, David. (1999). How to Share an Intention. In *The Possibility of Practical Reason* (pp. 200-220). Oxford: Oxford University Press.
- Warneken, Felix and Tomasello, M. (2006). Altruistic Helping in Young Infants and Chimpanzees. *Science*, 311: 1301-1303.
- Warneken, Felix and Tomasello, M. (2007). Helping and Cooperation at 14 Months of Age. *Infancy*, 11(3), 271-294.
- Warneken, Felix, Chen, F., and Tomasello, M. (2006). "Cooperative Activities in Young Children and Chimpanzees". *Child Development*, 77(3), 640-663.