Economics and Philosophy, **17** (2001) 21–38

SYMPOSIUM ON AMARTYA SEN'S PHILOSOPHY: 2
UNSTRAPPING THE STRAITJACKET OF 'PREFERENCE': A COMMENT ON AMARTYA SEN'S CONTRIBUTIONS TO PHILOSOPHY AND ECONOMICS

Elizabeth Anderson
University of Michigan, Ann Arbor

1. AMARTYA SEN'S CRITIQUE OF THE CONCEPT OF PREFERENCE

The concept of preference dominates economic theory today. It performs a triple duty for economists, grounding their theories of individual behavior, welfare, and rationality. Microeconomic theory assumes that individuals act so as to maximize their utility – that is, to maximize the degree to which their preferences are satisfied. Welfare economics defines individual welfare in terms of preference satisfaction or utility, and social welfare as a function of individual preferences. Finally, economists assume that the rational act is the act that maximally satisfies an individual's preferences. The habit of framing problems in terms of the concept of preference is now so entrenched that economists rarely entertain alternatives.

In this commentary, I would like to explore and extend Amartya Sen's critique of the concept of preference. A critique of a concept is not a rejection of that concept, but an exploration of its various meanings and limitations. One way to expose the limitations of a concept is by introducing new concepts that have different meanings but can plausibly contend for some of the same uses to which the criticized concept is typically put. The introduction of such new concepts gives us choices about how to think that we did not clearly envision before. Before envisioning these alternatives, our use of the concept under question is

https://doi.org/10.1017/S0266267101000128 Published online by Cambridge University Press

dogmatic. We deploy it automatically, unquestioningly, because it seems as if it is the inevitable conceptual framework within which inquiry must proceed. By envisioning alternatives, we convert dogmas into *tools*: ideas that we can *choose* to use or not, depending on how well the use of these ideas suits our investigative purposes.

Within economics, the representation of human behavior and normative judgements in terms of the concept of preference has long functioned as a dogma. Sen has done more than any other economist to convert this dogma into a tool, by offering alternative tools for economists to use, for both explanatory and normative purposes. A good starting point for exploring Sen's instrumentalizing of the concept of preference for *explanatory* purposes is his classic paper, 'Behavior and the Concept of Preference' (Sen, 1973). In that paper, he argued that there is a profound conceptual ambiguity at the heart of the economic theory of rational choice. Economists use the same concept of 'preference' to perform three distinct tasks: (a) to describe a person's choices; (b) to represent whatever motives underlie a person's choices; and (c) to represent a person's welfare. Sen pointed out that these are conceptually distinct, and that in particular one is not entitled to infer that a particular choice advanced the individual's welfare just because she made it voluntarily. Sen was later to elaborate this distinction, between preference in the wide sense of whatever states of affairs one values, and preference in the narrow sense of personal welfare or self-interest, as the distinction between agency and welfare (Sen, 1985).

His second point was more momentous. He argued that even if we consider preference in the wide sense, a person might not choose in such a way as to maximally satisfy her preferences. That is, she may act on some principle other than the maximization of her utility, even where utility was understood in the wide sense as encompassing any states of affairs she might value. Sen illustrated this point by contrasting two alternative explanations of cooperation in a prisoner's dilemma. First, a person could be sympathetic, and care about the welfare of the other party to the dilemma. This is a case of maximizing utility.[1] But second, a person could be motivated by moral principle, a social norm of responsibility, or what Sen was later to call 'commitment' (Sen, 1977). In the second case, the individual still prefers to serve less time in prison rather than more, and does not personally care about how well-off the other party is. Her choice cannot therefore be rationalized in terms of maximizing her utility, either in the narrow (self-interest) or the wide

[1] As Sen described the case, which he labeled 'sympathy', it was a matter of maximizing utility in the narrow, egoistic sense. The sympathetic party feels pain at observing the suffering of others, and helps them in order to relieve his own pain. This is distinct from the case of genuine altruism, where an individual helps out of direct concern for the other. This would be a case of maximizing utility in the broad sense.

sense. Instead, she decides to act on a principle of choice other than to maximally satisfy her personal preferences. She *suspends* individualistic calculation of consequences, acting *as if* she cared personally about the other party, even though she does not.[2]

Thus, for explanatory purposes, Sen instrumentalized the concept of preference in two ways: first, by *disambiguating* the concept, replacing it with three distinct concepts (choice, underlying motive, and welfare), and second, by articulating an alternative model of behavior, commitment, that was not framed in terms of preference satisfaction at all. In place of one dogma, he gave us four tools.

Sen's wide-ranging critique of the normative uses of 'preference' is even more well-known. Within normative economics, the concept of preference is used to (1) make judgements about individual welfare; (2) make judgements about the overall good of society; and (3) articulate a principle of rational choice (utility maximization). Sen has criticized excessive reliance on the concept of preference on all three fronts.

With respect to judgements of individual welfare, Sen points out that none of the three *explanatory* concepts of 'preference' (choice, underlying motive, and perceived self-interest) quite does the job of offering a good measure of a person's well-being. Preference conceived as choice or motive fails to do so, because people have motives wider than and sometimes even counter to their self-interest, and often choose accordingly. Preference understood as perceived self-interest also often fails as a measure of welfare, particularly with respect to the severely disadvantaged. Seriously disadvantaged people often adapt their self-interested preferences to their limited opportunities: they lower their aspirations to avoid frustration. 'The extent of a person's deprivation . . . may not at all show up in the metric of desire-fulfillment, even though he or she may be quite unable to be adequately nourished, decently clothed, minimally educated, and properly sheltered' (Sen, 1992, p. 55).

With respect to judgements of the overall good of society, Sen argues that utility information alone is insufficient to ground ethical or socially rational evaluations. This of course follows from the fact that utility is not a good measure of welfare. But in addition, the amount of freedom

[2] I compress some of Sen's intellectual history here to sharpen the conceptual point. In 'Behavior and the Concept of Preference' (Sen, 1973) and 'Rational Fools' (Sen, 1977) Sen did not clearly distinguish, within the category of commitment, (a) maximizing utility in the broad sense, (as in altruistic or utilitarian action) from (b) action that is properly represented as choosing the best option within conventional or ethical constraints (for example, Kantian ethics) whose authority is accepted by the agent. In 'Maximization and the Act of Choice' (Sen, 1997, pp. 769–71) marks this distinction clearly. In this commentary, I confine my use of the term 'commitment' to motives of type (b), which Sen correctly argues are not properly represented as maximizing utility in any sense, narrow or broad.

people enjoy is important, over and above how much welfare they get from choosing their most preferred option (Sen, 1985). It also independently matters that social outcomes be produced by just, fair, and rights-respecting procedures (Sen, 1995, p. 13). And fairness in the distribution of outcomes matters, too, beyond the total 'amount' of welfare enjoyed in society (Sen, 1995, pp. 9–10). Again, in place of one dogma, Sen gives us many tools.

All of these observations call for enriching the informational basis of individual and social welfare judgements beyond that provided by the concept of preference. Sen has proposed 'capabilities' and 'functionings' as appropriate objective measures of freedom and welfare, a judgement with which I concur (Anderson, 1999).

What about judgements of rationality? Here, I think, there is a lacuna in Sen's work. Sen's analysis of utility maximization in prisoner's dilemma situations certainly suggests that the individual who always acts on his preferences can be a fool and a social misfit in circumstances where acting on social norms of cooperation brings about better consequences for all (Sen, 1977). Yet, Sen does not propose an alternative, non-preference-based conception of rationality in terms of which committed action makes sense. To be sure, committed action is often socially and ethically desirable. But this only raises the question of how it can be rational for the individual to act on socially and ethically desirable principles, when so acting does not advance the satisfaction of her broad or narrow preferences. Moreover, not all committed action is desirable from a moral point of view. Recall that the original setting for the prisoner's dilemma involves two presumably guilty co-conspirators who have an interest in getting away with their crime. Although, given their right against self-incrimination, they are within their rights in remaining silent, one can hardly deem their silence as morally desirable. From a moral point of view, it would be better if either or both confessed to their crimes. It is only from a point of view including the criminals alone that 'cooperation' in this prisoner's dilemma is desirable.

In this essay, I shall explore this lacuna in Sen's work. I shall argue that a full understanding of the *rationality* of committed action requires us to enrich the information basis of the theory of rationality beyond the concept of individual preference, in two ways. First, we need to devise a non-preference-based conception of *reasons for action*. Second, we need robust conceptions of *collective agency* and *individual identity*. We can find hints of both ideas in Sen's work, which I shall develop further. Briefly, committed action turns out to be action on principles (reasons) that it is rational for *us* (any group of people, regarded as a collective agent) to adopt, and thus that it is rational for any individual who *identifies* as a member of that group to act on. We can then build on this idea to get from the rationality to the morality of committed action. If it would be

rational for a collective encompassing *all of humanity* to adopt a certain principle of committed action, then action on that principle is *morally right*. It is then rational for anyone who identifies as a member of this cosmopolitan community of humanity – we could call it the Kingdom of Ends – to act on such a principle. I thus propose to develop some of Sen's ideas in a more Kantian direction than he has been willing to go. In the last section of this commentary, I shall consider some implications of this analysis for understanding the plight of women across the world.

2. WHY THE ECONOMIC THEORY OF RATIONAL CHOICE CANNOT GROUND A GENERAL SOLUTION TO PRISONER'S DILEMMAS

To see more clearly how committed action is problematic from the standpoint of rational choice theory, let us get the phenomenon clearly into focus. Suppose some people face a situation in which the payoffs take the form of a prisoner's dilemma. That is, for each party to the dilemma, non-cooperation will bring about a better or preferred state of affairs than cooperation, regardless of what the other parties to the dilemma do. Non-cooperation *dominates* cooperation. However, if everyone cooperated, this would bring about a better state of affairs from each person's perspective than if no one cooperated. The puzzle is to understand how it can be rational for any of the parties to cooperate under these conditions.

Proposed cooperative solutions to prisoner's dilemmas within preference-satisfaction theories of rational choice try to show that the payoffs recorded in the dilemma do not reflect all the preferences of the parties. When the parties' full preferences are taken into account, they really would maximize their utilities by cooperating. There are two main ways to broaden people's preferences in the required way. First, one may suppose that the parties have sympathetic or altruistic preferences. They care not just about the payoffs to themselves, but about the payoffs to others. Second, one may suppose that the act of cooperation is valued intrinsically, as a consequence in itself, and not just valued for the sake of the other consequences it brings about.

Neither proposal shows that it is rational to cooperate in a prisoner's dilemma. Rather, they assert that the payoffs to the parties, once fully accounted for, do not really have the structure of a prisoner's dilemma. But it is not so easy to evade the problem of prisoner's dilemmas in these ways.

Consider first the altruistic evasion. This supposes that if only people cared about other's interests, prisoner's dilemmas would not arise. This is an error. The key feature of the payoffs that generates prisoner's dilemmas is not the fact that people care only about the payoffs to themselves. It is the fact that any single person's action,

https://doi.org/10.1017/S0266267101000128 Published online by Cambridge University Press

considered in isolation, has an expected marginal payoff of zero, or close to zero, with respect to the socially desirable outcome. Such a situation can therefore occur even if everyone has altruistic or public-spirited preferences. Thus, it arises in almost any n-person public goods case, even when people value public goods for the sake of others besides themselves and think they ought to be provided.

Consider, for example, the case of voting. Democracy would collapse if the people did not go to the polls, so, assuming democracy is good for the people, mass voting can be regarded as a public good. Moreover, from the perspective of the supporters of any particular candidate on the ballot, the election of that candidate is also a public good. That is, supporters suppose that the election of their candidate would be better for society, not just for themselves, and may well prefer their candidate for this public-spirited reason. Yet when deciding whether to vote, each partisan who accepts the principle of expected utility reasons as follows: regardless of how my fellow partisans vote, the chances that *my* ballot will make a difference to the outcome of the election are negligible. Therefore, if there is the slightest inconvenience to me (or – thinking altruistically now – inconvenience to others!) from my voting, this will certainly outweigh the expected marginal positive impact of my voting. But there is always some inconvenience. So I ought not to vote. The conclusion follows not because each partisan is selfish, but because each partisan correctly reasons that her marginal impact on the outcome each prefers from a public-spirited standpoint is negligible.

What generates prisoner's dilemmas, then, is people's acceptance of a principle of rational choice that has an act-consequentialist form. As long as people judge the value of their action in terms of its expected marginal causal impact, their cooperation in prisoner's dilemmas cannot be counted on, even if their underlying preferences are altruistic.

The second strategy for evading prisoner's dilemmas within an act-consequentialist framework is to postulate that the parties value cooperative action intrinsically, apart from its consequences. This is Sen's favored strategy for resolving the paradox of voting: to postulate that people enjoy the act of voting in itself, apart from its consequences (Sen, 1997, p. 750).[3] I do not believe that this attempt to rationalize voting makes sense. If the act of marking ballots with one's preferences really did have no further consequences, it would be absurd for people to

[3] In his replies to a version of this paper delivered at the American Philosophical Association Central Division Meetings (Chicago, April 22, 2000), Sen observed that he had offered this only as an explanation of why some people vote, not as a consideration that could make voting rational. My subsequent comments therefore address not Sen himself, but anyone who thinks attaching a value to the act of voting itself could make voting rational.

value it. Suppose an oligarchy announced that, henceforth, the people would be allowed to mark their preferences for candidates on secret ballots. Only the ballots would be burned without being counted, and candidates would be selected by the oligarchy using its traditional undemocratic methods. Here would be an opportunity for people to 'vote' for its own sake, apart from its consequences. But only a fool would value it.

One might object: but that is not really voting. This would be right. But that is just to admit that the act of voting makes no sense apart from an appreciation of its causal role in selecting public officeholders. And this, I would bet, is how most voters view the matter as well. When they go to the polls, they generally do so with the end in view of helping their favored candidates get elected. And the same could likely be said about any actions, such as paying taxes, that collectively result in the production of public goods, although any of these actions taken in isolation have negligible marginal impact. People would not, in general, find any value in such acts if there were no causal connection between them and the production of the public goods in question.

In any event, such an *ad hoc* solution to prisoner's dilemmas does not have the general features needed to vindicate the rationality of cooperation. In 'Rational Fools', Sen suggests that those who fail to cooperate in prisoner's dilemmas, however much they may be maximizing their expected utilities, are acting foolishly. But it is hardly foolish to not prefer the act of cooperating in itself, apart from its consequences. What is foolish about non-cooperators is not their preferences, which are perfectly understandable, but their principle of rational choice. And what makes *that* principle foolish is its act-consequentialist structure. Any principle of rational choice that evaluates an individual's act solely according to its marginal causal impact on valued outcomes will meet the same difficulties. This is one powerful reason why many people are drawn away from act-consequentialism toward rule-consequentialism, or toward non-consequentialist frameworks.

3. THE RATIONAL BASIS OF COMMITTED ACTION

When people face a genuine prisoner's dilemma – that is, one that retains a PD payoff structure even when consequences are valued unselfishly and any intrinsic preference for performing the cooperative act is factored in – cooperation can only be rationalized in terms of a non-act-consequentialist principle of rational choice. Thus, people's cooperation in such cases must be based on a principle of choice *other* than the maximum satisfaction of their preferences. This is what I, following Sen, have been calling 'committed action'. Sen agrees that some people do engage in committed action for reasons that are not

https://doi.org/10.1017/S0266267101000128 Published online by Cambridge University Press

properly represented in terms of maximizing utility (Sen, 1997, pp. 769–71). Such action is rather a matter of following social conventions, norms, or conceptions of ethical duty. If we regard cooperation in some such cases as socially desirable, we need to find an alternative account of rational action that does not define it in terms of its marginal causal impact on desired outcomes. The question is, then, what other principle of choice there could be, and how to understand it as a *rational* principle.

Sen offers us several hints in disparate writings from which we can build a coherent account. First, he suggests that what people do depends on their understanding of their ***identities***, which may be constituted by membership in various social groups. For example, he points to data suggesting that Indian women tend to conceive of themselves more as members of their families than as individual selves, and choose accordingly (Sen, 1990, pp. 125–6). People may also identify with their occupation, the firm where they are employed, various associations and clubs, their nation, caste, religion, and so forth, and choose on the basis of these identities. Second, in discussing the solution to one-shot prisoner's dilemmas, he suggests that the parties can reach the collectively desirable action by 'treating as ''the unit of selection'' their ***joint*** strategy'. This treatment 'would entail a violation of the standard formulation of individual rational choice' (Sen, 1994, p. 387). Third, Sen has stressed the importance of ***discussion*** in changing the bases upon which people act (Sen, 1995, p. 18; Sen, 1999).

We can integrate Sen's disparate hints into a unified account of the rationality of committed action as follows. Suppose the parties to a prisoner's dilemma ***identify*** with one another as common members of a social group. Then they would pose to themselves a different practical question. Each would ask, not 'What should *I* do?', but rather 'What should *we* do?'. To ask the latter question is to deliberate from a standpoint that one can coherently regard everyone else in the group taking up as well. It is to regard oneself as acting in concert with the other parties, as a single body. Any group of people whose members refer to one another as 'we' and who, in virtue of that fact, see themselves as ready to be jointly ***committed*** to acting together, will properly regard the object of their choice to be a single ***joint strategy***. They will thereby constitute themselves as members of a single collective agent (Gilbert, 1990).

How would such a body evaluate different proposed policies or reasons for action? They would ***discuss*** them together, and try to reach a common point of view from which to assess them. In the classic prisoner's dilemma scenario, the parties are not able to discuss what they should do. Nevertheless, each can still take up the standpoint of collective deliberation, try to figure out what the outcome of such

discussion would be, and act accordingly. The key to figuring out this outcome is that it is a constitutive principle of a collective agent (a 'plural subject' or 'we') that whatever can count as a reason for action for one member of the collective must count as a reason for all. That is, in regarding themselves as members of a single collective agency, the parties are committed to acting only on reasons that are universalizable to their membership.

The universalization principle *rules out* the principle of maximizing expected utility (individual preference satisfaction) as an acceptable principle of rational choice for members of a collective agency who constitute the parties to a prisoner's dilemma. To make this demonstration vivid, consider the case of members of a political party, P, who agree that the best outcome for all would be to elect their candidate, A. As we have seen, according to the principle of expected utility, even this shared preference does not give a reason for any of the members of P to vote for A, if each personally finds voting inconvenient. For each member reasons that the expected marginal impact of his vote on the preferred outcome is so negligible that even a trifling inconvenience is enough to outweigh it. Is this reasoning valid, from the standpoint of P? It is valid, only if the members of P could jointly accept a trifling inconvenience as a reason for *all* the members of P not to vote. It is evident that they could not accept this as a reason not to vote. For if every member of P accepted this as a reason not to vote, few members of P would vote, and this would defeat their joint aim of electing A. To act on the principle of expected utility would be self-defeating from the standpoint of the collective agency. The principle of expected utility is therefore invalid for members of the group.

This argument does not turn on the members of the group having a common aim prior to collective deliberation. It turns on the fact that they accept as reasons for action only those considerations that each person would be willing to accept as reasons for everyone to act. In effect, each person asks, 'what reasons do *we* have to act?' Only the reasons that *we* can share are reasons on which people who identify as 'we' can accept as a ground for their action. Because *we* cannot will that each person try to free-ride on the efforts of others, we cannot accept the reasoning that supports unconditional non-cooperation in prisoner's dilemmas. Upon eliminating non-universalizable policies, the members of the group find that whatever jointly acceptable policies remain require them to, in effect, aim at what would be best for all. This is a conclusion, not a presupposition, of their deliberations.

If the only alternative to unconditional non-cooperation were a policy of unconditional cooperation, then the parties would will that everyone cooperate. In reality, matters are more complicated, because the parties would need to consider various policies of mixed and conditional

cooperation.[4] There is no general way to tell, in advance of discussion among the parties, what joint policy would be rational for all to accept. It would make sense to suppose, however, that they would begin with a presumption in favor of cooperating, and then consider what kinds of conditions all could accept as excusing members of the collective from the duty to cooperate.

In taking up the perspective of collective agency, the parties to an n-person prisoner's dilemma can see their actions as jointly advancing a desirable goal even though none of their actions taken in isolation has a positive marginal impact on that goal. Although *I* may not be able to regard *my* action taken alone as making a difference to the goal, *we* can regard *our* actions taken together as doing so. Insofar as I identify with the group, my reason for acting is: to do my part in advancing what we are willing together.

This argument for cooperating in prisoner's dilemmas shows how it *can* be rational to cooperate. It does not show that it is categorically *irrational* to follow the principle of expected utility, or to fail to cooperate in prisoner's dilemmas.[5] Rather, I make the following priority claim:

> *The Priority of Identity to Rational Principle:* what principle of choice it is rational to act on depends on a prior determination of personal identity, of who one is.

The validity of the principle of expected utility (maximizing the satisfaction of one's personal preferences) is *conditional* on regarding oneself as an isolated individual, not a member of any collective agency. In contexts where one regards oneself as a member of a social group, this principle will in general be invalid, because it is not universalizable among the members of the group. What is universalizable would be various principles of committed action (not necessarily of unconditional commitment).

This argument does not show that we must identify with any particular social group. *A fortiori*, it does not require that we act on our ascribed social identities of gender, race, caste, ethnicity, nationality, and so forth – that is, the group identities we have in virtue of the ways other people classify us. The concept of identity at stake in the theory of

[4] If it is known that only 80% cooperation in an n-person PD would be sufficient to produce the collectively desirable result, all might accept a mixed strategy – say, to toss a coin with an 80% chance of landing heads, and to cooperate if it turns up heads (Elster, 1989). More sensibly, they might accept a joint policy of conditional cooperation – for each to cooperate, if it imposes no more than minor inconvenience, but not, say, if one is gravely ill.

[5] Nor should we be eager to show that cooperation in prisoner's dilemmas is always rationally required: not all cooperation in prisoner's dilemmas is socially desirable (consider, for example, industrial cartels).

rationality is *practical*, not ascriptive.[6] From both a rational and a moral point of view, it would often be far better if we rejected our parochial ascriptive identities as bases of practical (action-governing) identification. The argument does not require that the people with whom one practically identifies be determined prior to or independently of the practical predicament at hand. Practical identification with others does not require any prior acquaintance or relationship. It only requires that we see ourselves as solving a problem by joining forces. As soon as one says 'Let's . . .' and the others manifest their willingness to go along, they have adopted a common practical identity as a social group with a shared goal. A shared intention is sufficient to constitute individuals as a social group with a common practical identity, and the only constraints on whom one may share an intention with are practical (that is, the conditions must be such that such sharing is possible). We could therefore find ourselves with good reason to practically identify and cooperate with perfect strangers.

What the priority of identity to rational principle does do is establish a rational permission to identify with others and join in a common agency. The argument does not claim that all action is or ought to be based on some group identity. One's practical identity for certain choices may be simply as an individual, with perhaps idiosyncratic interests or needs. If practical reason also permits regarding oneself as an isolated individual, then, for all I have argued, the principle of expected utility could well apply.[7] The principle of expected utility and the principle of group universalizability are, on the view I have developed so far, both conditionally valid. Their validity is conditional on the agent's self-conception, as an individual or a group member.[8]

Thus it appears that to adjudicate between these principles in any particular case, we need further principles of rational self-identification.

[6] I borrow the concept of practical identity from Christine Korsgaard (1996, 100–7).

[7] In other works, I have argued against preference-based theories of rational choice in favor of a rational attitude theory of rational choice (Anderson, 1993; Anderson, 1996). Rational attitudes such as love, respect, and admiration are more fundamental than preferences, but they do not yield a structure of preferences that satisfy the axioms of rational choice. I set this issue aside for the purposes of this commentary. In the next section, I will argue that the sort of individuals we all can be will not permit acting exclusively on the principle of expected utility.

[8] Postmodernists tirelessly remind us that the identities and boundaries of the self are not fixed, but contingent and changing. Although this refrain has been so often repeated that it has nearly become a mantra, here is an important place where it should be heeded. Of course it does not follow from this that the dimensions along which the self fractures (or rather, along which it unites with others) must involve the standard American quadruplet of race, class, gender, and sexual orientation. These are extremely important dimensions of social, political, and economic inequality, but they compete with numerous other dimensions of identity, and often do not form a desirable ground of practical identification.

In what contexts is it rational to identify with others, or, on the other hand, to conceive of oneself as an independent agent? *Which* others is it rational to identify with?

I want to avoid either of two extreme positions. One would say that we should identify with whoever else we stand in collective action problems, such as prisoner's dilemmas. The difficulty with this position is that in identifying with one group, we may thereby preclude identifying with another group. The prisoners who maintain a tacit conspiracy of silence in the paradigmatic prisoner's dilemma scenario are helping one another, but not society at large. Yet, it is hardly irrational for a criminal to identify with the larger society, feel remorse for his crimes, and confess because of that identification.

The other extreme position would say that identification with others would only be rational after the members of a group have manifested their willingness to join together, conditional on the others manifesting a like willingness. If this were so, then it would be irrational for people to cooperate in 'classic' prisoner's dilemma scenarios, where the parties are not allowed to communicate with one another in advance of making their choices. I think this condition on rational group identification is too strong. Common knowledge of everyone's (rational) conditional willingness to join together is sufficient, but not necessary, to make identification with the others rational. Where such common knowledge is absent, it could still make sense to take others *on trust*. In this case, one regards oneself as part of an imagined common agency, in the hope that others will join and make it real by cooperating.

We need hardly consider the suggestion that each agent should be a pure individualist at all times. To achieve most of the functionings constitutive of a person's wellbeing, and most of the larger projects worth pursuing, requires cooperation with others. The principle of expected utility would seem to apply comprehensively only to hermits.

On the other hand, particular collective agencies may fail to survive rational scrutiny, in which case uncritical identification with them would be irrational. (One might still *critically* identify with them, with the aim of reforming them so they operate on rationally acceptable principles.)

To make progress on the question of principles of rational identification, I suggest that we turn to a case study. Sen's own work on gender and cooperative conflict in the family is an apt place to look for illumination. His work shows how *both* actions on the principle of personal utility maximization *and* actions on a principle of family group identification systematically disadvantage women, often in extreme ways. It would seem that neither principle offers an adequate perspective from which to secure justice for women. The quest for a larger perspective, I shall suggest, puts a Kantian twist on some of the grand themes of Sen's life work.

4. WOMEN, COOPERATIVE CONFLICT, AND RATIONAL IDENTIFICATION

Women around the world are systematically disadvantaged in their access to income, wealth, and employment opportunities outside the home. In many parts of the world, especially Asia, these disadvantages are severe enough that women suffer substantially higher rates of malnutrition, morbidity, and mortality than men. Sen (1990) has estimated that 100 million women are 'missing' in Asia due to excess mortality stemming from material deprivation, parental neglect, and selective abortion.

Much of this material inequality can be traced to the institution of marriage and the gendered division of domestic labor (Sen, 1989; Sen, 1990; Okin, 1989). Because women have more unpaid domestic, child-rearing, and eldercare responsibilities, they have fewer opportunities to work outside the home for a wage. They therefore also invest less in acquiring the skills required for outside employment, and so are eligible for lower-paying jobs than men are. Women's assumption of domestic responsibilities frees men to work longer hours at income-generating activity and to acquire more human capital. The income husbands generate is thus not the product of their efforts alone, but a matter of joint production between them and their wives: were wives not performing the lioness's share of domestic labor, their husbands would have to devote some of their time to this work, and could only accept the kinds of employment open to workers with substantial domestic responsibilities. Of course, women share in the household income, so there is mutual advantage in the domestic division of labor. However, the division of the rewards from family cooperation is extremely unequal. Why do husbands and wives not share more equally in the rewards of their joint production?

The answer appears to involve elements of both economic rationality and commitment. To the extent that both husbands and wives are rational egoists, the answer is supplied by bargaining theory. Under common knowledge of their preferences and alternatives, the parties would bargain their way into a Nash equilibrium in which the division of gains from cooperation is heavily influenced by the vulnerability of the parties in the breakdown position (how badly off each would be if there was no cooperation). The worse off a woman would be on her own, the worse deal she gets in marriage. Women face substantial obstacles to developing and fully employing their income generating potential, due to sex discrimination in employment, legal and customary barriers against women seeking employment outside the home, lower access to education, legal barriers to female inheritance and property ownership, and the gendered division of paid employment, which

reserves most higher-paid jobs and jobs on promotional ladders to men. Divorce laws provide at best minimal support for divorced women, as well. Thus, the cost of exiting marriage is higher for women than for men, and men therefore enjoy a threat advantage that they can exploit within marriage.

Bargaining theory no doubt explains part of the material disadvantage that women suffer. In an insecure marriage, a woman's knowledge of how badly off she would be if her husband left her no doubt quells many complaints she would otherwise voice about the division of benefits and burdens in her marriage. However, the theory clashes with several features of women's circumstances and motives. Marriage is conceived as a realm of love and obligation, distinct from market exchange. This difference is marked in part by norms against naked bargaining among marriage partners. The woman who drives a hard bargain, who insists on an explicit quid pro quo for her services, is marked as a prostitute, ineligible for marriage. In societies that practice arranged marriages, especially when dowry is involved, the woman has little bargaining power at the time of betrothal, since she functions more as the object of bargaining rather than as a party to the negotiations. Social norms against women's bargaining and severe limitations on women's choice of partner make her less able to secure a share of household resources than if she were free to bargain as a rational egoist.

In addition to these factors, Sen argues that women's motivations are not those of the rational egoist. Their wide preferences include the interests of other people, often to such a high degree that they have difficulty conceiving of their own interests as distinct from those of their family members. When Indian women are asked about their own welfare, they typically answer in terms of how well their family is doing (Sen, 1990, pp. 125–6). The difficulty they have in perceiving the distinctiveness of their own interests further reduces their ability to claim a share of domestic resources.

Women also manifest committed motivations, in accepting social norms that devalue their contribution to domestic resources. Real-world voluntary distributions require legitimation in terms of a conception of justice. Considerations of desert or productive contribution therefore play a role, beyond self-interest, in determining the division of gains from cooperation. Women and men share a perception of relative contributions that gives greater credit to men's wage-earning than to women's unpaid domestic labor and lower-paid employment. The wife is cast as dependent on her husband, obscuring the ways in which his productivity depends on her providing for his physical needs in kind (cooking, cleaning, sewing, shopping, obtaining water and firewood, etc.) as well as on her assuming responsibility for caring for his children and sometimes his elderly parents as well. In accepting norms that

devalue their contributions, women again get less from marriage than they would have if they were purely self-interested.

Sen's work on gender and the division of family resources thus places in social context the two types of cooperative motivation he identified as possible solutions to collective action problems – altruistic preferences, and commitment to social norms seen as right, legitimate, or obligatory. Women accept a lesser share of family resources not just because (a) they have little bargaining power, but because (b) they think that they ought not act like a hard bargainer, (c) have a hard time seeing themselves in this role in any event, given the difficulty they have conceiving of their interests as distinct from their family's interests, and (d) because they think they deserve, and hence ought to accept, only the little they are getting. Women are even worse off than they would be if they were rational egoists, bargaining as economic models suppose.

Does this mean that women ought rather to conceive of themselves as rational egoists than as family members? It is not clear that someone can function in a spousal or parental role with the self-conception of an egoist. In any event, women would still get a raw deal in marriage even if they were to conceive of themselves this way.

Luckily, we do not have to choose between these two dismal options, of rational egoism or commitment to an unjust form of group agency. The perspective of rational egoism proves highly useful for criticizing oppressive commitments. But it points beyond itself. Bargaining theory recommends that, if women are to improve their situation, they ought to seek employment outside the home. This will reduce the cost of exit from marriage and thereby improve women's prospects within marriage. But outside employment has additional effects not predicted by bargaining theory. First, given that outside income is more salient to husbands and wives as a productive contribution to household resources, wives acquire a greater perceived claim of desert to larger shares of those resources. Second, outside employment puts women into contact with diverse others who do not identify solely as family members. Such contact provides information about wider opportunities outside the home, and can inspire women to take a more critical stance toward their domestic identity and commitments.

Third, outside employment gives women opportunities to acquire new identities, besides their family-bound identities as wives and mothers. They now identify as workers in a cooperative enterprise distinct from the family. This is a new form of collective identity. Once they have more than one collective identity, whose constitutive commitments may not be fully coordinated with one another, they may find themselves needing to acquire yet another identity. This is the identity of an individual self, who is authorized to adjudicate the conflicts among its various constitutive collective identities.

This scenario suggests a different genealogy of the individual from that assumed in standard economic models. Within economic theory, the individual is assumed to be given, with interests and preferences defined independently of and prior to her joining any group. Whether it is rational for her to join a collective agency then depends upon whether so joining would advance the satisfaction of her individual preferences. She joins if, and only if, joining would maximize her expected utility.

This model is unrealistic in societies where women do not have an option *not* to join certain groups, and lack the option to join others. Where women's identities are *comprehensively* defined in terms of their family roles – as daughter, wife, mother – they are given no choice about whether to so identify themselves, and have no other options to identify with other groups. Where women are deprived of such choices in how they identify, they have no opportunity to become the individuals who could choose otherwise. To become such individuals, women need mobility between different types of collective agency. They need to be free to move from the family sphere to other spheres of social organization – outside employment, politics, women's associations, friendships unregulated by their male relatives, and so forth. Thus, if the freedom to *function as an individual* is an important kind of agency capability for a human being, then we cannot assume that women around the world already function as individuals, but must dramatically revise the norms constitutive of 'women's' roles to make this possible.

I am arguing that individuality – identification as an individual person – emerges out of a certain kind of social order. Such an order is defined by multiple, distinct spheres of social life, none of which comprehensively defines anyone's agency, individual freedom of mobility among those spheres, and individual membership in multiple spheres. Only when these social conditions are in place can people become individuals, understood as agents authorized to set their own priorities, on their own, according to an autonomously defined self-conception. Before that point, their priorities are set by the commitments of the collective agency to which they belong, and hence it is not up to them to set their priorities on their own. On the view of individuality I propose, then, acquiring a self-conception as an individual requires actually conceiving of oneself as a committed member of multiple social groups, among whose claims one must adjudicate in allocating one's own efforts.

But the freedom to determine one's own priorities in committing oneself to various groups depends on those groups limiting their demands on their members in ways that enable their members to identify with and function as members of multiple groups. Social groups do not exist in pre-established harmony; different groups often make incompatible demands on their members. One therefore cannot expect

groups to limit their demands on their members simply on the basis of considerations available internally to each group's perspective. To harmonize the demands of different groups requires adoption of a perspective that can coordinate them all. We could see this as a collective action problem that groups face with one another. To solve it requires that we transcend our various parochial identities and identify with a community that comprehends them all. Such a community is what Rawls (1971, pp. 527ff.) calls a 'social union of social unions'. Were we to expand this community of identity to the whole of humanity, we would deliberate from the standpoint of what Kant (1964, pp. 100–3) called 'the kingdom of ends'. This is the point at which rationality coincides with morality.

I have no argument that would show that identification as a member of a universal community of humanity – a kingdom of ends – is rationally required. But, as economic development proceeds, we find ourselves more often generating problems, especially environmental, that need to be solved, and can only be solved, within a global system of cooperation. Identification with humanity as such may therefore become an historically urgent task.[9] A move to cosmopolitan identification has some powerfully attractive features. Besides providing a comprehensive perspective from which the collective action problems worth solving could be solved, a universal commitment to act from this perspective would secure the conditions for *everyone* being able to achieve an identity and agency as individuals. To see 'us' as comprehending all of humanity, is, however, to commit ourselves to placing significant constraints on appeal to preferences as reasons for action. The only sort of individual that *everyone* can be is one who identifies with multiple collective agencies as well as with humanity as a whole, and who therefore accepts multiple commitments, not grounded in individual preferences, as reasons for action. To understand the nature of rational choice, then, we need to enrich our information basis beyond individual preferences, and include the ideas of *identity*, *collective agency*, and *reasons for action*, where

[9] My reasoning in this section is dialectical rather than transcendental. In transcendental mode, we take our current, parochial practical identities as given and show that we rationally cannot accord them any authority without also identifying with humanity as such. For an example of this Kantian style of argument, see Korsgaard (1996, 120–5). In dialectical mode, we begin with our current, parochial identities and show how they generate practical problems that cannot be solved within the confines of those identities, or experiences that cannot be understood in their terms. The quest for a perspective that can make sense of our experiences and solve our problems leads to more and more expansive, cosmopolitan identifications, in an historical rather than a purely logical process. This is Hegel's way. For a superb illustration of how this dialectical process worked in one episode of American history see Oakes (1990) (arguing that the institution of slavery was undone by its internal contradictions, which simultaneously denied and demanded the moral agency and accountability of slaves).

the test for valid reasons is universalizability among those with whom one rationally identifies.

REFERENCES

Anderson, E. 1993. *Value in Ethics and Economics*. Harvard University Press
Anderson, E. 1996. 'Reasons, attitudes, and values: replies to Sturgeon and Piper'. *Ethics*, 106:538–54
Anderson, E. 1999. 'What is the point of equality?' *Ethics*, 109:287–337
Elster, J. 1989. *The Cement of Society*. Cambridge University Press
Gilbert, M. 1990. 'Walking together: a paradigmatic social phenomenon'. *Midwest Studies in Philosophy*, 15:1–14
Kant, I. 1964. *Groundwork for the Metaphysics of Morals*. Trans. H. J. Paton. Harper
Korsgaard, C. 1996. *The Sources of Normativity*. Cambridge University Press
Oakes, J. 1990. *Slavery and Freedom*. Alfred A. Knopf
Okin, S. 1989. *Justice, Gender, and the Family*. Basic Books
Rawls, J. 1971. *A Theory of Justice*. Harvard University Press
Sen, A. 1973. 'Behavior and the concept of preference'. *Economica*, 41:241–59
Sen, A. 1977. 'Rational fools: a critique of the behavioral foundations of economic theory'. *Philosophy and Public Affairs*, 6:317–44
Sen, A. 1985. 'Well-Being, agency, and freedom: the Dewey Lectures 1984'. *Journal of Philosophy*, 82:169–221
Sen, A. 1989. 'Cooperation, inequality, and the family'. *Population and Development Review*, 15 (Supplement: Rural Development and Population: Institutions and Policy):61–76
Sen, A. 1990. 'Gender and Cooperative Conflicts'. In *Persistent Inequalities*, 123–49. I. Tinker (ed.). Oxford University Press
Sen, A. 1990. 'More than a hundred million women are missing'. *New York Review of Books*, 20 December, 61–6
Sen, A. 1992. *Inequality Reexamined*. Harvard University Press
Sen, A. 1994. 'The formulation of rational choice'. *American Economic Review*, 84:385–90
Sen, A. 1995. 'Rationality and social choice'. *American Economic Review* 85:1–24
Sen, A. 1997. Maximization and the act of choice. *Econometrica*, 65:745–79
Sen, A. 1999. *Development as Freedom*. A. A. Knopf