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

Manuscript Number: ROPP31

Title: Rationality and the unit of action

Article Type: Joint Action: What is Shared?

Keywords: Action; reasons; team thinking; joint action; plans

Corresponding Author: Dr Christopher Woodard,

Corresponding Author's Institution:

First Author: Christopher Woodard

Order of Authors: Christopher Woodard

Abstract: This paper examines the idea of an extended unit of action, which is the idea that the reasons for an individual action can depend on the qualities of a larger pattern of action of which it is a part. One concept of joint action is that the unit of action can be extended in this sense. But the idea of an extended unit of action is surprisingly minimal in its commitments. The paper argues for this conclusion by examining uses of the idea of an extended unit of action in four theoretical contexts. It also explains why the idea of an extended unit of action need not involve magical thinking, and discusses possible replies to an objection based on a worry about recklessness.

Suggested Reviewers:

Title: Rationality and the unit of action

Author: Christopher Woodard

Affiliation/contact:
Associate Professor in Philosophy
Department of Philosophy, University of Nottingham, Nottingham, NG7 2RD, UK
christopher.woodard@nottingham.ac.uk
Tel. +44 (0) 115 951 5844
Fax. +44 (0) 115 951 5840

Acknowledgements:
I am very grateful to the late Susan Hurley, the late Michael Bacharach, and Alex Gregory for very helpful discussion of these issues.

Omitted references

Woodard, C. (2003). Group-based reasons for action. *Ethical Theory and Moral Practice*, 6: 215-229.
Woodard, C. (2008a). A New Argument Against Rule Consequentialism. *Ethical Theory and Moral Practice*, 11: 247-61.
Woodard, C. (2008b). *Reasons, Patterns, and Cooperation*. New York: Routledge.
Woodard, C. (2009). What's wrong with possibilism. *Analysis*, 69: 219–226.

Rationality and the unit of action

Abstract: This paper examines the idea of an extended unit of action, which is the idea that the reasons for an individual action can depend on the qualities of a larger pattern of action of which it is a part. One concept of joint action is that the unit of action can be extended in this sense. But the idea of an extended unit of action is surprisingly minimal in its commitments. The paper argues for this conclusion by examining uses of the idea of an extended unit of action in four theoretical contexts. It also explains why the idea of an extended unit of action need not involve magical thinking, and discusses possible replies to an objection based on a worry about recklessness.

Keywords: Action; reasons; team thinking; joint action; plans

One concept of joint action is of an action composed of parts such that the reasons to perform each part depend on the qualities of the whole. For example, *electing the government* might be considered a joint action in this sense, because it might be thought that each government-supporting voter's reason for casting her vote depends on the effects of the election as a whole. The crucial idea here is that the focus of our attention should be on the part-whole relationship—not any cause-effect relationship—between the individual vote and the election. On this way of thinking of each voter's action, the reason to perform it is not that it *causes* the election of the government, but that it is *part* of the election of the government. I will explain further below what difference this makes.

This concept of joint action—which I shall refer to as the idea of an *extended unit of action*—is a component of some much richer concepts of joint action. But in itself it is a surprisingly thin idea, or so I shall argue in this paper. Few conditions must be satisfied for individual actions to be related in such a way that the reasons for each

depend on the qualities of the whole pattern of action of which they are parts. The agents concerned needn't share any goals, or be willing or intending to cooperate with each other, for example.

My method will be to survey several different contexts in which the idea of an extended unit of action has been deployed within philosophy and allied disciplines. Discussions within each context tend to operate largely in isolation from each other. Pointing out the common elements at work in each may help us to keep an open mind about what is involved in the idea of an extended unit of action. What appears to be an obvious requirement of the idea in one context may not appear so obvious in another context. The result of keeping this open mind, I will claim, is that we should not build much into the concept. None of this, of course, is meant to deny that there are *other* concepts of joint action which are much richer and have more demanding conditions.

Section 1 surveys a range of uses of the idea of an extended unit of action in four different contexts. Section 2 distils the essential features of the idea, and explains why it need not involve any magical thinking. Section 3 defends the minimal version of the idea as I have outlined it from the objection that it involves a kind of recklessness. Section 4 concludes. Throughout, I'll use the phrase 'the actor' to refer to an individual agent whose normative reasons for action are under discussion.

1. A range of views in four contexts

1.1 Perhaps the most familiar context for the idea of an extended unit of action is in debates about the structure of consequentialism. The contrast between Act Consequentialism and Rule Consequentialism is very well known. Act Consequentialism gives an account of reasons in terms of the goodness of individual actions. In contrast, Rule Consequentialists explain reasons in terms of the goodness of widespread compliance with (or acceptance of) sets of rules. If the explanation is given in terms of compliance it is clear that this amounts to use of an extended unit of action in the sense

at stake here. These forms of Rule Consequentialism aim to account for reasons for or against individual actions in terms of the goodness or badness of widespread compliance with sets of rules. But compliance with a set of rules may be understood as a very extended pattern of action, consisting of every occurrence of compliance. Thus these forms of Rule Consequentialism employ the idea of an extended unit of action (_____ 2008a: 260).

Other variants of consequentialism also employ the idea of an extended unit of action. Collective Consequentialism explains reasons for or against individual actions in terms of the goodness of larger patterns of action performable by some relevant group (Mulgan 2001). Plan Consequentialism explains reasons in terms of the goodness of a larger pattern of action performable by the actor herself (Feldman 1997; McClennen 1985,1990). Cooperative Consequentialism explains reasons in terms of the goodness of the best pattern of action performable by all cooperative agents (Regan 1980; Hurley 1989: 136-159, 1991, 1994, 2005). Other variants may also use the idea of an extended unit of action. Any so-called 'indirect' version of consequentialism that explains reasons in terms of something that amounts to a pattern of action uses this idea.

I shall describe such theories from a fairly abstract point of view, glossing over many of their details. My aim is to draw broad comparisons between uses of this same idea in different proposals and contexts, and to identify certain fundamental questions facing any such use. As such, I will not always try to describe the fine detail of different proposals—though for other purposes these details may of course be crucial.

1.2 Consider now a second context. There is a debate in deontic logic between so-called 'actualists' and 'possibilists' (Bergström 1966; Sobel 1976; Goldman 1976; Jackson and Pargetter 1986; Jackson 1987; Zimmerman 1996: Ch. 6). These two camps disagree about the concept of moral obligation. Actualists claim that obligation depends on facts about how the actor's environment would respond to the choices she could make. These facts could include facts about her own response to those choices. In one of

the central examples, Professor Procrastinate has been invited to review a book (Jackson and Pargetter 1986: 235). He faces a sequence of choices. *Now* he must either accept or reject the invitation; if he accepts now, *later* he will face a choice between writing or not writing the review. The best outcome would be reached by accepting now and writing later. The next best outcome would be reached by declining now. The worst outcome would be reached by accepting now, but not writing later. Though he could accept or decline, and he could write or not write, as a matter of fact he would not write later, were he to accept now. Actualists believe this fact about his own response to accepting the invitation implies that he ought to decline. Possibilists deny this, and claim that it lets Procrastinate off the hook too easily. They claim that what he ought to do now depends on how he could best respond later, not on how he would respond later. Thus, they treat the actor's own response differently from the response of other parts of her environment. They claim that we should focus on the *best* response *she could* make, but on the response the rest of her environment *would* make.

The debate between actualists and possibilists is usually couched in terms which obscure its relevance to the idea of an extended unit of action. It tends to be described as being about *which* consequences of an *individual* action matter for obligation. Actualists claim, of course, that it is the 'actual' consequences that matter—where this means the consequences the action would have, were it performed. Possibilists claim, in contrast, that 'merely possible' consequences matter. But note that they are interested only in certain very specific merely possible consequences: namely, those that would occur were the action in question performed, and the actor were to perform *some other action she could perform*. Possibilists are not interested in any other merely possible consequences, such as those that would follow were the weather different than it will be, or were some other agent to behave in some way he will not. Only those possible consequences that are accessible by some combination of actions performable by the actor are said to matter.

For this reason it is more helpful to conceive of the debate between actualists and possibilists as being about whether an extended unit of action, which includes only actions performable by the actor, generates reasons for her to act now. Rather than being about which consequences of individual actions matter (the 'actual' versus the 'merely possible'), it is better to think of it as being about the relevant unit of action, the ('actual') consequences of which matter. Putting things this way explains which particular 'merely possible' consequences are the subject of the dispute, and thereby makes better sense of possibilist views (Jackson 1987: 106; _____ 2009). Possibilist versions of consequentialism are thus best understood as versions of Plan Consequentialism.

1.3 A third context in which we find the idea of an extended unit of action is in debates about the foundations of game theory and decision theory. Orthodox views in these areas define rational action in terms of the consequences of the various individual actions open to the actor.[1] As is well known, however, these accounts of rationality face various difficulties in certain collective action problems and their intrapersonal analogues, and this has led some theorists to propose revised accounts of rationality incorporating use of extended units of action (Gold and Sugden 2007; Bacharach 2006; McClennen 1985, 1990).

The prisoner's dilemma (PD) is probably the best known example of a collective action problem. Consider the two-player, one-shot version of this game shown in Figure 1.

		Column: C	Column: D
Row	C	3, 3	1, 4

[1] Of course there is an important difference between *expected* and *actual* consequences. But since this issue is orthogonal to those discussed in this paper, I will ignore it.

D	4, 1	2, 2

Figure 1: A two-person, one-shot PD

Row's payoffs are given by the first figure in each cell, while Column's payoffs are given by the second. A higher figure denotes a better payoff, by the lights of the player concerned. Each player faces a choice between two actions: *C* and *D*. Each player's choice is causally independent of the other's. The result of each action depends on the choice of the other player, in the way shown by the matrix. Thus, if both players choose *D*, the result is an outcome that each values at 2. If Row chooses *D* while Column chooses *C*, the result is an outcome that Row values at 4, but which Column values at 1; and so on.[2]

As is familiar, there is a simple and powerful argument to the effect that the rational choice for each player is *D*. First consider things from Row's point of view. Suppose Column were to choose *C*. Then the result of Row choosing *C* would be worse, by Row's lights, than the result of Row choosing *D*. Now suppose Column were to choose *D*. Again, the result of Row choosing *C* would be worse, by Row's lights, than the result of Row choosing *D*. So, whatever Column chooses, Row does better to choose *D*. Choosing *D* is Row's *best reply*, whatever Column chooses—which is to say that *D* is a *dominant strategy* for Row. If we consider things from Column's point of view, the situation is entirely symmetrical. Once again, choosing *D* is dominant, since it is the best reply to the other's choice, whatever that is. Plausibly, rationality requires choice of dominant strategies when they are available. So, plausibly, rationality requires that each player choose *D* in this case.[3]

[2] We need not ask why each player values these outcomes in these ways. It could be a matter of self-interest, altruism, theology, or something else. All that is necessary is that the outcomes are valued in the ways shown, and depend on the players' actions in the ways shown.

[3] Howard (1988) discusses a case in which each player can detect the other's decision rule, and argues that it would be rational to choose *C* if each player has a rule which says roughly 'choose *C* iff the other has this same rule'. He also shows that this need not involve any regress.

However, the result of both players choosing *D* is an outcome that each values less than each values the result of both choosing *C*. That is, the result of each player choosing *C* is *stronglypareto superior* to the result of each player choosing *D*.[4] Hence, doing the thing that rationality plausibly requires in this case leads the players away from an outcome that is possible and that both agree is better. This seems to be bad news: either for rationality itself, or for our understanding of what it requires (Parfit 1987: 53–55; Colman 2003).

What, if anything, has gone wrong? One diagnosis is as follows. The concepts of *best-reply* and *dominant strategy* both conform to the standard view of practical reasons, since they make claims about the actor's reasons based on an assessment of the difference some *one* action at a time would make, holding other actions constant. A best-reply, after all, is a choice that maximises the actor's payoff given the other player's choice. A strategy is dominant just in case it picks out a best-reply no matter what the other player chooses. Both concepts therefore examine the differences that could be made to the outcome by single actions.[5] The argument for choosing *D* thus makes use of a series of what we might call *horizontal* and *vertical* comparisons between cells in the payoff matrix. Holding Row's choices constant, it uses horizontal comparisons to examine the differences Column's choices could make. Holding Column's choices constant, it makes vertical comparisons to examine the differences Row's choices could make. It may consider several such comparisons. What it never does is make the crucial *diagonal* comparison between the result of both players choosing *D* and the result of both choosing *C*.

The concept of pareto superiority applies to this diagonal comparison, of course—but pareto superiority is a guide to value, not to reasons for action. Orthodox game theory and decision theory use no diagonal comparisons in their claims about reasons, because doing so would involve violating the standard view of reasons, to

[4] *X* is *weakly* pareto superior to *Y* when some prefer *X* and none prefer *Y*. *X* is *strongly* pareto superior to *Y* when all prefer *X* to *Y*.

[5] The same is true of the concept of Nash equilibrium, and David Lewis's concept of a 'coordination equilibrium' (1986: 14).

which they are committed. If we try to derive a reason to choose *C* in the PD from the comparison of <C, C> with <D, D>, for example, we are saying that there is a reason to choose *C* because it is part of a larger unit of action that is favourable.[6] This contradicts the standard view.

In light of this, some theorists propose revising the orthodox account of rationality in game theory and decision theory. Robert Sugden has developed an account of 'team thinking' which makes use of the idea of an extended unit of action. In Sugden's theory, it is rational to engage in team thinking when others are doing so as well, such that every member of some group identifies with the group as a team and seeks to maximise some shared team utility function. When those conditions are met, she has a reason to play her part in the profile of actions by the members of the group that would maximise the shared utility function (Sugden 1993: 84-89, 2000: 190-203; Gold and Sugden 2007; Bacharach 1999, 2006; Colman et. al. 2008). If we assume that the players in a PD satisfy these conditions, *and* that the shared utility function ranks <C, C> as uniquely best, then each player has a reason to choose *C*. In this way the theory of team thinking combines two important ideas: an extended unit of action, such that the actor's reason to choose *C* is derived from the goodness of the profile of actions that produces <C, C>; and a team-level evaluation of the outcomes, such that <C, C> is ranked as best from the 'team's point of view', as we might put it. I'll return briefly to this second idea later. The resulting account of rationality helps explain intuitions about reasons for action not just in the PD but in a range of collective action problems (Gold and Sugden 2007).[7]

1.4 The final context I wish to discuss is Kantian ethics. One reason to include it is to counteract the impression that may have been gained that the idea of an extended unit of action is relevant only to theories that are, in a broad sense, consequentialist. Each of

[6] '<C, C>' here denotes the outcome produced by both players' choice of *C*.
[7] McClennen (1985; 1990) proposes a theory of 'resolute choice' which modifies decision theory by adopting an intrapersonally extended unit of action. This helps to explain some intuitions about rational sequential choice in cases that are intrapersonal analogues of collective action problems.

the proposals I've discussed in the other three contexts used the idea of the good consequences of a larger unit of action to make claims about reasons to perform parts of it. But there is nothing essentially consequentialist about the idea of an extended unit of action.

Indeed, this idea may be essential in Kantian ethics. One respect in which it seems important is in the idea of an imperfect duty. According to Kant some duties, such as the duty not to lie, are perfect; others, such as the duty to develop one's talents, are imperfect (Cummiskey 1996: 114-122). This contrast seems to be a matter of laxity in guiding individual actions. Perfect duties set strict limits, or require certain things on specific occasions. Imperfect duties don't do this. For example, I can comply with the duty to develop my talents even while I am skipping the faculty research seminarto play a valueless video game, so long as I develop my talents sufficiently on other occasions. Because it is imperfect, this duty does not imply that any particular action is forbidden or required.[8] Instead, imperfect duties seem to imply that *lives*, or anyway extended periods of lives, with certain characteristics are forbidden or required. A life without development of one's own talents, for example, is forbidden. Though I can skip the research seminar this time, I cannot pass up similar opportunities every time they arise—for then I would have lived an impermissible life.

This picture of imperfect duties suggests that the idea of an extended unit of action is required to explain how imperfect duties ever give rise to reasons to perform individual actions. If I have a reason to go to the research seminar because of the imperfect duty to develop my own talents, for example, that reason must have the following form: going to the seminar is a constituent of one of the permitted lives, of which I am required to live one. Thus, a reason to perform individual action *A* is said to depend on the rightness of some larger unit of action (in this case, a series of actions

[8] There are limiting cases which complicate this story. Presumably, the duty forbids committing suicide. I shall set these complications aside.

constituting a permitted life). Kantian ethics thus seems to make important use of the idea of an extended unit of action.[9]

1.5 This brief survey of proposals in four contexts suggests the following points. First, it is clear that the idea of an extended unit of action is recurrently attractive. As I've said, each of these discussions tends to occur without much reference to any of the others. This suggests the following sociological speculation: the same idea somehow occurs to people who are interested in rationality as being part of an attractive view for some purpose, more or less independently of its having occurred to others working on related questions in different contexts. Whether or not that speculation is true, the idea of an extended unit of action is certainly recurrently attractive, and I take that to be evidence that it merits attention. It would be good to understand this idea and the issues it raises.

Second, the proposals fall into various types. One obvious dividing line is between views that limit the extended unit to the actor's own behaviour, and those which allow it to extend across individuals. Possibilism, Plan Consequentialism, and Kantian imperfect duties use the idea for the actor's own behaviour only. Rule Consequentialism, Collective Consequentialism, Cooperative Consequentialism, Team Reasoning, and (perhaps) other elements of Kantian ethics use the idea for other individuals' behaviour. Obviously, this issue is extremely important for the character of the resulting views. A second dividing line is more specific. This is between views that impose specific *eligibility constraints* on the behaviour that can form part of the unit of action, and those that do not. Cooperative Consequentialism, for example, disallows any behaviour that is not the behaviour of a cooperative agent. Team Thinking requires not just cooperativeness, but also that every agent identifies as part of the team, and shares the team-level objective. Rule Consequentialism, on the other hand, extends the unit of

[9] If Parfit (2011) is right that we can think of Kantian contractualism as converging with Rule Consequentialism, and I am right that Rule Consequentialism depends centrally on the idea of an extended unit of action, we might conceive Kant's ethics as using this same idea more generally, not just with respect to imperfect duties.

action to include the behaviour of all those capable of performing the type of action in question, whether or not they are cooperatively disposed or identify as part of the group of all such agents. Third, some views marry the idea of an extended unit of action with the idea of group-level evaluation, while others do not. Certainly these two ideas are conceptually distinct, though it might be possible to argue that the first is implausible without the second (_____ 2003: 215).

One assumption about eligibility might explain why some views restrict the unit of action to the individual actor, and perhaps why some views require group-level evaluation. This is the assumption that the unit of action mirrors the unit of agency. On this *mirroring assumption*, the unit of action could extend across individuals only if several individuals could together comprise a supra-individual agent. Some of the proposals we have surveyed might tacitly accept this assumption. This could help explain, for example, why possibilists extend the unit of action only to the actor's own behaviour, or why Team Thinking requires team identification and shared team-level evaluation. Note, though, that other proposals clearly do not accept the mirroring assumption.

Thus, abstracting from the details of a particular proposal in one context so that we can see the similarities between different proposals across contexts helps us to identify controversial presuppositions. And, indeed, later I shall challenge the assumption that the unit of action mirrors the unit of agency.[10]

2. The commitments of the idea of an extended unit of action

Suppose we are persuaded that the idea of an extended unit of action is interesting. What exactly does it involve? What theoretical commitments does it bring?

[10] I learned the concept of what I call the 'unit of action' from Susan Hurley. She tended to use the phrase 'unit of agency' instead to describe this concept (1989: 145-148). This usage has been somewhat influential. However, I prefer 'unit of action' because it makes it easier to see that the mirroring assumption is a substantive thesis. In later work (e.g. 2005), Hurley used the phrase 'unit of activity'.

At its most basic, the idea of an extended unit of action is that there can be a reason for or against some action, *A*, because it is part of larger pattern of action, *P*, which is good or right. As I shall sometimes say, it is the idea that there are *pattern-based* reasons (_____ 2008b). Such reasons exist, if at all, because*A* is a part rather than a cause of*P*(Hurley 1989: 148). That is, the idea is not that there is a reason to perform *A* because it will bring about *P*, or make *P* more likely. Instead it is that the goodness or rightness of *P* provides reasons to perform its parts, just in virtue of their being parts of it. There may of course be constraints on which patterns of action can support such reasons—eligibility constraints, as I'll call them. But the crucial point for now is that the idea of pattern-based reasons, or of an extended unit of action, is not the commonplace idea that there can be a reason to do something because it will lead to some sequence of actions that is good or right. As we'll see shortly, adopting an extended unit of action in fact involves suspending our usual predictive stance towards parts of the actor's environment. This is a puzzling feature of the idea, which I shall try to explain.

Another important preliminary point is that the idea is not that there can be units of *agency* other than individuals. The idea of an extended unit of action does not specify the unit of agency that could perform the action *A*, or the pattern *P*. Most commonly, perhaps, *A* is an action performable by some individual at some particular time. The larger pattern, *P*, could then be an extended plan of action performable by that individual over time, or a pattern of action performable by some group of individuals, *G*. These two natural applications of the idea of pattern-based reasons give us the idea of *plan-based* and *group-based* reasons for individuals, respectively. Such reasons would remain reasons for *individuals* to act at specific times, however. They would not be reasons for individuals-over-time to act, or for groups to act.

Thus the idea of pattern-based reasons, or of an extended unit of action, is compatible with, but does not presuppose, the idea that there can be units of agency other than individuals. It does presuppose that the larger pattern, *P*, is performable in

some sense. But that could be merely a matter of each of its parts being performable. However, if there could be supra-individual agents, we could generalise the idea of an extended unit of action. Some supra-individual agent *S* might perhaps have reasons to perform some action *H* because it is part of some yet-larger pattern of action *J*. *J* could be a plan performable by *S* over time, or a pattern performable by some group of which *S* is a member. In this way, the idea of an extended unit of action is compatible with a wide range of views about the kinds of agents there could be.

We can now characterise the standard view of reasons for action, from which the idea of pattern-based reasons is a departure, more precisely. According to this view, the unit of action is always identical to the action in which we're interested. If we're interested in the reasons for or against *A*, the unit of action is *A* itself, according to the standard view. Thus we can call the reasons that exist according to this view *act-based reasons*. As with the idea of an extended unit of action, the claim that the unit of action is identical to *A* does not presuppose, but is compatible with, the idea that *A* is an action performable by some supra-individual agent. Thus we should be careful to remember that our topic is not whether collective agents exist or have reasons. It is whether the goodness or rightness of larger patterns, perhaps performable (in some sense that is not specified by the bare idea) by groups, can give rise to reasons to perform their parts. According to the standard view, they cannot. The idea of an extended unit of action denies this, and thereby asserts the possibility of pattern-based reasons of some kind.

To see what's involved, let us consider a specific example. Suppose we're interested in Art's reasons for action. His problem is whether to go to the café or the park for lunch. He'd prefer meeting Susie to not doing so, and meeting her in the park to meeting her in the café. She has the same preferences, *mutatis mutandis*. Unfortunately, neither can remember the arrangement they made, and they have no way of communicating. All these facts are common knowledge amongst them.

According to the standard view, Art's reasons depend on the goodness or rightness of each of his options, namely *café* and *park*. Let's assume that neither of these

options is intrinsically right or wrong; so we're interested only in their goodness. The goodness of each of them depends on Susie's choice. If she chooses *café*, Art does best to choose *café* too. If she chooses *park*, on the other hand, he does best to choose *park*. If he has a credible prediction of her choice he may be able to settle on one of these options. However, the symmetry of their situation may make this difficult. His prediction is likely to depend on Susie's reasons, if she is rational. If it is based *entirely* on an assessment of her reasons, as in game theory, then he is locked in a circle, since his reason to choose one option rather than the other depends on her reason to do likewise, and her situation is exactly symmetrical to his (Regan 1980: Chapter 2; Hurley 1989: 154-155; Sugden 2000: 179-181). Thus game theory may suffer from a problem of indeterminacy in simple coordination problems such as this (though that is not endemic to all theories of rationality that conform to the standard view, since there could be some other basis for the prediction of the other agent's behaviour).

Now suppose that we claim that the unit of action for Art's choice between *café* and *park* includes not just his behaviour but Susie's. (Shelve for the moment any concerns about the rationale for doing this; we're just trying to see what's involved in the idea.) This means that we consider the rightness or goodness of the various possible combinations of choices they could make. Once again we can assume that no combination is intrinsically right or wrong, so that we have only their goodness to consider. Of the possible combinations, both choosing *park* is, they both agree, the best. Thus (assuming Art's reasons in this case depend on his preferences) we may say that Art has a pattern-based reason to choose *park*, since that is his part in the best pattern of action he and Susie could perform.

Note the following two very important points about this train of thought. The first is that there is no danger of falling into the circle that orthodox game theory can find itself trapped in, since Art's reasons are not said to depend on Susie's reasons. Art's reasons depend instead on what it would be best for them both to do (as do her reasons). This is part of the attraction of the idea of an extended unit of action. However,

the second point highlights the cost of the idea, while explaining the first point. The reason that there is no danger of falling into the circle is that Art's reasons are not said to depend on *how Susie would react to his choice*. The circle arises when we say that one agent's reasons depend on how another agent would react, and our prediction of that reaction depends on assessing the second agent's reasons, in cases where those reasons are symmetrical to the first agent's reasons(Sugden 2000: 191). It's obvious that we can avoid this at the first step, by denying that the first agent's reasons depend on how the second agent would react. But many find that an absurd claim, and this explains most of the resistance to the idea of an extended unit of action.

So if we're to take this idea seriously we need to examine the alleged absurdity more closely. One suspicion might be that extending the unit of action to include other agents' behaviour involves some sort of magical thinking (Elster 1989: 192-202). In particular, it might be thought to rely on the tacit assumption that the actor has the power to determine what other relevant agents will choose. For example, if we suppose that Art has a pattern-based reason to choose *park* because it is part of the best pattern performable by him and Susie, it might be thought that we are supposing that his choice would bring it about that the best pattern is realised. That would amount to believing that, by choosing *park*, he could cause Susie to do the same.

In some specific situations this sort of control of another's response might be possible. But of course it is not generally possible, which is why this sort of assumption is appropriately described as 'magical thinking'. However, the idea of an extended unit of action is in no way committed to such thinking. Art's pattern-based reason to choose *park*, supposing he has one, does not depend on any claim about how Susie would respond to that choice. As I explained above, it is supposed to follow from the fact that his going to the park is part of the best pattern they could perform. It thus depends on claims about what each of them could do. But it does not rely on any claims about what either of them *would* do. To suppose otherwise is to misunderstand the nature of the idea of an extended unit of action, or of pattern-based reasons.

Far from relying on predictions based on dubious magical thinking, pattern-based reasons don't rely on predictions (about behaviour that is part of the unit of action) at all. Though we usually adopt a predictive stance towards the behaviour of relevant parts of our environment when deliberating, we suspend this stance towards behaviour that is part of the unit of action. The standard view of practical reasons recognises this with respect to the options open to the actor, of course. No-one thinks that in deliberating whether to choose A or $\sim A$, I should try to predict whether I will choose A or $\sim A$. Predicting *other* behaviour by myself might be pertinent; but predicting the very thing under deliberation is not. This is not because the nature of human freedom makes such predictions impossible. It is just because such predictions are irrelevant to the question being asked. In deliberating about A or $\sim A$, I am asking myself which of these options is best or right. Answering 'I will choose A' or 'I will choose $\sim A$' simply fails to address the question. For this reason, deliberation about a piece of behaviour 'crowds out' prediction of that piece of behaviour.[11]

The same goes for your choice of B or $\sim B$, if that behaviour of yours is part of the unit of action for my choice of A or $\sim A$. When I include your behaviour in the unit of action, I address myself to the goodness or rightness of the various possible combinations of A, $\sim A$, B, and $\sim B$. If that is the question I'm strictly interested in, it is irrelevant to ask whether you will choose B or $\sim B$. Of course, prediction of whether you would respond to my choice of A with B or $\sim B$ may well be relevant to whether I should be interested in a unit of action including your behaviour. Likewise, it may be relevant to the significance for my reasons of the rightness or goodness of a pattern involving your behaviour. Those are different questions, however (to which I will return below, in §3). Similar remarks go of course for my own future behaviour insofar as I include that in an extended unit of action.

So the central commitments of the idea of an extended unit of action are as follows. The basic idea is that the goodness or rightness of a pattern of action can

[11] The phrase is Levi's (1997: ix).

provide reasons for performing parts of that pattern. It is crucial that these reasons do not depend on causal claims about the relationship between the part—the action for which there is said to be a reason—and the completion of the pattern. Pattern-based reasons depend on parthood relations, not causal relations, between the constituent action and the favoured pattern. This explains why deliberation about pattern-based reasons involves suspending our normal predictive stance towards all parts of the pattern. Because we are not asking whether the pattern would come about, but whether it would be good or right, it is irrelevant to predict that this or that part of it would or would not be performed.

3. Recklessness

I've explained that the idea of an extended unit of action, or of pattern-based reasons, involves the idea that we should suspend our normal predictive stance towards all parts of the actor's environment that are included in the unit of action. This is not always fully appreciated because it is all too easy to slip into the confusion that pattern-based reasons depend on causal relations between favoured patterns and their constituent parts. Once we have cleared up this confusion, however, the outlook may seem *worse* for the idea of pattern-based reasons. One imagines a sceptic saying, 'if *that's* what you mean, let's hear no more of it!'.

I've already responded to the charge of magical thinking. The other main way that the idea of an extended unit of action can seem absurd is that it can seem reckless. Claiming that there is a reason to do such and such, because this is part of a good or right pattern of action, without making the reason conditional on predictions about completion of the pattern, seems like an invitation to bring forward cases in which acting unilaterally on such reasons would have disastrous consequences. For example, suppose that the best pattern of behaviour by drivers on UK roads involves their all driving on the right hand side of the road (since that would minimise the risk of

accidents when they drive elsewhere). Surely it's crazy to claim that any individual driver has a reason to drive on the right, where that reason does not depend on any prediction about whether the others will do so too?

Other than engaging in unrestricted bullet-biting, there are at least four ways for the advocate of pattern-based reasons to respond to the worry about recklessness. One is to claim that pattern-based reasons exist only in specific contexts, where the contexts are described in ways such that acting on these reasons does not seem reckless. The leading version of this idea restricts pattern-based reasons to cooperative contexts. For example, Regan's 'Cooperative Utilitarianism' bids each agent to '[hold] himself ready to co-operate with whoever else is willing and able to co-operate . . . [and then to identify] the other co-operators and [do] his part in the best possible pattern of behaviour for the class of co-operators (including himself) in view of the behaviour of the non-cooperators' (Regan 1980: 136). Though this describes a decision procedure, it would be natural for someone holding this view to claim also that there are no pattern-based reasons except when others are cooperative.

Of course, one must be careful to define 'cooperative' and 'potential cooperator' in ways that do not reintroduce the predictive stance. But this seems possible. Regan, for example, defines cooperativeness in terms of decision procedures. Thus, when deliberating using Cooperative Utilitarianism we seek to recognise others' decisions procedures rather than seeking to predict their choices as such (Regan 1980: 142, 178).[12] As he says, '[w]hat each member of [the group of cooperators] needs to know about each other member . . . is just that each other member will eventually act on the best plan *as he* (the other member) *sees it* when he comes to the last step [of deliberation] . . . It is not necessary for one member . . . to know what other members will do under any other description than this'(Regan 1980: 142, emphasis in the original). Providing we can define 'cooperative' in some appropriate way such that we

[12] Recognition of decision procedures is a central feature of Howard's treatment of one-shot PDs, too (1988).

can identify cooperators without predicting their ultimate choices, claims about pattern-based reasons can be conditional on cooperativeness even though they are not conditional on predictions of specific behaviour. The former kind of conditionality may seem sufficient to avoid recklessness.

A variant of this first response to the worry about recklessness claims that pattern-based reasons exist only when all the agents involved together satisfy conditions of joint agency—which presumably include, but go beyond, the sort of cooperativeness that does the work in the previous proposal(Sugden 1993: 87). This is one way of motivating the mirroring assumption described in §1, and it may explain why some claim that the unit of action can be extended only intrapersonally. However, once we've seen the possibility of restricting pattern-based reasons to cooperative contexts, it's not clear that considerations of recklessness speak in favour of further restriction to situations in which there is a prospect of joint agency. This further restriction just seems gratuitous. It's not clear, for example, that others' intentions to engage in a joint action with me further reduces the risk of my performing my part in a favoured pattern, as compared with their intentions simply to cooperate with me in producing the best pattern.[13]

A second possible response to the worry about recklessness is to adopt some device designed to steer agents away from disaster. Rule Consequentialists do this when they argue that one of the rules in the best set instructs the actor to break the other rules if necessary to avoid disaster (Hooker 2000: 98-99; _____ 2008a). In normal cases, Rule Consequentialism claims that we each have pattern-based reasons that are unconditional on any facts about others' responses. I should tell the truth even if you won't, because of the good consequences of widespread truth-telling. However, Rule Consequentialists typically recognise that there can be unusual cases in which following rules that are usually benign would have disastrous consequences. By permitting or

[13] There may be other arguments for the mirroring assumption. However, I do not know of any that seem convincing. For discussion, see _____ (2008b: 74-80).

requiring the agent to break those rules in such cases, a disaster-prevention rule ameliorates the worry about recklessness. The upshot is a theory with a dual stance on whether the responses others would make affect the actor's reasons. In normal cases, they do not; but in cases where disaster looms, they do.

A third response to the worry about recklessness is to restrict pattern-based reasons to kinds of action that seem important enough for the apparent recklessness of performing them unilaterally not to seem a sufficient objection to doing so. This is one way of understanding deontological ethical theories such as Kant's. These theories pick out certain kinds of action that, if universally performed, would constitute a pattern of action that is good or right. They then claim, of those specific actions, that one should perform them unilaterally, despite the consequences that may follow. In saying this, they are not thereby committed to general endorsement of pattern-based reasons. Only certain patterns of action are eligible to support reasons, and not all patterns meet that test. *Not lying* might meet it, for example, even though *driving on the right hand side of the road* does not.

These three kinds of response to the worry have their advantages and disadvantages. Elsewhere I have advocated a fourth kind of response (_____ 2003, 2008a, 2008b). If we are *pluralists* in the sense of permitting more than one unit of action for a single decision problem, we can claim that pattern-based reasons sometimes conflict with act-based reasons in a single case. This means that we can give an account of pattern-based reasons that is not restricted to any particular context or kind of action, while allowing that facts about how other agents would respond are vitally important, since they affect the nature and strength of possibly countervailing act-based reasons. This sort of structure has the advantage of being able to explain intuitions about conflicting reasons across a wide range of cases, and also of finding a natural home for facts about others' responses. The natural home for such facts is in claims about act-based reasons. So long as we are pluralists, we do not have to find some surrogate device for avoiding recklessness. How you would respond matters to my reasons, just

not to any pattern-based reason I have in respect of a unit of action that includes your response. But since such a pattern-based reason must be considered alongside my other reasons, facts about your response may affect its relative strength.

My aim here is not to argue for the superiority or even adequacy of this fourth response to the worry about recklessness.[14] Instead my aim is twofold. First, I want to point out that the worry about recklessness does not amount to a knockdown objection to the idea of an extended unit of action. There are several possible responses to the worry. Second, advocates of some theory in some local context that makes use of one of the methods described here for avoiding recklessness, or some other method, should not be quick to assume that this is the only way of doing so. Again, there are several possible responses to the worry. One advantage of drawing connections between the different proposals using the idea of an extended unit of action that are made in different theoretical contexts should be to encourage further debate about the relative merits of all of these possibilities.

4. Conclusion

The idea of an extended unit of action, which is equivalent to the idea of pattern-based reasons, recurs in many theoretical contexts in which rationality is discussed. It stands opposed to the standard view of normative reasons for action, according to which these depend only on the goodness or rightness of each individual action. Pattern-based reasons are reasons for an action that depend on the goodness or rightness of some larger pattern, of which the immediate action is a part. They thus depend on parthood relations, not causal relations, between the favoured pattern and the action under consideration.

[14] One debt it incurs is to give some account of how act-based and pattern-based reasons may interact in a single case. See _____ (2008b: 107-118).

I suggested that the fact that this idea recurs in different contexts should make us interested in it. This fact is not evidence that the idea is correct, but it is evidence that it is worth understanding the idea well. It also suggests that there may be theoretical benefits in abstracting from the details of specific proposals made in one context or other, to consider the fundamental features of the idea as it appears across these different contexts. One such feature is that pattern-based reasons are not conditional on facts about whether the rest of the pattern would be realised. This explains why, when deliberating according to some extended unit of action, we suspend our normal predictive stance in respect of any behaviour that forms part of the unit. Doing this is not an instance of magical thinking—which would be, on the contrary, a matter of retaining the predictive stance, but making unjustified predictions on the supposition that the actor has the power to control responses to her choices.

However, clarification of this point leads to a different objection. It can seem reckless to suspend the predictive stance. In line with my general method in this paper, I surveyed four different responses to this worry that could be made by advocates of pattern-based reasons. I did not argue for any of these responses, but merely claimed that, in light of their existence, the worry about recklessness should not be treated as a knockdown objection to the idea of an extended unit of action, and nor should advocates of one specific version of this idea assume that their particular response is the only genuine candidate. Pointing out the variety of ways of constructing a theory of pattern-based reasons should lead those of us interested in the idea to examine the different possibilities more closely. Doing that might also enable us to export insights gained in one context to one or more of the others.

References

(Self-identifying references omitted)

Bacharach, M. (1999). Interactive Team Reasoning: A Contribution to the Theory of Co-operation.*Research in Economics*, 53: 117–47.

Bacharach, M. (2006). *Beyond Individual Choice. Teams and Frames in Game Theory*. Gold, N. and Sugden, R. (eds.). Princeton: Princeton University Press.

Bergström, L. (1966).*The Alternatives and Consequences of Actions*. Stockholm: Almqvist and Wiskell.

Colman, A. M. (2003). Cooperation, Psychological Game Theory, and Limitations of Rationality in Social Interaction.*Behavioral and Brain Sciences*, 26: 139–53.

Colman, A. M., Pulford, B. D., and Rose, J. (2008). Collective rationality in interactive decisions: Evidence for team reasoning. *ActaPsychologica*, 128: 387-397.

Cummiskey, D. (1996). *Kantian Consequentialism*. New York: Oxford University Press.

Elster, J. (1989).*The Cement of Society*.Cambridge: Cambridge University Press.

Feldman, F. (1997). World utilitarianism. In Feldman, F. (1997).*Utilitarianism, Hedonism, and Desert*. Cambridge: Cambridge University Press: 20-35.

Gold, N. and Sugden, R. (2007).Theories of Team Agency. In Peter, F. and Schmid, H. (eds.) (2007). *Rationality and Commitment*. Oxford: Oxford University Press: 280-312.

Goldman, H. S. (1976). Dated Rightness and Moral Imperfection.*The Philosophical Review*, 85: 449–87.

Hooker, B. (2000).*Ideal Code, Real World*.Oxford: Clarendon Press.

Howard, J. V. (1988). Cooperation in the Prisoner's Dilemma.*Theory and Decision*, 24: 203–13.

Hurley, S. L. (1989).*Natural Reasons*. New York: Oxford University Press.

Hurley, S. L. (1991). Newcomb's Problem, Prisoners' Dilemma, and Collective Action.

Synthese, 86: 173–96.

Hurley, S. L. (1994). A New Take from Nozick on Newcomb's Problem and Prisoners' Dilemma.*Analysis*, 54: 65–72.

Hurley, S. L. (2005).Social Heuristics That Make Us Smarter.*Philosophical Psychology*, 18: 585–612.

Jackson, F. (1987). Group Morality. In Pettit, P., Sylvan, R., and Norman, J. (eds.) (1987). *Metaphysics and Morality.Essays in Honour of J. J. C. Smart*. Oxford: Basil Blackwell: 91-110.

Jackson, F. and Pargetter, R. (1986). Oughts, Options, and Actualism. *The Philosophical Review,* 95: 233-255.

Levi, I. (1997).*The Covenant of Reason*.Cambridge: Cambridge University Press.

Lewis, D. (1986).*Convention*. Oxford: Basil Blackwell; originally published 1969.

McClennen, E. F. (1985). Prisoner's Dilemma and Resolute Choice, in Campbell, R. and Sowden, L. (eds)*Paradoxes of Rationality and Cooperation*.Vancouver: The University of British Columbia Press: 94–104.

McClennen, E. F. (1990).*Rationality and Dynamic Choice*.Cambridge: Cambridge University Press.

Mulgan, T. (2001).*The Demands of Consequentialism*.Oxford: Clarendon Press.

Parfit, D. (1987).*Reasons and Persons*. Reprint with corrections. Oxford: Clarendon Press.

Parfit, D. (2011).*On What Matters*.Oxford: Oxford University Press, forthcoming.

Regan, D. (1980). *Utilitarianism and Co-operation*. Oxford: Clarendon Press.

Sobel, J. H. (1976). Utilitarianism and Past and Future Mistakes.*Noûs*, 10: 195–219.

Sugden, R. (1993). Thinking as a team: Towards an explanation of nonselfishbehavior. *Social Philosophy and Policy*, 10: 69-89.

Sugden, R. (2000). Team Preferences. *Economics and Philosophy*, 16: 175-204.

Zimmerman, M. J. (1996).*The Concept of Moral Obligation*. Cambridge: Cambridge University Press.